Epidemiology of Metabolite Profile and Prostate Cancer Risk

Thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017



Julie Andersen Schmidt

Keble College

University of Oxford

Supervisors

Dr Ruth Travis, Dr Naomi Allen, Professor Tim Key

Nuffield Department of Population Health

Copyright © 2017 by Schmidt.

All rights reserved.

Abstract

Epidemiology of Metabolite Profile and Prostate Cancer Risk

Julie Andersen Schmidt, Keble College, University of Oxford, DPhil, Michaelmas 2017

Insulin-like growth factor-I (IGF-I) is the only known potentially modifiable risk factor for prostate cancer. Intake of dietary protein, especially from dairy products, might also be associated with risk and with circulating IGF-I, but it is not clear if amino acids play a role in these relationships. Moreover, investigations of circulating concentrations of metabolites might reveal novel risk factors for prostate cancer. This thesis investigates plasma concentrations of amino acids and other metabolites in relation to protein intake, IGF-I, and prostate cancer risk using data from the European Prospective Investigation into Cancer and Nutrition (EPIC).

To characterise plasma metabolite profile in men consuming markedly different amounts and types of animal products (meat-eaters, fish-eaters, vegetarians and vegans), cross-sectional analyses of 392 men in the EPIC-Oxford sub-cohort were conducted. Of 21 amino acids, six varied significantly by diet group, and the metabolite profile of vegans was different from those of other diet groups owing to lower concentrations of several glycerophospholipids and sphingolipids.

In a case-control study nested within EPIC, with a mean follow-up time of seven years, the relationship of plasma metabolites with risk of prostate cancer overall, by time to diagnosis, by tumour characteristics, and with risk of prostate cancer death, was investigated. Data from 1,077 matched sets suggested that seven metabolites, from various classes, were associated with risk of prostate cancer overall ($p < 0.05$). After correction for multiple testing, 12 glycerophospholipids were inversely associated with risk of advanced prostate cancer (the strongest $\text{OR}_{\text{1SD}} = 0.54$; 95%CI: $0.40 - 0.72$). In multivariate analyses, including data from 1,593 matched sets, principal component analysis (PCA) and treelet transform (TT) were used to identify patterns in metabolite profile, five of which were associated with risk of more aggressive tumour sub-types (high grade, advanced and aggressive disease) and/or prostate cancer death. There was a $\approx 50\%$ lower risk of advanced and high grade prostate cancer in men with metabolite profiles characterised by high glycerophospholipids and sphingolipids (for advanced $\text{OR}_{\text{TT, top vs bottom third}} = 0.48$; 95%CI: $0.31 - 0.74$), with similar results for high grade and PCA).

To investigate if associations between protein intake and circulating IGF-I may be mediated by plasma amino acid concentrations, cross-sectional analyses of amino acid concentrations with protein intake and IGF-I concentrations were carried out in 1,697 and 1,142 control participants, respectively, from the nested case-control study. Dairy protein intake was positively associated with concentrations of branched-chain amino acids and several other essential amino acids, while plant protein intake was strongly associated only with histidine. Serum IGF-I was positively associated with arginine and inversely with ornithine and certain amino acid ratios.

In conclusion, men with different dietary habits with respect to the consumption of protein types have different amino acid and metabolite profiles, and metabolite concentrations may be associated with risk of more high-risk prostate cancer sub-types (high grade, advanced and aggressive disease) and prostate cancer death. Further large-scale studies are needed to determine if metabolites play a role in aetiology or are markers of sub-clinical prostate cancer.

Acknowledgements

Firstly, I would like to thank the participants in the European Prospective Investigation into Cancer and Nutrition (EPIC) for providing blood samples and information on diet and lifestyle. Without their time and effort, the research presented in this thesis would not have been possible to conduct.

I am deeply grateful for the opportunity to study at the Cancer Epidemiology Unit (CEU), Nuffield Department for Population Health (NDPH), under the supervision of Dr Ruth Travis, Dr Naomi Allen and Prof Tim Key. Their rigorous and encouraging approach to research and supervising is inspiring and has made it a very enjoyable experience. I would also like to thank Georgina Fensom for her valuable contribution to this thesis via data preparation, analysis and presentation, and Paul Appleby for sharing his extensive knowledge about the EPIC datasets and analyses. A further thank you goes to the laboratory teams involved in the preparation of blood samples and the metabolomics assays. Firstly, the team responsible for all metabolomics assay at the laboratory at Nutrition and Metabolism Section, Biomarkers Group at the International Agency for Research on Cancer, Lyon, France – especially Dr Augustin Scalbert, Dr Sabina Rinaldi and David Achaintre for sharing their knowledge and providing high quality data. Secondly, Sarah Tipper and her team in the previous CEU laboratory and currently NDPH Wolfson Laboratories, University of Oxford, for preparing the samples for assaying and for letting me be involved in the lab.

I have enjoyed many inspiring discussions with my colleagues at CEU. The support and friendship from Sarah Floud and Kathryn Bradbury have been especially rewarding. Furthermore, I would like to thank Sarah for painstaking proof-reading of this thesis.

Last but not least, I thank my friends and family for their continuing support. A special thanks goes to Anton, who reminds me that there is more to life than epidemiology, and to Jakob Blaavand, who's companionship and understanding, as well as mathematical knowledge, have been an immense support.

The work for this thesis was funded by Cancer Research UK (grant numbers: 570/A16491 and C8221/A19170), World Cancer Research Fund (WCRF UK), as part of the World Cancer Research Fund International grant programme (grant number: 2014/1183). My fees were covered by the Nuffield Department of Clinical Medicine and the NDPH.

Contents

Abstract i

Acknowledgements iii

List of Tables xi

List of Figures xvii

List of Abbreviations xxi

1 Introduction 1

1.1 Rationale . . . 1

1.1.1 Research aims . . . 1

1.2 Outline of the thesis . . . 2

1.3 Role of the author in the preparation of this thesis . . . 2

1.4 Publications from this thesis . . . 3

2 Prostate cancer epidemiology and metabolomics 5

2.1 Introduction . . . 5

2.2 Prostate cancer epidemiology . . . 5

2.2.1 Geographical trends . . . 5

2.2.2 Temporal trends . . . 6

2.2.3 Tumour heterogeneity, PSA-testing and implications for analyses . . . 6

2.2.4 Established risk factors . . . 7

2.2.5 Potential risk factors . . . 9

2.2.6 Hypothesis on dairy protein, amino acids, IGF-I and prostate cancer risk 15

2.3 Metabolomics . . . 17

2.3.1 Analytical techniques . . . 17

2.3.2 Metabolomics in epidemiology . . . 18

2.3.3 Metabolites and prostate cancer risk . . . 19

2.4 Conclusion . . . 19

3 Data sources 21

3.1 Study population . . . 21

3.1.1 Recruitment . . . 21

3.1.2 Exclusion criteria . . . 22

3.2 Data collection . . . 22

3.2.1 Blood collection and processing . . . 23
3.2.2 Dietary data . . . 23
3.2.3 Other data . . . 25
3.2.4 Data storage, management and quality control . . . 25
3.3 Ethics . . . 25
3.4 Follow-up for cancer incidence and mortality . . . 26
3.4.1 Classification of prostate cancer and tumour sub-types . . . 26
3.5 Metabolomics assay . . . 26
3.5.1 The assay procedure . . . 27
3.5.2 Metabolite nomenclature . . . 28

4 Amino acid intakes and concentrations by diet group 33
4.1 Introduction . . . 33
4.1.1 Previous studies . . . 33
4.2 Aim . . . 34
4.3 Methods . . . 34
4.3.1 Study population . . . 34
4.3.2 Measurement of dietary intake . . . 35
4.3.3 Blood collection and laboratory analysis . . . 36
4.3.4 Covariates . . . 37
4.3.5 Statistical analysis . . . 37
4.4 Results . . . 39
4.4.1 Covariates by diet group . . . 39
4.4.2 Amino acid intakes and concentrations by covariates . . . 39
4.4.3 Intake of amino acids . . . 40
4.4.4 Plasma concentrations of amino acids . . . 40
4.4.5 Correlations between dietary factors and plasma concentrations of amino acids . . . 41
4.5 Discussion . . . 42
4.5.1 Main findings . . . 42
4.5.2 Intake of amino acids . . . 42
4.5.3 Amino acid concentrations . . . 43
4.5.4 Strengths and limitations . . . 44
4.5.5 Potential impact of findings . . . 44
4.6 Conclusion . . . 45
4.7 Tables . . . 46

4.8 Figures . . . 53

5 Metabolite concentrations by diet group 55

5.1 Introduction . . . 55

5.1.1 Previous studies . . . 55

5.2 Aim . . . 55

5.3 Subjects and methods . . . 56

5.3.1 Study population and data collection . . . 56

5.3.2 Blood collection and laboratory analysis . . . 56

5.3.3 Exclusions . . . 58

5.3.4 Dietary and anthropometric data . . . 58

5.3.5 Statistical analysis . . . 58

5.4 Results . . . 63

5.4.1 Participant characteristics, nutrient intake and blood sample characteristics by diet group . . . 63

5.4.2 Metabolite concentrations by covariates . . . 63

5.4.3 Individual metabolite concentration by diet group . . . 64

5.4.4 Correlations between metabolites . . . 67

5.4.5 Patterns in metabolite profile by diet group . . . 67

5.4.6 Variability in metabolomics data explained by diet group and other variables 69

5.5 Discussion . . . 69

5.5.1 Main findings . . . 69

5.5.2 Other studies . . . 70

5.5.3 Strengths and limitations . . . 71

5.5.4 Further research . . . 73

5.6 Conclusion . . . 73

5.7 Tables . . . 74

5.8 Figures . . . 105

6 Individual metabolite concentrations and prostate cancer risk 117

6.1 Introduction . . . 117

6.1.1 Prostate cancer . . . 117

6.1.2 Metabolomics . . . 117

6.1.3 Studies of metabolomics and prostate cancer . . . 117

6.1.4 How metabolites might be involved in prostate cancer . . . 118

6.2 Aim . . . 119

6.3 Methods . . . 119

6.3.1 Study population . . . 119
6.3.2 Blood collection and laboratory analysis . . . 120
6.3.3 Exclusions of metabolites and participants . . . 121
6.3.4 Statistical analysis . . . 122
6.4 Results . . . 124
6.4.1 Participant characteristics . . . 124
6.4.2 Distribution of metabolite concentrations . . . 125
6.4.3 Correlations . . . 125
6.4.4 Metabolite concentrations by covariates . . . 125
6.4.5 Risk of overall prostate cancer . . . 126
6.4.6 Risk of prostate cancer sub-types . . . 127
6.4.7 Risk of death from prostate cancer . . . 128
6.4.8 Sensitivity analyses . . . 128
6.5 Discussion . . . 128
6.5.1 Main findings . . . 128
6.5.2 Other studies . . . 129
6.5.3 Possible mechanisms . . . 131
6.5.4 Strengths and limitations . . . 132
6.6 Conclusion . . . 134
6.7 Tables . . . 135
6.8 Figures . . . 194

7 Metabolite patterns and prostate cancer risk 209
7.1 Introduction . . . 209
7.1.1 Previous studies . . . 209
7.2 Aim . . . 210
7.3 Methods . . . 210
7.3.1 Study population . . . 210
7.3.2 Blood collection and laboratory analysis . . . 212
7.3.3 Exclusions of metabolites and men . . . 212
7.3.4 Statistical analysis . . . 213
7.4 Results . . . 216
7.4.1 Participant characteristics . . . 216
7.4.2 Principal component analysis . . . 216
7.4.3 Treelet transform . . . 218
7.5 Discussion . . . 221

7.5.1 Main findings . . . 221
7.5.2 Other studies . . . 223
7.5.3 Potential mechanisms . . . 224
7.5.4 Strengths and limitations . . . 226
7.5.5 Further work . . . 228
7.6 Conclusion . . . 228
7.7 Tables . . . 230
7.8 Figures . . . 263

8 Intake of protein and amino acid concentrations 271
8.1 Introduction . . . 271
8.1.1 Previous studies . . . 271
8.2 Aim . . . 272
8.3 Methods . . . 272
8.3.1 Study population . . . 272
8.3.2 Dietary data . . . 272
8.3.3 Blood collection and laboratory analysis . . . 273
8.3.4 Statistical analysis . . . 273
8.4 Results . . . 276
8.4.1 Participant characteristics . . . 276
8.4.2 Intake of protein types and amino acid concentrations . . . 276
8.5 Discussion . . . 279
8.5.1 Main findings . . . 279
8.5.2 Previous studies and plausibility of findings . . . 280
8.5.3 Strengths and limitations . . . 282
8.5.4 Future work . . . 283
8.6 Conclusion . . . 284
8.7 Tables . . . 285
8.8 Figures . . . 314

9 Concentrations of amino acids and insulin-like growth factor-I 319
9.1 Introduction . . . 319
9.1.1 Previous studies . . . 319
9.2 Aim . . . 320
9.3 Methods . . . 320
9.3.1 Study population . . . 320
9.3.2 Blood collection and laboratory analysis . . . 320

9.3.3 Measurement of amino acids . . . 320
9.3.4 Statistical analysis . . . 321
9.4 Results . . . 323
9.4.1 Correction of IGF-I concentration . . . 323
9.4.2 Participant characteristics . . . 323
9.4.3 Concentrations of amino acids and IGF-I . . . 323
9.5 Discussion . . . 325
9.5.1 Main findings . . . 325
9.5.2 Other studies . . . 325
9.5.3 Plausibility of findings . . . 325
9.5.4 The role of amino acid concentrations in the association between dietary protein intake and IGF-I concentration . . . 326
9.5.5 Strengths and limitations . . . 327
9.5.6 Further work . . . 328
9.6 Conclusion . . . 328
9.7 Tables . . . 330
9.8 Figures . . . 339

10 Summary and recommendations for future research 341

10.1 Summary of findings . . . 341
10.1.1 Findings across analyses . . . 342
10.1.2 The thesis in context of existing literature . . . 343
10.2 Methodological considerations . . . 344
10.2.1 The role of chance . . . 344
10.2.2 The role of bias . . . 345
10.2.3 The role of confounding . . . 348
10.2.4 Summary . . . 348
10.3 Future research . . . 348
10.3.1 Other metabolomics platforms . . . 348
10.3.2 Temporality of the association between metabolites and prostate cancer risk 349
10.3.3 Genomics . . . 349
10.3.4 Tumour tissue . . . 349
10.3.5 Collaborative work . . . 350
10.4 Conclusion . . . 350

References 351

List of Tables

3.1 Abbreviations and full names of metabolites 29

4.1 Coefficients of variation of blinded pooled plasma samples 46
4.2 Characteristics, nutrient intakes and blood sample characteristics by habitual diet group 47
4.3 Amino acid intakes and concentrations significantly associated with participant characteristics and factors related to blood sampling and processing 48
4.4 Intakes of amino acids by habitual diet group 49
4.5 Plasma concentrations of amino acids by habitual diet group 50
4.6 Spearman's rank correlations between dietary intakes and plasma concentrations of amino acids 51
4.7 Spearman's rank correlations between intake of protein (% of energy) from food types and plasma amino acid concentrations 52

5.1 Summary of overall coefficients of variation 74
5.2 Summary of non-quantifiable measurements 74
5.3 The effect of excluding metabolites and men with missing metabolite concentration on the number of men with complete information for the included metabolites . . . 74
5.4 Characteristics and intakes of selected nutrients in 379 men in EPIC-Oxford by diet group 75
5.5 Metabolites significantly associated with participant characteristics and factors related to blood sampling and processing in EPIC-Oxford 76
5.6 Geometric mean metabolite concentrations by diet group and percentage difference compared to meat-eaters in EPIC-Oxford 81
5.7 Geometric mean metabolite concentrations in vegans and non-vegans and percentage difference compared to non-vegans in EPIC-Oxford 89
5.8 Geometric mean metabolite concentrations in non-vegan diet groups and percentage difference compared to meat-eaters in EPIC-Oxford 95
5.9 Mean principal component score by diet group and percentage difference compared to meat-eaters in EPIC-Oxford – main model 103
5.10 Mean principal component score by diet group and percentage difference compared to meat-eaters in EPIC-Oxford – alternative mode 104

6.1 Completeness and coefficients of variation for the metabolomics assay and corresponding exclusions of metabolites for assay round I 135

6.2 Characteristics of 1,077 prostate cancer cases and 1,077 controls in EPIC 138
6.3 Distribution of metabolite concentrations by case-control status in EPIC 139
6.4 Partial correlation coefficients between concentrations of total PSA and metabolites in EPIC 144
6.5 Metabolites significantly associated with covariates in 1,077 control participants from EPIC 149
6.6 Risk of prostate cancer per 1 standard deviation increase in metabolite concentrations in 1,077 matched case-control sets in EPIC 155
6.7 Departure from linearity of the association between metabolites and prostate cancer risk 158
6.8 Risk of prostate cancer in relation to fifths of metabolite concentrations in 1,077 matched case-control sets in EPIC 163
6.9 Metabolite concentrations and risk of prostate cancer by time to diagnosis in 1,077 matched case-control sets in EPIC 168
6.10 Metabolite concentrations and risk of prostate cancer by histological grade in 902 matched case-control sets in EPIC 173
6.11 Metabolite concentrations and risk of prostate cancer by stage in 664 matched case-control sets in EPIC 178
6.12 Metabolite concentrations and risk of aggressive prostate cancer in 654 matched case-control sets in EPIC 183
6.13 Metabolite concentrations and risk of death from prostate cancer in 127 matched case-control sets in EPIC 188
6.14 Risk of overall prostate cancer in 985 matched cases control sets after excluding the overlap in men with the analysis in EPIC-Heidelberg 191

7.1 Completeness and coefficients of variation for the metabolomics assay and corresponding exclusions of metabolites for assay round II 230
7.2 Stability assessment for treelet transform 233
7.3 Sign patterns of treelet components for stability assessment 234
7.4 Characteristics of 1,593 prostate cancer cases and 1,593 controls in EPIC 237
7.5 Variance in metabolite concentrations explained by the first seven components from principal component analysis and treelet transform 238
7.6 Loadings for the first seven principal components 239
7.7 Principal component scores and risk of overall prostate cancer in 1,593 matched case-control sets in EPIC 242

7.8 Principal component scores and risk of prostate cancer by time to diagnosis in 1,593 matched case-control sets in EPIC . . . 243
7.9 Principal component scores and risk of prostate cancer by tumour grade in 1,224 matched case-control sets in EPIC . . . 244
7.10 Principal component scores and risk of prostate cancer by tumour stage (localised and advanced) in 1,002 matched case-control sets in EPIC . . . 245
7.11 Principal component scoress and risk of prostate cancer by tumour stage (non-aggressive and aggressive) in 1,002 matched case-control sets in EPIC . . . 246
7.12 Principal component scores and risk of death from prostate cancer in 172 matched case-control sets in EPIC . . . 247
7.13 Loadings for the first seven treelet components . . . 248
7.14 Pearson's correlation coefficients between treelet component scores . . . 250
7.15 Loadings for the first seven treelet components at cut-level 102 . . . 251
7.16 Loadings for the first seven treelet components at cut-level 108 . . . 254
7.17 Pearson's correlation coefficients between principal and treelet component scores . . . 256
7.18 Treelet component scores and risk of overall prostate cancer in 1,593 matched case-control sets in EPIC . . . 257
7.19 Treelet component scores and risk of prostate cancer by time to diagnosis in 1,593 matched case-control sets in EPIC . . . 258
7.20 Treelet component scores and risk of prostate cancer by tumour grade in 1,224 matched case-control sets in EPIC . . . 259
7.21 Treelet component scores and risk of prostate cancer by tumour stage (localised and advanced) in 1,002 matched case-control sets in EPIC . . . 260
7.22 Treelet component scores and risk of prostate cancer by tumour stage (non-aggressive and aggressive) in 1,002 matched case-control sets in EPIC . . . 261
7.23 Treelet component scores and risk of death from prostate cancer in 172 matched case-control sets in EPIC . . . 262

8.1 Characteristics of the study population by thirds of total protein intake (% of energy) 285
8.2 Partial correlation coefficients between protein intake (% of energy) and amino acid concentrations and ratios in 1,697 men in EPIC – minimally-adjusted model . . . 286
8.3 Geometric mean amino acid concentrations and ratios by fifths of total protein intake in 1,697 men in EPIC – minimally-adjusted model . . . 288
8.4 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in total protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model . . . 290

8.5 Geometric mean amino acid concentrations and ratios by fifths of animal protein intake in 1,697 men in EPIC – minimally-adjusted model 291
8.6 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in animal protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model 293
8.7 Geometric mean amino acid concentrations and ratios by fifths of dairy protein intake in 1,697 men in EPIC – minimally-adjusted model 294
8.8 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in dairy protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model 296
8.9 Geometric mean amino acid concentrations and ratios by fifths of plant protein intake in 1,697 men in EPIC – minimally-adjusted model 297
8.10 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in plant protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model 299
8.11 Partial correlation coefficients between protein intake (% of energy) and amino acid concentrations and ratios in 1,697 men in EPIC – multivariable-adjusted model . . . 300
8.12 Geometric mean amino acid concentrations and ratios by fifths of total protein intake in 1,697 men in EPIC – multivariable-adjusted model 302
8.13 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in total protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model 304
8.14 Geometric mean amino acid concentrations and ratios by fifths of animal protein intake in 1,697 men in EPIC – multivariable-adjusted model 305
8.15 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in animal protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model 307
8.16 Geometric mean amino acid concentrations and ratios by fifths of dairy protein intake in 1,697 men in EPIC – multivariable-adjusted model 308
8.17 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in dairy protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model 310
8.18 Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in plant protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model 311

8.19 Geometric mean amino acid concentrations and ratios by fifths of plant protein intake in 1,697 men in EPIC – multivariable-adjusted model 312

9.1 Measured and corrected IGF-I concentration by follow-up phase 330
9.2 Characteristics of the study population by thirds of corrected IGF-I concentration . 331
9.3 Partial correlations of amino acid concentrations and ratios with corrected IGF-I in 1,142 men in EPIC 332
9.4 Geometric mean corrected IGF-I concentration by fifths of amino acid concentrations and ratios in 1,142 men in EPIC – minimally-adjusted model 333
9.5 Percentage difference in corrected IGF-I concentration for an 80 percentile increase in amino acid concentrations and ratios in 1,142 men in EPIC – minimally-adjusted 335
9.6 Geometric mean corrected IGF-I concentration by fifths of amino acid concentrations and ratios in 1,142 men in EPIC – multivariable-adjusted model 336
9.7 Percentage difference in corrected IGF-I concentration for an 80 percentile increment in amino acid concentrations and ratios in 1,142 men in EPIC – multivariable-adjusted model 338

List of Figures

1.1 Research questions 2

2.1 Incidence rates of prostate cancer by age-group and geographical region 7
2.2 Stimulation of proliferation and cell survival by amino acids and IGF-I via the PI3K/Akt/mTOR pathway 16

3.1 Countries and study centres in EPIC 22
3.2 EPIC-Oxford baseline food frequency questionnaire 24

4.1 Agreement between estimated protein intakes using the UK and US nutrient tables . 53
4.2 Correlation matrix for plasma amino acid concentrations 54

5.1 Exclusion of metabolites 105
5.2 Statistical significance of the associations between diet group and metabolite concentrations in 379 men from the EPIC-Oxford 106
5.3 Correlation coefficients of metabolites with cholesterols and apolipoproteins 107
5.4 Correlation coefficients between plasma metabolite concentrations 108
5.5 Scree plot for PCA of metabolite concentrations – main model 109
5.6 Loadings for the first four principal components by metabolite class – main model . 110
5.7 Score plot for principal component 1 and 3 by diet group – main model 111
5.8 Loading plot for principal component 1 and 3 – main model 112
5.9 Loading plot for principal component 1 and 2, and 2 and 3 – alternative model . . . 113
5.10 Score plots for principal component 1 and 2, and 2 and 3 by diet group – alternative model 114
5.11 Variability in metabolite concentrations explained by diet group and other characteristics 115

6.1 Flowchart of exclusions of metabolites and participants 194
6.2 Partial correlation coefficients between metabolite and total PSA concentrations . . 195
6.3 Variability in metabolite concentrations explained by covariates 196
6.4 Statistical significance of associations between metabolite concentrations and risk of overall prostate cancer in EPIC 197
6.5 Odds ratios for overall prostate cancer risk by concentration of selected metabolites in EPIC 198
6.6 Statistical significance of associations between metabolite concentrations and prostate cancer risk by time to diagnosis in EPIC 199

6.7 Statistical significance of the associations between metabolite concentrations and risk of high grade prostate cancer in EPIC . . . 200
6.8 Odds ratios for high grade prostate cancer by concentration of selected metabolites in EPIC . . . 201
6.9 Statistical significance of associations between metabolite concentrations and risk of advanced stage prostate cancer in EPIC . . . 202
6.10 Odds ratios for advanced stage prostate cancer by concentration of selected metabolites in EPIC . . . 203
6.11 Statistical significance of the associations between metabolite concentrations and risk of aggressive prostate cancer in EPIC . . . 204
6.12 Odds ratios for aggressive prostate cancer by concentration of selected metabolites in EPIC . . . 205
6.13 Statistical significance of associations between metabolite concentrations and risk of death from prostate cancer in EPIC . . . 206
6.14 Odds ratios for death from prostate cancer by selected concentration of metabolites in EPIC . . . 207

7.1 Flowchart of exclusions of metabolites and participants . . . 263
7.2 Cluster tree for treelet transform . . . 264
7.3 Cross-validation of cut-level for treelet transform . . . 265
7.4 Scree plot for treelet transform . . . 265
7.5 Scree plot for principal component analysis . . . 266
7.6 Loadings for the first seven principal components . . . 267
7.7 Loadings for the first seven treelet components . . . 269

8.1 Mean concentrations of branched-chain amino acids by fifths of protein intakes in 1,697 men in EPIC . . . 314
8.2 Mean concentrations of selected essential amino acids by fifths of protein intakes in 1,697 men in EPIC . . . 315
8.3 Mean concentrations of selected non-essential amino acids by fifths of protein intakes in 1,697 men in EPIC . . . 316
8.4 Mean concentrations of selected non-standard amino acids by fifths of protein intakes in 1,697 men in EPIC . . . 317

9.1 Schematic overview of the associations of selected amino acids with protein intake and circulating IGF-I concentration observed in this thesis . . . 327

9.2 Corrected IGF-I concentration by fifths of selected amino acid concentrations in 1,142 men in EPIC 339
9.3 Corrected IGF-I concentration by fifths of selected amino acid ratios in 1,142 men in EPIC 340

10.1 Association of selected amino acid concentrations with diet, IGF-I concentration and risk of prostate cancer as observed in this thesis 343

List of Abbreviations

a	Acyl
aa	'Diacyl' if preceded by PC; 'Amino acids' if preceded by anything else, e.g. essential
ae	Acyl-alkyl
ANOVA	Analysis of variance
Apo	Apolipoprotein
Arg	Arginine
Asn	Asparagine
Asp	Aspartate
ATBC	Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study
BMI	Body mass index
CEU	Cancer Epidemiology Unit
CI	Confidence interval
Cit	Citrulline
COMETS	COnsortium of METabolomics Studies
CUP	Continuous Update Project
CV	Coefficient of variation
C*x:y*	x and y denote the total number of carbon atoms and double bonds, respectively, in acylcarnitines, glycerophospholipid and sphingolipids
DNA	Deoxyribonucleic acid
EHNBPCCG	Endogenous Hormones, Nutritional Biomarkers and Prostate Cancer Collaborative Group
EPIC	European Prospective Investigation into Cancer and Nutrition

ESI	Electrospray ionization
FDR	False discovery rate
FFQ	Food frequency questionnaire
FIA	Flow injection analysis
FWER	Family wise error rate
GC	Gas chromatography
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
HDL	High density lipoprotein
HPLC	High performance liquid chromatography
His	Histidine
ICC	Intra-class correlation coefficient
ICD-10	10th revision of the International Statistical Classification of Diseases and Related Health Problems
IARC	International Agency for Research on Cancer
Ile	Isoleucine
IGF-I	Insulin-like growth factor-I
Kurt	Kurtosis
LC	Liquid chromatography
Leu	Leucine

Lys	Lysine
Met	Methionine
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
mTOR	Mammalian target of rapamycin
NOS	Nitric oxide synthase
OH	Hydroxy
OR	Odds Ratio
Orn	Ornithine
p_{adj}	Adjusted p-value, i.e. the p-value after correction for multiple testing
PC	Phosphatidylcholine
PC#	Principal component number #
PCA	Principal component analysis
p_{diff}	p-value for difference
Phe	Phenylalanine
p_{het}	p-value for heterogeneity
PI3K-Akt-mTOR	Phosphatidylinositol 3-Kinase-Akt-mammalian target of rapamycin
PLCO	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial
Pro	Proline
PSA	Prostate-specific antigen
PTEN	Phosphatase and tensin homolog
r	Correlation coefficient

RNA Ribonucleic acid

RR Relative risk

SD Standard deviation

SELECT Selenium and Vitamin E Cancer Prevention Trial

Ser Serine

Skew Skewness

SM Sphingomyelin

Thr Threonine

TNM Tumour-Node-Metastasis

Trp Tryptophan

TC Treelet component

TT Treelet transform

Tyr Tyrosine

t4-OH-Pro trans-4-hydroxyproline

UPLC Ultra-high performance liquid chromatography

US United States of America

USDA United States Department of Agriculture

UK United Kingdom

Val Valine

WCRF World Cancer Research Fund

WHO World Health Organization

1. Introduction

1.1 Rationale

Prostate cancer is the second most commonly diagnosed cancer in men worldwide [1] but no potentially modifiable risk factors, besides circulating insulin-like growth factor-I (IGF-I) concentration [2], have been established. IGF-I stimulates cell proliferation through the PI3K-Akt-mTOR pathway [3], the activity of which might be influenced by dietary protein composition, as branched-chain amino acids can upregulate mTOR activity [4, 5]. In epidemiological research, IGF-I concentration has been associated with dietary intake of protein, especially from dairy products [6, 7] which have also been weakly positively related to prostate cancer risk [8, 9].

Metabolomics may help us to understand the mechanisms underlying these associations and to identify novel risk factors for prostate cancer [10, 11]. Metabolomics is the identification and quantification of metabolites (i.e. low-molecular weight reactants, intermediates or products of biochemical reactions) in a biological system [12, 13], and since metabolite concentrations are affected by dietary, lifestyle, environmental and genetic factors, the measurements provide a snapshot of biological activity [11, 14].

1.1.1 Research aims

The main aims of this thesis are to investigate the prospective association between circulating metabolite concentrations and risk of prostate cancer and if dietary and hormonal factors might be linked to this association. More specifically, the research questions listed below will be investigated using data from the European Prospective Investigation into Cancer and Nutrition (EPIC). A schematic representation of the research questions is shown in Figure 1.1 on the following page.

- Do plasma concentrations of amino acids and other metabolites vary between meat-eaters, fish-eaters, vegetarians and vegans?
- Are plasma concentrations of metabolites associated with risk of prostate cancer?
- Are intakes of protein from different dietary sources related to plasma concentrations of amino acids?
- Are there associations between circulating concentrations of amino acids and IGF-I?

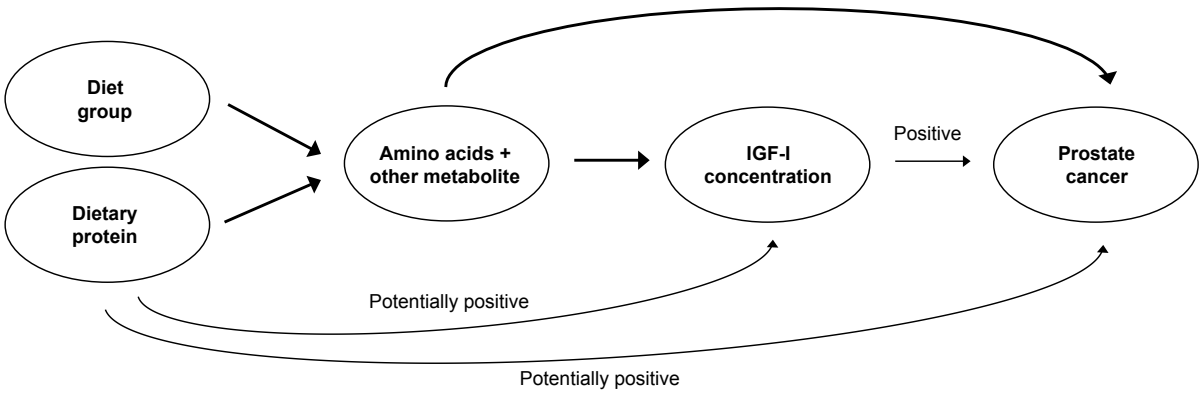


Figure 1.1: Research questions
Thick arrows represent associations which will be investigated in this thesis, while thin arrows represent previously studied associations which provided background for the thesis research questions.

1.2 Outline of the thesis

The first three chapters provide background information for the subsequent six empirical chapters. **Chapter 1** presents the rationale, research aims and an overview of the thesis. **Chapter 2** provides a review of prostate cancer epidemiology and explains what metabolomics is and why it is relevant for investigations of prostate cancer risk. **Chapter 3** describes the data sources for this thesis, i.e. EPIC, EPIC-Oxford and the metabolomics platform used.

Chapters 4 and **5** describe investigations into circulating concentrations of amino acids and metabolites, respectively, in meat-eaters, fish-eaters, vegetarians and vegans in cross-sectional analyses. **Chapters 6** and **7** provide the results of a nested case-control study of pre-diagnostic metabolite concentrations and risk of prostate cancer; in the former chapter, associations with the individual metabolites are examined, while in the latter chapter, patterns in metabolic profile are investigated using multivariate methods. **Chapters 8** and **9** presents results from cross-sectional analyses of amino acid concentrations in relation to dietary protein intake and to circulating IGF-I concentration, respectively.

In **Chapter 10**, the main findings and methodological considerations are summarised, and future work is outlined.

1.3 Role of the author in the preparation of this thesis

The idea of studying circulating concentrations of metabolites in relation to prostate cancer risk in EPIC was conceived by Ruth Travis and Timothy Key, who also secured the funding for the work from Cancer Research UK and the World Cancer Research Fund. After the conception of the main aim for the DPhil, I have been involved in all aspects of the research and have led

the literature search, data analysis and interpretation, and write up, while supervised by Ruth, Tim and Naomi Allen.

I conducted all searching and summarising of the literature, and I have written all the chapters. However, co-authors have read and provided advice on the manuscripts for publication on which three of the chapters in this thesis are based (Chapters 4 to 6); thus this work might not be considered fully my own.

All metabolomics assays were conducted by personnel at the laboratory at Nutrition and Metabolism Section, Biomarkers Group at the International Agency for Research on Cancer (IARC), Lyon, France. I prepared plasma samples from follow-up phase I for shipment to IARC and coordinated with the laboratory for all assays. Furthermore, I have visited the laboratory to discuss the assay with laboratory staff and to see how the assay is conducted. Assays for IGF-I were carried out at IARC and the Cancer Epidemiology Unit (CEU), University of Oxford before I started the DPhil [15, 16]. All other data, besides intake data on amino acids, were provided from the EPIC-wide and EPIC-Oxford databases by Paul Appleby, CEU statistician. I estimated the amino acid intakes for EPIC-Oxford participants as described in Chapter 4. Moreover, I was responsible for the cleaning of metabolite data and preparation of analytic datasets for Chapters 4 and 5, while datasets for the remaining empirical chapters (i.e. Chapters 6, 7, 8 and 9) were prepared by Georgina Fensom, CEU statistician.

I conducted all statistical analyses for Chapters 4, 8 and 9. For the remaining empirical chapters, I developed the analysis plan including all details of the analyses and Georgina provided technical assistance for efficient implementation in STATA and tabulating of results. Furthermore, the principal component partial R-square analysis presented in Chapter 5 was conducted by Marion Carayol, IARC postdoctoral fellow.

I designed all figures and tables. However, the code for the forest plots and Manhattan-style plots in Chapter 6 and the scatter plots in Chapters 8 and 9 was written by Georgina. The code to transform tabulated results into the appropriate format for the former two plot types was written by Jakob Blaavand, who also wrote the code for the EPIC map in Chapter 3. I produced the figures in all the other chapters and all the tables.

1.4 Publications from this thesis

Some of the work presented in this thesis (Chapters 4 to 6) has been published in peer reviewed journals as three papers of which I am the first author. The papers describe two cross-sectional analyses of circulating concentrations of amino acids [17] and metabolite concentrations [18], respectively, by diet group using data from EPIC-Oxford, and a nested case-control study on metabolite concentrations and risk of prostate cancer in EPIC [19].

The work described in Chapters 7 to 9 will be written up for publication when data from more matched case-control sets become available.

2. Prostate cancer epidemiology and metabolomics

2.1 Introduction

Prostate cancer is the second most commonly diagnosed cancer among men worldwide [1] but still few risk factors have been established. Among the established risk factors, only insulin-like growth factor-I (IGF-I) is potentially modifiable.

Metabolomics is the identifying and quantifying of small molecules in a biological system [12], and might prove to be a useful tool for identifying novel and potential modifiable risk factors for prostate cancer and for elucidating mechanisms underlying the associations between dietary elements, circulating IGF-I and risk of prostate cancer.

In this chapter, an overview of the epidemiology of prostate cancer will be given, along with an introduction to metabolomics.

2.2 Prostate cancer epidemiology

2.2.1 Geographical trends

Worldwide, 1.1 million new cases of prostate cancer were diagnosed in 2012 (the most recent update of Globocan), which makes prostate cancer the second most commonly diagnosed cancer in men [1]. The incidence rates vary more than 25-fold between regions. The highest rates are seen in high-income regions, such as Western and Northern Europe, North America and Australia/New Zealand (85 to 112 per 100,000), making prostate cancer the most commonly diagnosed male cancer in these regions. Rates are also high in some less developed regions including the Caribbean and Southern Africa (80 and 62 per 100,000, respectively). In contrast, the lowest incidence rates are found in Asia (9.4 per 100,000).

Compared to incidence, mortality rates differ less between regions, with a ten-fold difference between men in the Caribbean (29.3 per 100,000) and South-Central Asia (2.9 per 100,000) [1]. High mortality rates are also found in some African regions (ranging from 18.7 to 24.4 per 100,000 in Eastern and Southern Africa, respectively), while intermediate rates are seen in regions with the highest incidence (ranging from 9.8 to 14.5 in Northern America and Northern Europe, respectively). Overall there were 307,000 prostate cancer deaths worldwide in 2012 and prostate cancer ranked fifth among causes of cancer deaths in men.

The apparent discrepancy in geographical variation between incidence and mortality might partly be explained by differences in methods for detection (including testing using prostate-specific antigen (PSA)) and the availability of treatment [20].

2.2.2 Temporal trends

There has been an upward trend in prostate cancer incidence for decades in several countries [1, 21]. Furthermore, in the early to mid 1990s, shortly after the introduction of PSA-testing, a marked peak in incidence of prostate cancer was observed in countries such as Australia, USA and Canada, due to the rapid uptake of PSA-test in these countries [1, 21, 22]. In other countries a more steady increase in incidence has been observed.

For mortality, the trend over time has been different to that of incidence. In high-income countries, rates have mostly increased until the 1990s [1], followed by a decreasing trend [1, 21]. This might partly be explained by early detection (although the effect of PSA-testing on mortality is debated [23, 24]) and improved treatment [21]. In contrast, mortality is still increasing in less developed countries, including in parts of Asia and Africa [1, 21]. It has been suggested that the increase reflects a transition towards a more Westernised lifestyle in these countries but it might also partly be explained by improvements in data collection [21].

Due to the ageing and growth of the world's population, the burden of prostate cancer is expected to increase in the future [20]; 1.85 million diagnoses and 544,000 deaths from prostate cancer are estimated for 2030 [1].

2.2.3 Tumour heterogeneity, PSA-testing and implications for analyses

Prostate cancer develops very slowly and some tumours are indolent, while others are more aggressive and potentially lethal [25]. With PSA-testing, prostate cancer can be picked up three to 12 years before clinical diagnosis and its use has led to diagnosis of more low risk tumours [25, 26]. The choice of seeking and/or having a PSA-test as part of other medical investigations might be related to prostate cancer risk factors [27]. All of this has implications for prospective analyses aiming to identify risk factors for prostate cancer, including those presented in this thesis (Chapters 6 and 7).

Firstly, tumours present but not yet diagnosed at blood collection might affect metabolite concentrations and thus observed associations may at least partly be due to reverse causality. However, metabolites indicative of sub-clinical disease might be of interest for early detection [11]. To distinguish metabolites which might be markers of sub-clinical prostate cancer from metabolites which might be relevant for prostate cancer aetiology, risk analyses in the present thesis are also presented stratified by time from blood collection to diagnosis.

Secondly, the marked difference in the impact (clinical and personal) of indolent and potential deadly prostate cancer highlights the need to emphasise investigations of risk factors for more aggressive tumour sub-types. To accommodate this, analyses of prostate cancer risk are

conducted for sub-types of tumours defined according to tumour stage and histological grade of prostate cancer, and risk of prostate cancer death is also investigated.

Thirdly, PSA-testing could introduce detection bias in epidemiological studies as PSA-testing is related to risk factors for prostate cancer [27], and indeed, metabolites associated with screen-detected and clinically-detected prostate cancer may differ [28, 29]. In EPIC, data on PSA-tests have not been routinely collected. However, PSA-testing has not been as common in Europe as in US but might be increasing [20]. Recent data from the UK suggest that the annual (2007) uptake of PSA-testing was 6% in the general male population [30], and that 20% (EPIC-Oxford; mid 2000s) to 29% (UK Biobank; late 2000s) of male cohort participants had ever had a PSA-test [27, 31].

2.2.4 Established risk factors

Age [1], ethnicity [32, 33], family history [34], genetic variants [35] and circulating concentrations of IGF-I [2] are the only established risk factors for prostate cancer.

2.2.4.1 Age

Age is the major risk factor for prostate cancer. The reported incidence increases greatly with age; for men aged 45 to 49 years the world-wide incidence rate is 6.6 per 100,000 and it rises to 27.6, 67.7, 138.3, 232.9, 308.5 and 392.6 per 100,000 for each consecutive 5-year age band (Figure 2.1) [1]. Furthermore, based on autopsies, and thus including undiagnosed cases, it

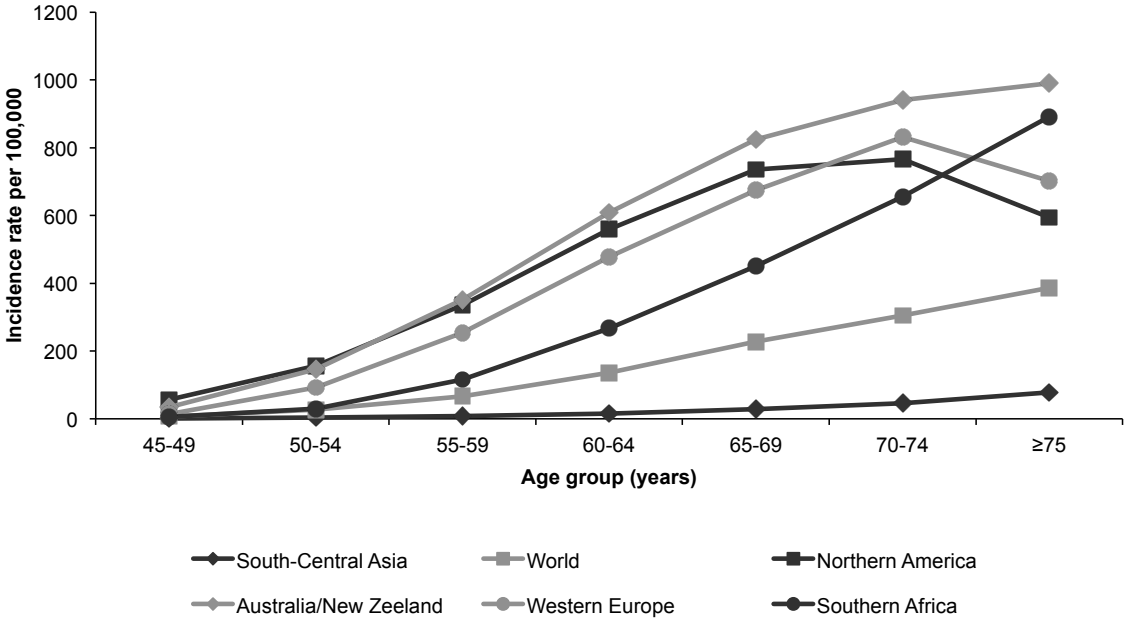


Figure 2.1: Incidence rates of prostate cancer by age-group and geographical region
Data are from GLOBOCAN 2012 [1] and rates are age-standardised.

is estimated that 8%, 33% and 91% of men of European descent in their 30s, 60s and 90s, respectively, have prostate cancer [32].

2.2.4.2 Ethnicity

As shown in Figure 2.1 on the preceding page, the incidence of prostate cancer varies markedly across the globe, and differences in prostate cancer incidence between men of different ethnic origins are also present within countries; in the US for example, the incidence rates in African Americans and Asian/Pacific Islanders are approximately 160% and 60%, respectively, of that in white men [22]. Such discrepancies within and between countries might be explained by differences in access to health care, uptake of PSA-testing, completeness of cancer registrations and genetic susceptibility to prostate cancer among others [36, 32], but autopsy studies confirm that prostate cancer is more common in men of African descent, followed by men of European descent and lowest men of Asian descent [32].

2.2.4.3 Family history

Prostate cancer is among the cancers with the largest heritability [37], which may be explained by common genetic, lifestyle and environmental factors.

A meta-analysis including 26 studies found a 2.5-fold risk of prostate cancer in men with a first degree relative with the disease compared to men without (RR = 2.48; 95%CI: $2.25 - 2.74$), while having two or more first degree relatives with prostate cancer was related to a 4.4-fold risk (RR = 4.39; 95%CI: $2.61 - 7.39$) [34]. Family history was more strongly related to early onset (before age 65 years; RR = 2.87; 95%CI: $2.21 - 3.74$) than later onset disease (RR = 1.92; 95%CI: $1.49 - 2.47$; $p_{het} = 0.002$).

2.2.4.4 Genetic factors

To date, more than 100 genetic variants (a few rare and many common variants) associated with prostate cancer risk have been identified [35, 37, 38]. Although each of the common variants is associated with a small increase in risk, when combining information from 100 genetic variants, the risk was 4.7 times higher in men in the top 1% of risk score compare to the average risk score (OR = 5.65; 95%CI: $4.83 - 6.62$) [35]. The identified variants explain about a third of the heritability in populations of European ancestry.

Some genetic variants have been related to aggressive prostate cancer but not exclusively; thus so far, genome-wide association studies have not helped to distinguish between susceptibility to aggressive and to non-aggressive disease [35, 39, 40].

2.2.4.5 Insulin-like growth factor-I

The Endogenous Hormones, Nutritional Biomarkers and Prostate Cancer Collaborative Group (EHNBPCCG) has published a pooled analysis of individual level data from 17 studies, covering most of the world's data on circulating concentrations of IGFs, including IGF-I, and prostate cancer risk [2]. When analysing IGF-I data from the 15 prospective studies including 7,682 cases, men in the top study-specific fifth compared to the bottom fifth had a 29% higher risk of prostate cancer (OR = 1.29; 95%CI: $1.16 - 1.43$); this association remained after adjusting for the other IGFs and the study results suggest that the association is linear. There was no evidence of heterogeneity in the association with prostate cancer by time to diagnosis or tumour sub-types. Furthermore, studies of genetic variants have also linked the IGF-pathway to prostate cancer risk [39, 41] and mortality [41, 42].

Moreover, the observational results are in line with experimental evidence showing that IGF-I has mitogenic and anti-apoptotic properties in both normal and cancer cells, including prostate cancer cells [43, 44]. IGF-I might be involved in several stages of cancer from the transformation of normal cells into malignant cells to metastatic invasion [43].

Finally, a thorough review and meta-analysis commissioned by World Cancer Research Fund (WCRF) UK of the association between milk intake, IGFs and prostate cancer risk including observational and experimental studies in both humans and animals found that there is moderate evidence for an association between IGF-I and risk of overall and aggressive prostate cancer [45].

2.2.5 Potential risk factors

In 2014, the WCRF published a systematic literature review of prospective observational studies and randomised controlled trials on prostate cancer in relation to foods, nutrients and physical activity as part of the Continuous Update Project (CUP) [46]. Based on the evidence emerging from this work and evidence on potential mechanisms, WCRF and the American Institute for Cancer Research concluded that positive associations of prostate cancer with body fatness and adult height are probable, while there is limited suggestive evidence of positive associations with intake of dairy products and calcium, and for inverse associations with circulating concentrations of selenium and α-tocopherol [9]. These potential risk factors for prostate cancer will be described below along with folate and lycopene, for which the evidence was too limited for conclusions to be drawn in the CUP but more evidence is now available. Diabetes and circulating concentrations of sex hormones, which might also be related with risk of prostate cancer but were beyond the scope of the CUP, will also be described.

Furthermore, the WCRF review concluded that the evidence was too limited to reach a conclusion for a wide range of other dietary factors, including specific sub-groups of: fruits,

vegetables, vitamins, minerals, macro-nutrients and drinks, and physical activity and vegetarian diets.

2.2.5.1 Adult attained height

The systematic literature review from WCRF found an increased risk of overall prostate cancer with height (OR = 1.04, 95%CI: $1.03 - 1.05$ per 5 cm in 34 studies); results for prostate cancer mortality and advanced/high grade prostate cancer were similar to the overall results [46]. Positive associations with high grade prostate cancer and death from prostate cancer have also been reported in a subsequent large (total $n_{cases} = 7,010$) cohort study within EPIC [47]. In contrast, two Mendelian randomisation studies of genetically predicted height observed no association with overall prostate cancer risk [48, 49] and an inverse association with advanced prostate cancer [48]. These results might thus not support a causal association between genetically determined height and prostate cancer although environmental determinants of adult height, e.g. early life factors, might explain the observed association of height and prostate cancer [48].

2.2.5.2 Body fatness

A recently published analysis in EPIC found a positive association of adiposity (BMI and waist circumference) with high grade and fatal prostate cancer [47], and a meta-analysis of observational studies found weight gain in adulthood to be positively associated with high-risk prostate cancer and death from prostate cancers [50].

In line with these results, several different measures of body fatness, i.e. body mass index (BMI, 45 studies), weight (14 studies), waist circumference (9 studies) and waist to hip ratio (5 studies), were all associated with a higher risk of more aggressive prostate cancer tumour types and prostate cancer death in the WCRF review [46]. For example, the risk of advanced/high grade and fatal prostate cancer increased by 8% and 11%, respectively, per 5 kg/m^2 increase in BMI ($RR_{\text{advanced/high grade}} = 1.08$; 95%CI: $1.04 - 0.12$; 23 studies and $RR_{\text{fatal}} = 1.11$; 95%CI: $1.06 - 0.17$; 12 studies).

It is not clear if the association between adiposity and more aggressive prostate cancer types is causal. Two Mendelian randomisation studies have not found significant associations between genetic score for BMI and risk of prostate cancer overall, by tumour sub-types [48, 51]. Moreover, the association between adiposity and prostate cancer risk in observational studies might partly be explained by detection bias [52]. In obese men, PSA concentrations are lower, digital rectal examination is more difficult and the likelihood of missing a tumour is greater due to the larger size of the prostate. This may lead to lower detection rates of early stage disease and higher rates of advanced prostate cancer.

2.2.5.3 Diabetes

While individuals with diabetes might be at higher risk of several cancers, including breast, colorectal, endometrial, and intrahepatic bile duct cancer, [53], an inverse association between diabetes and risk of prostate cancer might exist. A recent meta-analyses of 56 observational studies has found that men with diabetes have a modestly decreased risk of prostate cancer (RR = 0.88; 95%CI: $0.82 - 0.93$) [54]. Results were not materially changed when restricting the analysis to prospective studies. The inverse association was observed for both localised and advanced disease, as well as for low grade tumours (Gleason score < 7) but not for high grade disease (Gleason score ≥ 7). Likewise, subsequent results from EPIC showed a reduced risk of prostate cancer in men with diabetes (HR = 0.74; 95%CI: $0.63 - 0.86$) [55].

2.2.5.4 Circulating concentration of α-tocopherol

α-tocopherol is a form of vitamin E that has anti-oxidant functions [56]. Rich sources of α-tocopherol are nuts, seeds, vegetable oils and dietary supplements.

The WCRF review found small borderline inverse associations for circulating concentrations of α-tocopherol and risk of both overall (RR per 1 mg/L = 0.99; 95%CI: $0.98 - 1.00$; 9 studies) and advanced/aggressive prostate cancer (RR per 1 mg/L = 0.98; 95%CI: $0.97 - 1.00$; 4 studies) [46]. Subsequently, EHNBPCCG has published results from a pooled analysis of individual level data from 15 studies ($n_{cases} = 7,642$). In line with the systematic review, they showed a lower risk of overall and aggressive prostate cancer with higher concentrations of α-tocopherol; the ORs for the top versus bottom study-specific fifth were 0.86 (95%CI: $0.78 - 0.94$) for prostate cancer overall and 0.74 (95%CI: $0.59 - 0.92$) for aggressive prostate cancer [56].

There have also been several trials of vitamin E supplementation. Although the Alpha Tocopherol, Beta Carotene Prevention Trial (ATBC) study initially reported a protective effect of α-tocopherol on prostate cancer [57], this effect was not present at longer follow-up [58]. Similarly, no effect was reported in the Physicians' Health Study II [59], while a harmful effect was observed in the Selenium and Vitamin E Cancer Prevention Trial (SELECT) [60].

In summary, there might be an inverse association between circulating concentrations of α-tocopherol and prostate cancer risk including more aggressive tumour sub-types, but based on the currently available evidence from trials taking vitamin E supplements to reduce prostate cancer risk is not recommended.

2.2.5.5 Selenium concentration

The review by WCRF found small borderline significant decreased risks of overall prostate cancer (RR = 0.95; 95%CI: $0.91 - 1.00$; 9 studies) and advanced/high grade prostate cancer

(RR = 0.95; 95%CI: 0.89 − 1.00; 5 studies) in men with high serum or plasma selenium, while no prospective study of toenail selenium was identified [46]. Since then, a further meta-analysis and a pooled analysis have been published. Men with the highest serum selenium concentrations had a lower risk of overall, high grade and advanced prostate cancer in a meta-analysis of 17 studies, but the associations did not remain when restricting the analysis to the four cohort studies (three of which were prospective) [61]. A pooled analysis from EHNBPCCG of data from 15 prospective studies found strong inverse associations of selenium measured in blood and toenails with aggressive prostate cancer ($OR_{80\text{ percentile increase}} = 0.43$; 95%CI: $0.21 - 0.89$ and 0.18; $0.11 - 0.31$, respectively) [62]. For toenail selenium, inverse associations were also seen with overall and non-aggressive disease, while blood selenium was not associated with these outcomes. This may suggest that toenail concentration is a better marker of long term selenium levels.

The SELECT trial and two smaller trials in men at high risk of prostate cancer have reported no effect of selenium supplementation on prostate cancer risk [60, 63, 64]. However, the effect of supplementation might depend on the baseline level of selenium; a beneficial effect has been found, especially in men with low selenium status [65], while an increased risk of high grade prostate cancer has been reported in men with higher baseline concentrations [66].

In summary, a lower concentration of selenium might be associated with a higher risk of prostate cancer. More prospective studies using toenail concentration may advance our understanding of the association between selenium and prostate cancer risk.

2.2.5.6 Folate

Folate (vitamin B9) is involved in DNA methylation, repair and synthesis [67]. A recent EHNBPCCG analysis, including data from six prospective studies ($n_{cases} = 6,875$), found that men in the top study-specific fifth of circulating folate concentration had an increased risk of prostate cancer ($OR_{\text{top vs bottom fifth}} = 1.13$; 95%CI: $1.02 - 1.26$; $p_{trend} = 0.02$) and in particular of high grade prostate cancer for which men in the top fifth had a 2.3-fold risk compared to men in the bottom fifth ($OR_{\text{top vs bottom fifth}} = 2.30$; 95%CI: $1.28 - 4.12$; $p_{trend} = 0.001$) [67].

Two meta-analyses, including five prospective studies and seven mainly prospective studies, respectively, have also found increases in overall prostate cancer risk with higher blood folate concentrations ($OR_{\text{high vs low}} = 1.21$; 95%CI: $1.05 - 1.39$ and 1.43; $1.06 - 1.93$, respectively) [68, 69]. In contrast, no association with intake of folate was found.

Prior to these publications, the review from WCRF did not find an association between serum folate and prostate cancer risk, when pooling data from seven studies on overall prostate cancer risk [46]. Neither the two meta-analyses nor the review included results for tumour sub-types.

2.2.5.7 Lycopene

Lycopene is a carotenoid mainly found in tomatoes [56]. Data from 15 studies ($n_{cases} = 6,107$) in the EHNBPCCG showed no association with risk overall ($OR_{80\text{ percentile increase}} = 0.97$; 95%CI: $0.87 - 1.08$) but there was evidence of heterogeneity by tumour stage. The risk reduction was approximately 27% for both advanced and aggressive disease ($OR_{80\text{ percentile increase}} = 0.73$; 95%CI: $0.54 - 0.99$ and 0.72; $0.53 - 0.97$, respectively), while no associations were seen with localised and non-aggressive prostate cancer ($p_{het} = 0.04$ and 0.03 for localised versus advanced and non-aggressive versus aggressive prostate cancer, respectively).

Two meta-analyses, including prospective and retrospective studies, were also published in 2015 and both reported inverse associations of lycopene intake and blood concentrations with overall prostate cancer risk [70, 71].

In contrast, the evidence gathered by the WCRF on both serum (10 studies) and dietary (7 studies) lycopene did not support an association between lycopene and risk overall (OR per 10 μg/dL = 0.99; 95%CI: $0.96 - 1.01$, and per 5 mg/d = 0.98; $0.93 - 1.02$, respectively) or by prostate cancer sub-types [46]. Furthermore, a recent, relatively small ($n_{cases} = 217$ for lycopene) study from the UK Dietary Cohort Consortium pooled data from four prospective studies and did not find an association between dietary lycopene and risk of prostate cancer overall or sub-types [72].

2.2.5.8 Sex steroid hormones

Male sex steroid hormones or androgens (e.g. testosterone) are involved in the normal development and growth of the prostate gland [73]. Such hormones have long been hypothesised to be related to risk of prostate cancer partly because androgen-deprivation therapy has an effect on most early prostate cancers. A decade ago, no association was found between sex steroid hormones, i.e. testosterone, calculated free testosterone, dihydrotestosterone, dehydroepiandrosterone sulfate, androstenedione, androstanediol glucuronide, estradiol and calculated free estradiol, and risk of prostate cancer in a pooled analysis of 18 prospective studies ($n_{cases} = 3,886$) from the EHNBPCCG (called the Endogenous Hormones and Prostate Cancer Collaborative Group then) [73]. This work is currently being updated and results from 20 prospective studies ($n_{cases} = 6,933$) indicate that men in the bottom decile of free testosterone concentration have a decreased risk of prostate cancer overall (OR = 0.79; 95%CI: $0.70 - 0.88$), while men with higher concentrations all have similar risks (personal communication Eleanor Watts, Cancer Epidemiology Unit, University of Oxford). Men with the lowest free testosterone concentrations might also be at lower risk of low grade disease (OR = 0.76; 95%CI: $0.67 - 0.88$) but at higher risk of high grade prostate cancer (OR = 1.65; 95%CI: $1.04 - 2.63$; $p_{het} = 0.002$).

2.2.5.9 Intake of a high calcium diet

The results on calcium intake from the WCRF literature review have also been published in a journal paper, along with results on intake of dairy products [74]. For calcium intake, the findings indicated a small increase in risk of total prostate cancer with higher total intakes of calcium (RR per 400 mg/d = 1.02; 95%CI: $1.01 - 1.04$; 9 studies) and an increased risk of fatal prostate cancer (RR per 400 mg/d = 1.11; 95%CI: $1.02 - 1.21$; 2 studies) [46, 74]. No association was observed with prostate cancer sub-types, including non-advanced and advanced prostate cancer. The results suggested that the associations might be confined to calcium from dairy products – the major source of dietary calcium. The review reported a modest increased risk of overall prostate cancer with higher intakes of calcium from dairy products in many populations (RR per 400 mg/d = 1.06; 95%CI: $1.02 - 1.09$; 6 studies) [46, 74]. Similar but non-significant risk estimates were observed for non-advanced, advanced and fatal prostate cancer. In contrast, no association was seen with non-dairy calcium or calcium supplements.

Recently, the pooled analysis from the UK Dietary Cohort Consortium reported no association between calcium intake and risk of prostate cancer ($n_{cases} = 1,717$ for calcium intake) [72].

In summary, these results suggest that men who have a diet high in calcium might have a higher risk of prostate cancer. However, the association might primarily be with calcium from dairy products but not non-dairy calcium, thus compounds in dairy other than calcium might at least in part explain the association between calcium and risk [74].

2.2.5.10 Intake of dairy products

The review from WCRF found modest positive associations of intake of total dairy products (RR per 400 g/d = 1.07; 95%CI: $1.02 - 1.12$; 15 studies), total milk, low-fat milk and cheese with overall prostate cancer risk [46, 74]. However, no associations were observed with whole milk, yogurt and other dairy products including ice cream and butter. There was no heterogeneity in the association between dairy product intake and prostate cancer risk by tumour characteristics or death from prostate cancer (fatal versus non-advanced prostate cancer). However, a positive association between low-fat milk and non-advanced disease was reported (RR per 200 g/d = 1.09; 95%CI: $1.01 - 1.17$; 4 studies), while no association was seen with advanced disease (RR per 200 g/d = 1.02; 95%CI: $0.95 - 1.08$; 5 studies) [46, 74]. For other categories of dairy products, no associations were observed with tumour sub-types. It has been suggested that fat or calcium might explain the association between intake of dairy products and prostate cancer risk but these hypotheses are not supported by the evidence collected by the WCRF, because results varied by source of calcium and fat content of the dairy products [74].

The recent UK pooled analysis reported no association between intake of milk, cheese or yogurt, with no heterogeneity by stage of disease or detection method (PSA versus clinically) [72].

In summary, men who consume more dairy products might have a higher risk of prostate cancer, which might be explained by components of dairy other than calcium and fat.

2.2.6 Hypothesis on dairy protein, amino acids, IGF-I and prostate cancer risk

A compound of dairy products that might partly explain the association between intake of such foods and prostate cancer risk is protein. A study in EPIC found a 32% (95%CI: $1 - 72\%$) increase in overall risk and a 76% (95%CI: $6 - 195\%$) increase in risk of high grade disease per 35 g of protein derived from dairy products [8], but null findings have been reported in other studies [72, 75]. This potential association between dairy protein intake and prostate cancer risk might partly be explained by effects of dairy protein on IGF-I concentrations and the amino acid composition of the protein.

Firstly, there is considerable evidence that higher intake of dairy protein is related to higher circulating IGF-I concentration in humans [6, 7, 75, 76]. It is not clear if this is due to the specific amino acid composition of the protein, because limited literature is available on the association between intake of dairy protein, amino acids and IGF-I concentrations, especially in men. Nonetheless, total protein intake has been reported to be positively correlated with circulating concentrations of some amino acids including branched-chain amino acids [77], and lower concentrations of the non-essential amino acid arginine have been reported to be associated with higher concentrations of IGF-I in women [78]. Also in women, intake of protein low in essential amino acids (non-soya plant protein) was inversely associated with IGF-I concentration, while a positive association was observed for protein rich in essential amino acids (protein from animals and soya) [79]. Furthermore, vegan men and women, who do not consume animal products (including dairy) and thus have a diet lower in essential amino acids, [79, 80] have been observed to have lower IGF-I concentrations than meat-eaters and vegetarians [79, 81].

Secondly, results from the few prospective studies of prostate cancer risk and metabolite concentrations published to date [19, 28, 29, 82, 83, 84] have suggested that the amino acids alanine [82, 84], arginine [28, 29], lysine [82], methionine [19, 82], phenylalanine [82], proline [19, 84], trans-4-hydroxyproline [19] and tryptophan [29], and metabolites involved in metabolism of branched-chain amino acids (i.e. leucine, isoleucine and valine) [28] might be related to risk of prostate cancer. Post-diagnostic amino acid concentrations have also been reported to differ between prostate cancer cases and controls [11, 85].

Finally, the hypothesis that intake of dairy protein may be positively related to prostate

cancer risk, via the amino acid composition of the protein and IGF-I concentration, is also supported by experimental studies. Evidence in rodents suggests that the amino acid composition of dietary protein may be related to prostate tumour growth [4], potentially via influence on IGF-I concentration and activation of mTOR1 [4, 86] (Figure 2.2). Intake of essential amino acids might increase plasma IGF-I concentration in humans and rodents [86, 87, 88]. IGF-I stimulates the PI3K-Akt-mTOR pathway [3], which leads to increased proliferation and decreased autophagy. This pathway, which is upregulated in 30 to 50% of prostate cancers, may play a key role in prostate cancer initiation and progression [4, 89]. Moreover, the branched-chain amino acids can upregulate mTOR activity [4, 89, 5] by activating mTOR [89].

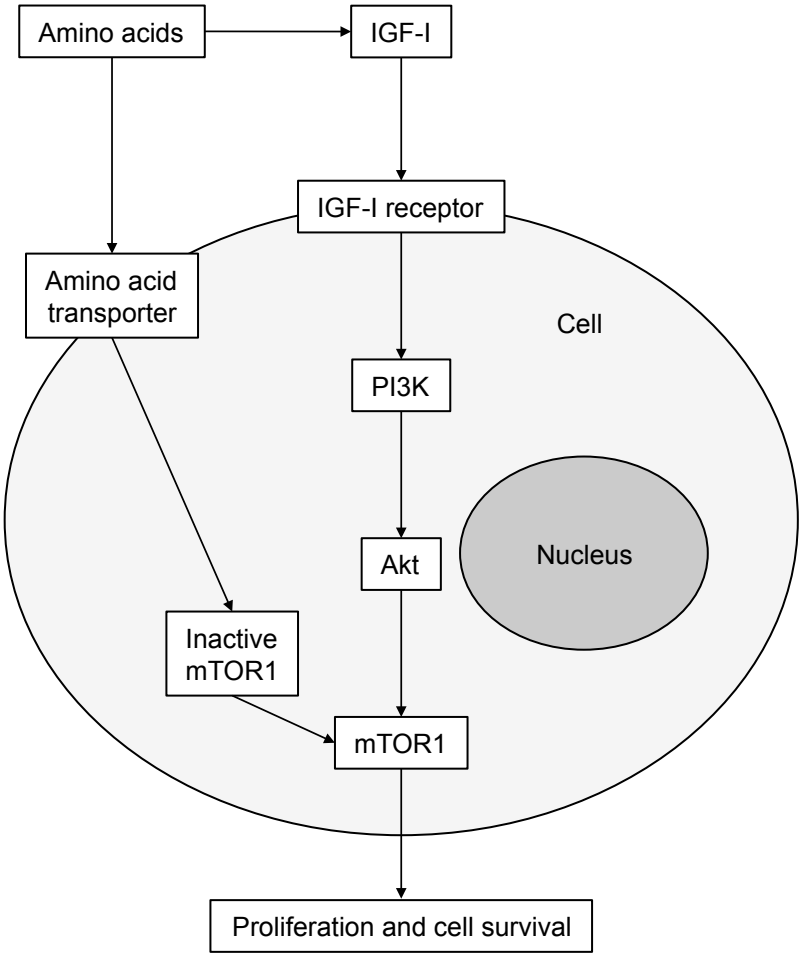


Figure 2.2: Stimulation of proliferation and cell survival by amino acids and IGF-I via the PI3K/Akt/mTOR pathway

Figure adapted from [4, 89]. Essential amino acids may increase IGF-I synthesis and IGF-I stimulates mTOR1 via the PI3K-Akt-mTOR pathway, leading to proliferation and cell survival. Branched-chain amino acids (i.e. isoleucine, leucine and valine) may stimulate mTOR1 activity by transforming the inactive form of mTOR1 to the active form.

In summary, a high intake of dairy protein might be related to an increased risk of prostate cancer due to elevated IGF-I concentration. While experimental animal research suggests that specific amino acids may be involved in this relationship, direct evidence in humans, especially men, is limited. Studies investigating the associations between intake of dairy protein and amino acid concentrations and concentrations of amino acids and IGF-I in men, as well as larger studies

investigating the association between amino acid concentrations and risk of prostate cancer risk are warranted.

2.3 Metabolomics

Metabolomics is one of the '-omics' technologies, like genomics, epigenomics or proteomics, and it focuses on identification and quantification of metabolites in a biological system such as a single cell or the human body [13]. Metabolites are low-molecular weight (up to 1000 to 1500 Dalton [13, 29]) reactants, intermediates or products of biochemical reactions and the sum of these constitutes the metabolome.

The term 'metabolomics' was first used by Fiehn and colleagues, working in the field of plant biology, in the early 2000s [90]. In 2002, Fiehn wrote:

> *"[...] a comprehensive analysis in which all the metabolites of a biological system are identified and quantified is needed. Since such an approach reveals the metabolome of the biological system under study, this approach should be called metabolomics."* [12].

A few years prior to this, in 1999, Nicholson and colleagues, working on drug toxicology, had defined the similar term, 'metabonomics', as

> *"the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification."* [91].

The two terms are now used as synonyms [13] but 'metabolomics' is the more commonly used term.

The metabolome can be viewed as a snapshot of the biological state of the entire system, physiological as well as pathological, and is affected by internal factors, including the genome, epigenome, transcriptome and proteome, as well as by external factors, such as diet, lifestyle, gut microbiota and environmental factors [11, 13, 14]. The human metabolome is estimated to comprise tens of thousands of metabolites and to date 74,507 have been listed in the Human Metabolome Database [92, 93].

2.3.1 Analytical techniques

No analytical approach can measure the entire human metabolome, but when applying an untargeted approach, the aim is to cover as many metabolites as possible with the chosen technique (hundreds to thousands) [94, 13]. In untargeted metabolomics both known and unknown metabolites are measured, often in a non-quantitative manner [13]. The main limitation of this approach is the difficulty in identifying unknown metabolites [11, 13].

In contrast, targeted approaches measure a limited, fixed number of known metabolites (tens to hundreds) and are often quantitative or semi-quantitative such that concentrations of the metabolites can be determined [13, 94]. Compared to the untargeted approaches, the targeted approaches are more accurate and precise but more labour intensive and may be biased towards metabolites for which internal standards to facilitate quantification are available [13].

For both targeted and untargeted approaches, several analytical platforms can be applied, with mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) being the most commonly used methods. Both methods are well established and have been used for decades [13]. Both these techniques can analyse various types of samples, including urine, plasma, serum and tissue.

In MS, metabolites are ionised and the measurement of metabolites is based on the mass to charge ratio of the ions [11, 13]. MS has high sensitivity (i.e. low limits of detection), resolution and accuracy, and the method is fast [13] and requires smaller sample volumes than NMR [11]. On the other hand, the identification of metabolites is complex and quantification requires internal standards. Moreover, the samples come in direct contact with the detector leading to potential contamination, and thus quality control samples are required.

The samples (after any necessary preparatory steps) can be injected directly into the mass spectrometer or as commonly done injected after using a separation method, e.g. gas chromatography (GC) and liquid chromatography (LC), [13, 95]. GC has a very high resolution and sensitivity but requires much pre-processing and run times are long; thus through-put using GC is not high. In contrast, traditional LC does not have a high resolution, but developments to this method, namely high and especially ultra-high performance liquid chromatography (HPLC and UPLC, respectively), have facilitated much better separation, sensitivity and lower run times, making these methods applicable for large-scale studies.

In NMR samples are placed in a magnetic field and metabolites are detected based on the emission of radio waves specific to each metabolite in response to changes in the magnetic field [96, 97]. The main strengths of NMR are that results are highly reproducible, it provides quantitative and structural information which aids the identification of metabolites, requires little sample preparation and preserves the samples [13]. In contrast, the main limitation is the lower sensitivity compared with that of MS [13].

2.3.2 Metabolomics in epidemiology

Due to recent advances in the technologies described above, it has become feasible in terms of time and cost to use metabolomics in large epidemiological studies [13].

The aims of using metabolomics in epidemiology are to identify novel risk factors for disease and biomarkers of disease and exposure, and to advance our understanding of disease mechanisms

[10]. Metabolomics may thus help us understand the mechanisms underlying the association between dairy protein, IGF-I and prostate cancer risk and to identify novel risk factors for prostate cancer [10, 11].

2.3.3 Metabolites and prostate cancer risk

Metabolomics is being used in the search for biomarkers to improve detection of prostate cancer and the prediction of treatment response and relapse. From such studies, which are human retrospective case-control studies or cell based studies, several metabolites have been reported to be potential markers of prostate cancer, including amino acids, choline containing compounds, acylcarnitines, lipids and metabolites related to energy production [11, 85]. Although the validation and replication of these results is still limited, promising results have been found; higher sensitivity and specificity of metabolite profiles than those for PSA have been reported for prostate cancer diagnosis and recurrence [11].

Less is known about how metabolite concentrations prior to diagnosis might be related to risk of subsequent development of prostate cancer. Two studies in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) have found strong inverse associations between energy (involved in the Krebs cycle) and lipid metabolites and risk of aggressive prostate cancer [28, 82], which might differ by tumour stage [83]. In line with the ATBC analyses, a study in EPIC-Heidelberg suggests that alterations in phosphatidylcholine concentrations (a type of phospholipids) might be linked to risk of prostate cancer [84]. In contrast to these findings, results from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) suggest that associations with lipid and energy metabolites might be in the opposite direction in a screened population possibly due to differences in case-mix [29].

These prospective studies were all of a limited size (n_{cases} = 74 to 380), few of the associations with individual metabolites remained after correction for multiple testing, only two used multivariate analytical approaches to deal with the high dimensionality of metabolomics data, and none investigated the risk of death from prostate cancer. Studies with larger sample sizes, applying multivariate statistical methods, are needed to provide more robust risk estimates and to understand associations between metabolite profile and prostate cancer development.

2.4 Conclusion

In conclusion, prostate cancer is a common disease in men, but relatively little is known about risk factors for this disease. Men who are older, are of African ancestry, have a family history of prostate cancer, have certain genetic variants or have higher circulating concentrations of IGF-I have a higher risk of prostate cancer. It is probable that taller men and men with higher body

fatness are also at higher risk of prostate cancer (more aggressive disease sub-types only for body fatness), while men with diabetes might have a lower risk of prostate cancer. There is limited suggestive evidence that men with higher concentrations of α-tocopherol and selenium have a lower risk of prostate cancer, and that men with diets rich in calcium and dairy products have a higher risk. The association with intake of dairy products might partly be explained by the protein content, as dairy protein intake may be linked to higher IGF-I concentration. It is not clear, however, if this is due to the amino acid composition of dairy protein. Finally, recently published and ongoing work suggests that folate and lycopene might be associated positively and inversely, respectively, with prostate cancer overall and/or more aggressive sub-types, while men with very low testosterone concentration might be at lower risk of overall and low grade prostate cancer and at higher risk of high grade disease.

Metabolomics is the identification and quantification of metabolites in a biological system and could help identify novel risk factors for prostate cancer and elucidate mechanisms underlying the observed associations between dietary protein, IGF-I and prostate cancer risk.

Little is known about pre-diagnostic metabolite concentrations and prostate cancer risk, and larger prospective studies are warranted. Furthermore, studies investigating circulating amino acid concentrations in relation to protein intake and to circulating IGF-I concentrations in men are lacking.

3. Data sources

All primary data used for this thesis are from the European Prospective Investigation into Cancer and Nutrition (EPIC). This large prospective multi-centre cohort was set up with the aim of investigating how diet, metabolism and lifestyle are related to risk of cancer in a population with large differences in dietary habits and cancer risks [98]. Oxford is one of the centres in EPIC and data from the EPIC-Oxford sub-cohort as well as data from other EPIC study centres, were used in this thesis.

3.1 Study population

The EPIC cohort comprises 520,000 men and women from 23 centres in 10 European countries (Denmark, France, Germany, Greece, Italy, Netherlands, Norway, Spain, Sweden and United Kingdom) [99]. Among the participants are 153,400 men from 19 centres in eight countries, as men were not recruited in France, Naples (Italy), Norway and Utrecht (Netherlands). Initially, seven core countries formed EPIC, with the three Scandinavian countries and the study centre in Naples, Italy joining later.

In EPIC-Oxford, 14,600 of the 65,000 participants are men [100].

The analyses in this thesis include data from EPIC-Oxford and from 14 other study centres in the core countries that recruited men (Germany, Greece, Italy, Netherlands, Spain and United Kingdom; Figure 3.1 on the following page).

3.1.1 Recruitment

Recruitment took place from 1992 to 2000 and participants were invited in person or by mail to take part [98]. The target population for each centre was chosen to ensure high rates of participation and long-term follow-up. In general, if follow-up was possible via population registries then members of the general population were invited, if not, specific groups, e.g. people with a health insurance plan, were recruited. The target age group was 35 to 74 years in most centres.

Unlike the other centres, EPIC-Oxford aimed to recruit large numbers of participants who did not consume meat, i.e. fish-eaters, vegetarians and vegans, to increase heterogeneity in the dietary exposure within the centre [100]. Individuals over the age of 20 were invited from across the UK between 1993 and 2001. Most of the 65,000 participants (89%) were recruited by postal recruitment via the Oxford Vegetarian Study and the Vegan Society, while the remaining participants were recruited via general practitioners' surgeries (mainly non-vegetarians).



Figure 3.1: Countries and study centres in EPIC

Countries participating in EPIC are marked on the map; dark and light grey indicate countries included and excluded in this thesis, respectively. Dots and the star (Oxford) indicate the location of the 15 study centres which contributed data for analyses in this thesis. The map was drawn using the Python programming language (Python Software Foundation, available at https://www.python.org/)

3.1.2 Exclusion criteria

In EPIC-wide analyses, men with prior cancer (except non-melanoma skin cancer), missing information on the date of cancer or end of follow-up, missing dietary or other lifestyle data, or an improbable total energy intake (in the top or bottom 1% of the estimated ratio of energy intake to energy requirement), were excluded. This left 142,000 men in the cohort.

In EPIC-Oxford, men were excluded if they had prevalent malignant cancer (registered or self-reported) at recruitment and were less than 20 or 90 years old or more. Further exclusion criteria specific to each analysis were applied and are described in the relevant chapters.

3.2 Data collection

Data on dietary intake, lifestyle, history of disease, anthropometry and biological samples were collected from study participants, who were asked to fill in questionnaires and/or to visit their study centre for further data collection.

3.2.1 Blood collection and processing

To facilitate measurements of nutritional status, metabolism, genetics and dietary biomarkers. blood samples were collected; 139,600 ($\approx 90\%$) of the male cohort members gave a blood sample.

Blood collection and processing was conducted according to a standardised protocol in the core centres. Identical tubes and other material for blood collection and storage was used in each centre. A total of 30 mL of blood was drawn from study participants. Whole blood was transported to a local laboratory protected from light at 5 to 10°C [15] and prepared for long-term storage within 24 hours. Processing included separating blood fractions, i.e. plasma (using citrate as an anticoagulant), serum and white and red blood cells, and aliquoting these into 28 0.5 mL straws. Half of the straws for each blood fraction was stored locally and the other half was stored at the International Agency for Research on Cancer (IARC), Lyon, France in liquid nitrogen at −196°C.

A deviation from the protocol occurred in EPIC-Oxford, where the transport of whole blood to a laboratory in Norfolk was done at ambient temperature, with a mean transit time of 1.5 days. The blood straws kept by EPIC-Oxford were stored in liquid nitrogen (−196°C) until 2011 and subsequently in electric freezers (−80°C).

3.2.2 Dietary data

3.2.2.1 EPIC-wide

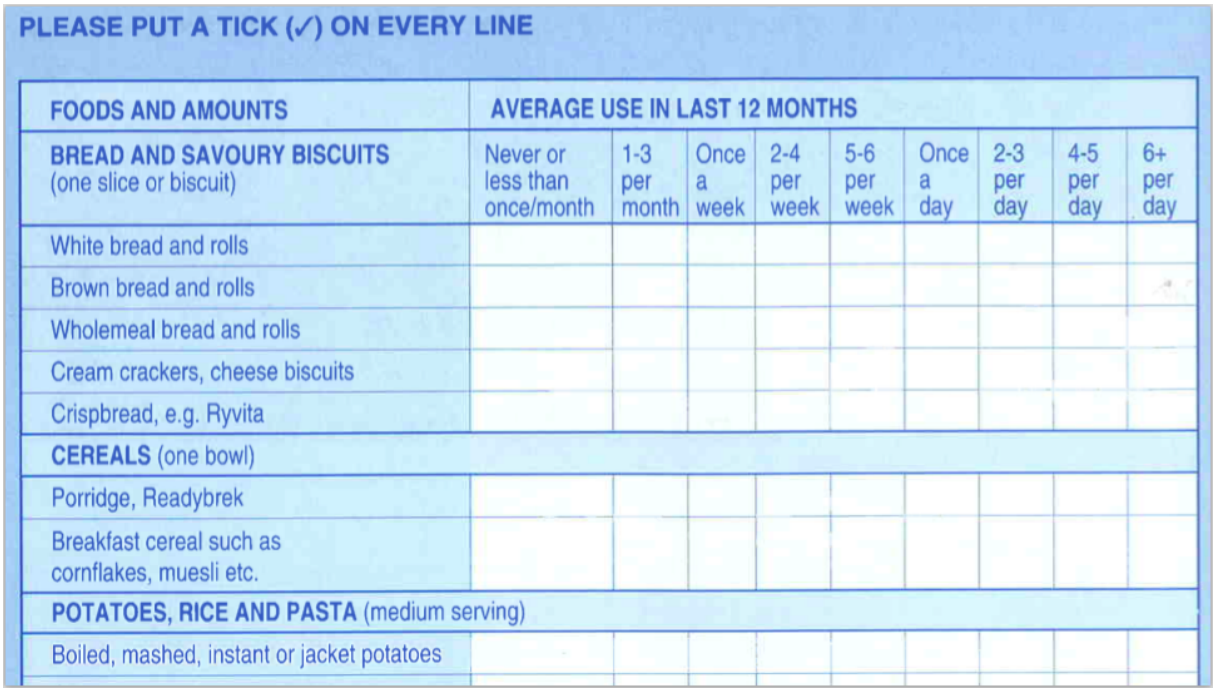
The collection of dietary data was designed to capture geographical differences in dietary intake and thus not standardised across countries. The usual dietary intake over the past 12 months was collected using a variety of methods, including quantitative dietary questionnaires, semi-quantitative food frequency questionnaires (FFQ) and 7-day dietary records [99]. In most centres, participants filled in the questionnaires or records themselves, while interviews were conducted in Spain. The relative validity of the measurement tool of choice was validated in each centre using multiple 24-hour dietary recalls or food records (UK), and in most centres also using biomarkers, including 24-hour urinary nitrogen for protein intake [101, 102, 103, 104, 105, 106, 107]. The relative validity was modest to good but varied by foods/nutrients and centre.

Nutrient intakes were estimated by multiplying the nutrient content of each food for a specific portion size by the frequency of consumption stated on the food questionnaire using country-specific nutrient databases.

3.2.2.2 EPIC-Oxford:

At recruitment, EPIC-Oxford participants filled in a validated semi-quantitative FFQ [107, 108] and were additionally asked: "Do you eat any meat?", "[...] fish?", "[...] dairy products?" and "[...] eggs?". Based on these questions, participants were categorised as meat-eaters (who answered 'yes' to meat, $n = 33,900$), fish-eaters (who did not eat meat but did eat fish, $n = 10,100$), vegetarians (who did not eat meat or fish, $n = 18,800$) and vegans (who answered 'no' to all four food groups, $n = 2,600$).

In the FFQ, participants were asked to report their average intake of 130 food items (113 of which were relevant for vegetarians and vegans) over the past 12 months in nine categories ranging from "Never or less than once/month" to "6+ per day" (Figure 3.2). The mean daily intake of food items was calculated by multiplying the frequency of consumption with a specified portion size (e.g. one slice of bread) mainly taken from Ministry of Agriculture, Fisheries and Food, Food portion sizes [109]. The mean daily nutrient intake was estimated by multiplying the daily intake by the nutrient content of the food item, using the 5th edition of 'McCance and Widdowson's The Composition of Foods' and its supplements [110, 111, 112, 113, 114, 115, 116, 117, 118, 119]. However, the amino acid contents of foods were taken from 'United States Department of Agriculture (USDA), National Nutrient Database for Standard Reference, release 26' [120], because limited data were available from 'McCance and Widdowson's The Composition of Foods'.



PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE IN LAST 12 MONTHS								
BREAD AND SAVOURY BISCUITS (one slice or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread and rolls									
Brown bread and rolls									
Wholemeal bread and rolls									
Cream crackers, cheese biscuits									
Crispbread, e.g. Ryvita									
CEREALS (one bowl)									
Porridge, Readybrek									
Breakfast cereal such as cornflakes, muesli etc.									
POTATOES, RICE AND PASTA (medium serving)									
Boiled, mashed, instant or jacket potatoes									

Figure 3.2: EPIC-Oxford baseline food frequency questionnaire
Excerpt of the semi-quantitative food frequency questionnaire used to assess dietary intake of 130 food and drink items at recruitment in EPIC-Oxford.

3.2.3 Other data

Participants filled in a non-dietary questionnaire about alcohol intake, physical activity, smoking, education, occupation, socioeconomic status and history of disease (own and family's) and treatments [99]. In the core countries, the same set of questions, although in different languages, were used.

The Cambridge Physical Activity Index was used to categorise participants as inactive, moderately inactive, moderately active and active based on their physical activity at work and their leisure activities, such as cycling, swimming, aerobics and jogging [121]. This index has been validated against energy expenditure (measured using 4-day heart rate monitoring) and cardio-respiratory fitness level (estimated using sub-maximum oxygen uptake tests), and was found to rank participants according to their physical activity level.

Anthropometric measurements, including height, weight and circumference of waist and hips, were measured at the study centre in the initial EPIC-centres, except in Oxford [99]. In EPIC-Oxford anthropometric data were self-reported. In a sub-sample of the cohort ($n = 4,808$) measurements were additionally taken and the correlation between measured and self-reported data was high (correlation coefficient (r) > 0.9) [122].

3.2.4 Data storage, management and quality control

Individual participant data, without information to identify participants, from each participating centre are stored centrally at IARC in the ORACLE database [99]. Centre-specific variables were standardised across all centres and rigorous quality control checks were performed before central EPIC data were stored.

Data on EPIC-Oxford participants are also stored locally at the Cancer Epidemiology Unit, University of Oxford.

3.3 Ethics

All participants gave written informed consent to participate in the EPIC study, and the EPIC study protocol was approved by the ethical committees of IARC, Lyon, France, and the participating centres.

The EPIC-wide project (described in Chapters 6 to 9) is approved by the Internal Review Board at the IARC (Project No. 14-09) and the local ethics committees in participating centres where required, i.e. Comité Ético de Investigación Clínica del Hospital Universitari de Bellvitge (Reference No. PR097/14).

The protocol for EPIC-Oxford was approved by a multi-centre research ethics committee (MREC/02/0/90) and the project described in Chapters 4 and 5 is covered by this approval.

3.4 Follow-up for cancer incidence and mortality

In most centres data on cancer diagnoses are obtained via linkage to population-wide cancer registries [99]. However, this is not possible in Germany and Greece, so a combination of active follow-up via participants or their next-of-kin and records from health insurance companies, and municipality-, hospital- and physician-based cancer and pathology registries is used. Similarly, mortality registries or active follow-up and death records are used to obtain data on deaths occurring in the cohort. Death from prostate cancer is defined as death with prostate cancer listed as the underlying cause of death on the death certificate.

3.4.1 Classification of prostate cancer and tumour sub-types

Prostate cancer was defined as code C61 in the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10). Cases were men diagnosed with prostate cancer after blood collection and prior to the end of follow-up.

The centralisation of follow-up information on prostate cancer and matching of cases with controls for nested case-control studies has been done in phases. The latest centralisation of follow-up (phase IV) was completed in 2015 and matching was finalised in 2016; hereafter, there were 6,778 prostate cancer cases and 6,778 matched controls in EPIC.

The histological grade (i.e. the degree of differentiation of the cancer cells) was defined using the Gleason score. Histological patterns of the tumour are scored from 1 to 5 and the sum of the two primary patterns give the Gleason score, ranging from 2 to 10, from the highest to the lowest degree of differentiation [123]. Tumours with a Gleason score < 8 or coded as well, moderately or poorly differentiated were classified as low-intermediate grade and tumours with Gleason score ≥ 8 or coded as undifferentiated were classified as high grade.

The cancer stage (i.e the extent to which the prostate cancer has spread) was primarily defined using the tumour-node-metastasis (TNM) staging system. The definition for localised cancer was $\leq$ T2 and N0/X and M0, or stage coded as localised, while advanced stage was defined as T3 $-$ 4 and/or N1 $-$ 3 and/or M1, or coded as advanced. The subset of advanced cancers scored as T4 and/or N1 $-$ 3 and/or M1 were classified as aggressive.

3.5 Metabolomics assay

All measurements of metabolite concentrations were performed at the laboratory at Nutrition and Metabolism Section, Biomarkers Group at IARC, using the Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbrcuk, Austria) [124]. This targeted assay, which covers all 20 standard amino acids, was chosen due to the *a priori* hypothesis around amino acids, IGF-I and prostate cancer risk (Section 2.2.6 on page 15), and the cost of the assay. In total up to 188

metabolites are covered and, besides amino acids, the assay measures acylcarnitines, biogenic amines, glycerophospholipids (all lysophosphatidylcholines and phosphatidylcholines), hexose (the sum of a range of monosaccharides with six carbon atoms, including glucose, fructose and galactose), and sphingolipids (all hydroxysphingomyelins and sphingomyelins).

3.5.1 The assay procedure

The metabolomics assay uses tandem mass spectrometry (MS/MS) coupled with liquid chromatography (LC) for amino acids and biogenic amines, and flow injection analysis (FIA) for other metabolites.

First, the assay is prepared by loading the 96-well plate with three zero-samples, seven calibrated standards, three internal quality control samples (low, medium and high concentration) and actual samples; 10 μL of sample is required [124]. To prepare for separation of individual analytes, derivatisation is performed for amino acids and biogenic amines, and extraction with organic solvent is carried out.

Second, the samples are processed. Amino acids and biogenic amines are assayed using LC-electrospray ionization (ESI)-MS/MS [125], (high performance LC was used in the assay for analysis of diet groups, while ultra-high performance LC was used in the assay for samples from prostate cancer cases and controls). All amino acids and biogenic amines are validated as fully quantifiable and standards in seven different concentrations are used to form the calibration curve from which the concentration of the sample is calculated using software provided by Biocrates Life Sciences. For acylcarnitines, glycerophospholipids, hexose and sphingolipids flow injection analysis (FIA)-ESI-MS/MS is used. Quantification is based on one-point calibration standard (as opposed to seven-point for the LC part). Standards specific to the individual metabolites are only commercially available for some metabolites measured using FIA; the kit includes standards specific to nine acylcarnitines (all of the format Cx, e.g. C3) and hexose, and one or two standard non-specific to each of the metabolite classes lysophosphatidylcholines, phosphatidylcholines and sphingomyelins. Thus, measurements of several acylcarnitines, all glycerophospholipids and all sphingolipids are semi-quantitative.

Finally, quality control checks are carried out [124]. Metabolites which performed poorly (e.g. 100% of samples had concentrations below the limit of detection (three times the median value of the three zero-samples) or readings for internal quality control samples were more than two to three standard deviations away from the mean) are excluded (personal communication David Achaintre, IARC). Data are then exported to spreadsheets which are sent to us from the laboratory.

3.5.2 Metabolite nomenclature

Table 3.1 on the next page shows a list of abbreviations and full biochemical names for the metabolites in this thesis. Fatty acid side chains in acylcarnitines, glycerophospholipids and sphingolipids were labelled 'Cx:y', where x and y denote the total number of carbon atoms and double bonds, respectively, in each molecule [126]. Acylcarnitines were abbreviated according to the fatty acid side chain. All glycerophospholipids were phosphatidylcholines, and sub-classes were separated by the number and type of fatty acids side chains. 'LysoPC a' denotes phosphatidylcholines with one acyl fatty acid side chain, 'PC aa' denotes two acyl side chains (diacyls), and 'PC ae' denotes one acyl and one alkyl side chain. Sphingolipids were all sphingomyelins (SM) with or without a hydroxy group (OH) attached and were also labelled according to the fatty acid side chain. Hexose is the sum of a range of monosaccharides with six carbon atoms, including glucose, fructose and galactose.

Table 3.1: Abbreviations and full names of metabolites

Abbreviation	Biochemical name
Acylcarnitines	
C0	Carnitine
C10	Decanoylcarnitine
C10:1	Decenoylcarnitine
C12	Dodecanoylcarnitine
C12:1	Dodecenoylcarnitine
C14	Tetradecanoylcarnitine
C14:1	Tetradecenoylcarnitine
C14:2	Tetradecadienylcarnitine
C16	Hexadecanoylcarnitine
C16:1	Hexadecenoylcarnitine
C18	Octadecanoylcarnitine
C18:1	Octadecenoylcarnitine
C18:2	Octadecadienylcarnitine
C2	Acetylcarnitine
C3	Propionylcarnitine
C3-DC (C4-OH)	Hydroxybutyrylcarnitine
C4	Butyrylcarnitine
C4:1	Butenylcarnitine
C5	Valerylcarnitine
C7-DC	Pimelylcarnitine
C8	Octanoylcarnitine
Amino acids	
Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartate
Cit	Citrulline
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine

Continues...

Table 3.1 continued

Abbreviation	Biochemical name
Met	Methionine
Orn	Ornithine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
t4-OH-Pro	trans-4-hydroxyproline
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
Biogenic amines	
ADMA	Asymmetric dimetylarginine
alpha-AAA	alpha-Aminoadipic acid
Creatinine	Creatinine
Kynurenine	Kynurenine
Met-SO	Methioninesulfoxide
Putrescine	Putrescine
Sarcosine	Sarcosine
SDMA	Symetric dimetylarginine
Tau	Taurine
Serotonin	Serotonin
Spermidine	Spermidine
Spermine	Spermine
Total DMA	Dimethylamine
Glycerophospholipids	
lysoPC a C14:0	lysoPhosphatidylcholine acyl C14:0
lysoPC a C16:0	lysoPhosphatidylcholine acyl C16:0
lysoPC a C16:1	lysoPhosphatidylcholine acyl C16:1
lysoPC a C17:0	lysoPhosphatidylcholine acyl C17:0
lysoPC a C18:0	lysoPhosphatidylcholine acyl C18:0
lysoPC a C18:1	lysoPhosphatidylcholine acyl C18:1
lysoPC a C18:2	lysoPhosphatidylcholine acyl C18:2
lysoPC a C20:3	lysoPhosphatidylcholine acyl C20:3
lysoPC a C20:4	lysoPhosphatidylcholine acyl C20:4
lysoPC a C24:0	lysoPhosphatidylcholine acyl C24:0
lysoPC a C28:0	lysoPhosphatidylcholine acyl C28:0
lysoPC a C28:1	lysoPhosphatidylcholine acyl C28:1
PC aa C26:0	Phosphatidylcholine diacyl C26:0
PC aa C28:1	Phosphatidylcholine diacyl C28:1
PC aa C30:0	Phosphatidylcholine diacyl C30:0
PC aa C32:0	Phosphatidylcholine diacyl C32:0
PC aa C32:1	Phosphatidylcholine diacyl C32:1
PC aa C32:2	Phosphatidylcholine diacyl C32:2
PC aa C32:3	Phosphatidylcholine diacyl C32:3
PC aa C34:1	Phosphatidylcholine diacyl C34:1

Continues...

Table 3.1 continued

Abbreviation	Biochemical name
PC aa C34:2	Phosphatidylcholine diacyl C34:2
PC aa C34:3	Phosphatidylcholine diacyl C34:3
PC aa C34:4	Phosphatidylcholine diacyl C34:4
PC aa C36:0	Phosphatidylcholine diacyl C36:0
PC aa C36:1	Phosphatidylcholine diacyl C36:1
PC aa C36:2	Phosphatidylcholine diacyl C36:2
PC aa C36:3	Phosphatidylcholine diacyl C36:3
PC aa C36:4	Phosphatidylcholine diacyl C36:4
PC aa C36:5	Phosphatidylcholine diacyl C36:5
PC aa C36:6	Phosphatidylcholine diacyl C36:6
PC aa C38:0	Phosphatidylcholine diacyl C38:0
PC aa C38:3	Phosphatidylcholine diacyl C38:3
PC aa C38:4	Phosphatidylcholine diacyl C38:4
PC aa C38:5	Phosphatidylcholine diacyl C38:5
PC aa C38:6	Phosphatidylcholine diacyl C38:6
PC aa C40:1	Phosphatidylcholine diacyl C40:1
PC aa C40:2	Phosphatidylcholine diacyl C40:2
PC aa C40:3	Phosphatidylcholine diacyl C40:3
PC aa C40:4	Phosphatidylcholine diacyl C40:4
PC aa C40:5	Phosphatidylcholine diacyl C40:5
PC aa C40:6	Phosphatidylcholine diacyl C40:6
PC aa C42:0	Phosphatidylcholine diacyl C42:0
PC aa C42:1	Phosphatidylcholine diacyl C42:1
PC aa C42:2	Phosphatidylcholine diacyl C42:2
PC aa C42:4	Phosphatidylcholine diacyl C42:4
PC aa C42:5	Phosphatidylcholine diacyl C42:5
PC aa C42:6	Phosphatidylcholine diacyl C42:6
PC ae C30:0	Phosphatidylcholine acyl-alkyl C30:0
PC ae C30:2	Phosphatidylcholine acyl-alkyl C30:2
PC ae C32:1	Phosphatidylcholine acyl-alkyl C32:1
PC ae C32:2	Phosphatidylcholine acyl-alkyl C32:2
PC ae C34:0	Phosphatidylcholine acyl-alkyl C34:0
PC ae C34:1	Phosphatidylcholine acyl-alkyl C34:1
PC ae C34:2	Phosphatidylcholine acyl-alkyl C34:2
PC ae C34:3	Phosphatidylcholine acyl-alkyl C34:3
PC ae C36:0	Phosphatidylcholine acyl-alkyl C36:0
PC ae C36:1	Phosphatidylcholine acyl-alkyl C36:1
PC ae C36:2	Phosphatidylcholine acyl-alkyl C36:2
PC ae C36:3	Phosphatidylcholine acyl-alkyl C36:3
PC ae C36:4	Phosphatidylcholine acyl-alkyl C36:4
PC ae C36:5	Phosphatidylcholine acyl-alkyl C36:5
PC ae C38:0	Phosphatidylcholine acyl-alkyl C38:0
PC ae C38:2	Phosphatidylcholine acyl-alkyl C38:2
PC ae C38:3	Phosphatidylcholine acyl-alkyl C38:3
PC ae C38:4	Phosphatidylcholine acyl-alkyl C38:4

Continues...

Table 3.1 continued

Abbreviation	Biochemical name
PC ae C38:5	Phosphatidylcholine acyl-alkyl C38:5
PC ae C38:6	Phosphatidylcholine acyl-alkyl C38:6
PC ae C40:1	Phosphatidylcholine acyl-alkyl C40:1
PC ae C40:2	Phosphatidylcholine acyl-alkyl C40:2
PC ae C40:3	Phosphatidylcholine acyl-alkyl C40:3
PC ae C40:4	Phosphatidylcholine acyl-alkyl C40:4
PC ae C40:5	Phosphatidylcholine acyl-alkyl C40:5
PC ae C40:6	Phosphatidylcholine acyl-alkyl C40:6
PC ae C42:1	Phosphatidylcholine acyl-alkyl C42:1
PC ae C42:2	Phosphatidylcholine acyl-alkyl C42:2
PC ae C42:3	Phosphatidylcholine acyl-alkyl C42:3
PC ae C42:4	Phosphatidylcholine acyl-alkyl C42:4
PC ae C42:5	Phosphatidylcholine acyl-alkyl C42:5
PC ae C44:3	Phosphatidylcholine acyl-alkyl C44:3
PC ae C44:4	Phosphatidylcholine acyl-alkyl C44:4
PC ae C44:5	Phosphatidylcholine acyl-alkyl C44:5
PC ae C44:6	Phosphatidylcholine acyl-alkyl C44:6
Hexose	
Hexose	Hexose
Sphingolipids	
SM (OH) C14:1	Hydroxysphingomyeline C14:1
SM (OH) C16:1	Hydroxysphingomyeline C16:1
SM (OH) C22:1	Hydroxysphingomyeline C22:1
SM (OH) C22:2	Hydroxysphingomyeline C22:2
SM (OH) C24:1	Hydroxysphingomyeline C24:1
SM C16:0	Sphingomyeline C16:0
SM C16:1	Sphingomyeline C16:1
SM C18:0	Sphingomyeline C18:0
SM C18:1	Sphingomyeline C18:1
SM C20:2	Sphingomyeline C20:2
SM C24:0	Sphingomyeline C24:0
SM C24:1	Sphingomyeline C24:1
SM C26:0	Sphingomyeline C26:0
SM C26:1	Sphingomyeline C26:1

C*x*:*y*, *'x'* and *'y'* are the total number of carbon atoms and double bonds, respectively, in the fatty acid side chain(s)

4. Amino acid intakes and concentrations by diet group

4.1 Introduction

Amino acids are the building blocks of proteins [127], are additionally utilised as a source of energy, and are necessary for the synthesis of a wide variety of compounds, including neurotransmitters, heme and DNA. Humans need daily supplies of protein including adequate amounts of essential amino acids, which cannot be synthesised endogenously [128]. The amino acid composition of animal protein better resembles the body's need than that of single sources of plant protein. However, as different plant proteins complement each other in terms of amino acid composition, diets without animal products can fulfil the amino acid requirement.

Although circulating concentrations of all amino acids are subject to homeostatic control [129], they are also affected by diet, metabolism, lifestyle and genetic factors [14]. Individuals who exclude some or all animal products from their diet, e.g. vegetarians and vegans, may thus have different intakes and circulating concentrations of amino acids compared to those of meat-eaters, due to differences in dietary protein sources, lifestyle and metabolism. Investigations of amino acid concentrations in such diet groups are of relevance in this thesis as studies suggest that vegans may have a lower risk of prostate cancer than meat-eaters [130, 131].

4.1.1 Previous studies

In addition to the paper published on the work presented in this chapter [17], only a few, small (including 5 to 30 participants in each group) previous studies have investigated the differences in dietary intakes or circulating concentrations of amino acids between diet groups, defined according to intake of animal protein, e.g. vegetarians versus omnivores.

Two cross-sectional studies have investigated intakes of amino acids in relation to diet group. Abdulla et al. [132, 133] measured content of essential amino acids in four 24-hours duplicate diets per participant from omnivores, vegetarians and vegans. No formal statistical tests between groups were presented, but leucine seemed lower in vegans, while phenylalanine plus tyrosine and threonine seemed higher in vegetarians than in the two other groups. Krajcovicová-Kudlácková et al. [80] found that vegetarians and vegans had lower intakes of methionine and lysine than omnivores based on a food frequency questionnaire (FFQ).

Three cross-sectional studies have compared plasma concentrations of amino acids between participants from different habitual diet groups. Anderson and Blendis [77] found lower concentrations of leucine, valine, phenylalanine and tyrosine in vegans than in vegetarians and participants eating a mixed diet. Rana and Sanders [134] compared an extensive list of amino

acid concentrations between omnivores and vegans, and found lower concentrations of leucine, valine, histidine, lysine, threonine, arginine, tyrosine and citrulline in vegans than omnivores. No differences in cysteine and methionine concentrations were found between vegans and non-vegetarians by Laidlaw et al. [135].

4.2 Aim

The aim of this analysis was to investigate differences in dietary intakes and plasma concentrations of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans in the Oxford arm of the European Prospective Investigation into Cancer and Nutrition (EPIC). The investigated amino acids were isoleucine, leucine, valine, histidine, lysine, methionine, phenylalanine, threonine, tryptophan, alanine, arginine, aspartate, glutamate, glycine, proline, serine and tyrosine in both diet and plasma; cystine in diet alone; and asparagine, glutamine, citrulline and ornithine in plasma only.

4.3 Methods

4.3.1 Study population

The EPIC-Oxford cohort comprises 65,000 participants including 14,600 men over the age of 20 who were recruited from across the United Kingdom between 1993 and 2001 [100]. The primary aim of this cohort is to investigate diet, lifestyle and risk of cancer and other diseases among people with a wide range of dietary intake, including sizable proportions of vegetarians and vegans. Most participants (89%) were recruited via postal recruitment, while the remaining participants were recruited via general practitioners' surgeries (mainly non-vegetarians). All participants gave written informed consent and the protocol for EPIC-Oxford was approved by a multi-centre research ethics committee (MREC/02/0/90).

4.3.1.1 Eligibility criteria for the current analysis

The eligibility criteria for the current analysis were: male sex; age 30 to 49; provision of a blood sample at recruitment; known smoking status and diet group; response to 80% or more of the relevant questions in the FFQ and a daily energy intake between 3.3 and 16.7 MJ (800 and 4000 kcal); and no prior cancer (excluding non-melanoma skin cancer), cardiovascular disease or treatment for any long-term illness or condition at recruitment.

Of the 110 eligible vegans (who do not eat any animal products), we selected all those aged 30 to 39 years, and randomly selected four out of every five aged 40 to 49. Additionally, eligible meat-eaters (who eat meat), fish-eaters (who do not eat meat but do eat fish) and vegetarians

(who do not eat meat or fish but do eat dairy products and/or eggs) were randomly selected in equal numbers within strata of the ages 30 to 39 and 40 to 49. A total of 392 men, 98 in each diet group, were included in this analysis.

4.3.2 Measurement of dietary intake

At recruitment, participants filled in a validated semi-quantitative FFQ [107, 108]. Participants were asked how often they consumed 130 foods and drinks, 113 of which were relevant to vegetarians and vegans. Mean daily food and drink intakes were estimated using specified portion sizes [109], and mean daily nutrient intakes were estimated mostly using data from the fifth edition of 'McCance and Widdowson's The Composition of Foods' and its supplements [110, 111, 112, 113, 114, 115, 116, 117, 118, 119]. Participants were asked if they ate meat, fish, dairy or eggs and the answers were used to categorise participants as meat-eaters, fish-eaters, vegetarians and vegans.

All items in the FFQ were categorised into seven mutually exclusive food types based on the main source of the protein of the food, i.e. meat products, fish products, dairy products, egg products, mixed animal products, non-soya plant products and soya products.

4.3.2.1 Estimation of amino acid intake

Data on amino acid content are only available from a limited number of foods ($n = 118$) from UK nutrition tables [136]. Mean daily intakes of 18 amino acids were therefore estimated using data from 'United States Department of Agriculture (USDA), National Nutrient Database for Standard Reference, release 26' [120]. This database included data for 18 standard amino acid for 4,929 foods, and 3,533 additional foods had data on some but not all 18 amino acids.

The USDA database was searched for food items similar to the food items in the FFQ. When choosing food items to best match the questions asked, focus was put on the availability of amino acid data and on the protein content and source. Thus, the coding might not be useful for estimation of intake of other nutrients. Appropriate food items with complete amino acid content was found for the majority of food items in the FFQ. However, for six food items no amino acid data were available in the USDA nutrient table, i.e. wine, fortified wine, low calorie fizzy drinks, squash and cordial, sugar and yeast extract. Of these, only yeast extract has a high protein content (23.9 g protein per 100 g [120]) and data for 'Marmite' were taken from the 4th edition of 'McCance and Widdowson's The Composition of Foods' [136]. The remaining five food items had negligible protein contents ($\leq$ 0.2 g protein/100 g [120]) and their amino acid contents were set to zero.

Aspartic/glutamic acid, aspartate/glutamate and asparagine/glutamine: The USDA nutrient tables provide information on aspartic and glutamic acids. At physiological pH these amino acids release a hydrogen molecule and are thus called aspartate and glutamate [137]. To use consistent terminology between intake and plasma concentrations, only aspartate and glutamate will be used in this thesis. The USDA nutrient tables do not include information on asparagine and glutamine due to the assay methods used [138]; acid hydrolysis transforms asparagine and glutamine to aspartate and glutamate, respectively, which are consequently overestimated in nutrient tables.

4.3.3 Blood collection and laboratory analysis

At recruitment, participants were also asked to have a blood sample taken at their local general practitioner's surgery; participants were not required to fast. Blood samples were sent at ambient temperature to the laboratory, where they were processed for long-term storage in liquid nitrogen (−196°C) until 2011 and subsequently in electric freezers (−80°C).

In 2013, plasma samples (with citrate as an anticoagulant) were transported on dry ice to the International Agency for Research on Cancer, Lyon, France, for assaying. Prior to the assay for the current analysis, samples had undergone two or three freeze-thaw cycles; samples from each diet group were equally distributed by number of cycles. The targeted metabolomics assay, Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbrcuk, Austria) was used to quantify 21 amino acids. These were separated by high performance liquid chromatography (HPLC) before injection into the mass spectrometer (QTRAP® 5500, AB Sciex, Framingham, Massachusetts, USA) .

For quality control, blinded samples of pooled plasma were included in all batches (four to six samples per batch and 26 in total). The lab technicians were blinded to diet group and quality control samples.

To assess the precision of the assay, coefficients of variation (CV) were calculated as the standard deviation divided by the mean of the quality control samples times 100, for each amino acid [139]; overall CVs were calculated based on all quality control samples and within-batch CVs were calculated as the average of CVs from each batch. The mean overall CV was 12.0%, covering a range from 5.3% to 17.3% for alanine and arginine at mean concentrations of 602 and 234 μmol/L, respectively (Table 4.1 on page 46). The lowest within-batch CV of 4.1% was also observed for alanine, while phenylalanine had the highest of 15.5% at a mean concentration of 113 μmol/L; the mean within-batch CV was 10.3%.

No samples had concentrations below the lower limit of detection, while for five amino acids, a few values ($n \leq 2$) were higher than the highest standard and were set equal to this value (400 μmol/L for leucine, histidine, threonine and ornithine, and 800 μmol/L for glutamate). For some

participants, it was not possible to assess the concentration of lysine ($n = 11$), phenylalanine ($n = 93$), tryptophan ($n = 2$), glutamine ($n = 2$), proline ($n = 45$) or ornithine ($n = 1$) accurately due to signal saturation of the detector in the mass spectrometer; thus no data were available. The number of unavailable concentrations did not differ significantly by diet group for lysine, phenylalanine or proline (no comparison was made for the other amino acids due to the low number of affected samples). Participants were excluded only from the analyses of amino acids for which they did not have data available.

4.3.4 Covariates

4.3.4.1 Personal characteristics

At recruitment, participants were asked additional questions on smoking habits, body size, physical activity, history of disease, treatments, education and job situation.

In addition to self-report, weight and height were also measured in a sub-sample of the cohort and there was good agreement between self-reported and measured values (correlation coefficient (r)> 0.9) [122]. Self-reported or measured values, if available, were used to calculate body mass index (BMI; kg/m^2).

4.3.4.2 Characteristics related to blood sampling and handling

At blood collection, time since last food or drink, use of medication or dietary supplements on the day, the date and time of day, were recorded. Date and time of day of blood processing and storage were also recorded. Finally, data were available on number of freeze-thaw cycles, whether or not samples were lipemic (cloudy sample due to lipids in the sample) and the degree of hemolysis (red colouring of the samples due to rupture of erythrocytes).

4.3.5 Statistical analysis

Participant characteristics, nutrient intakes and factors related to blood handling were compared across diet groups using the Kruskal-Wallis one-way of analysis of variation (ANOVA) and χ^2 test for continuous and categorical variables, respectively.

All amino acid intakes and concentrations were logarithmically transformed to approximate the normal distribution, except concentrations of tryptophan, aspartate and glutamine, which were closer to the normal distribution when on the original scale.

Geometric means and 95% confidence intervals (CI) of amino acid intakes and concentrations were computed for categories of personal characteristics, nutritional factors and factors related to the blood samples. Differences were tested using ANOVA adjusted for age ($30-34; 35-39; 40-44; 45-49$ years) and BMI ($< 22.5; 22.5-24.9; \geq 25$ kg/m^2; unknown) and allowing

for multiple testing using the Bonferroni method within each variable (see Section 4.3.5.1 for details).

Differences between diet groups in mean intakes and concentrations of amino acids were tested using ANOVA. The analyses of concentrations and intakes were both adjusted for age ($30-34; 35-39; 40-44; 45-49$ years), BMI ($<22.5; 22.5-24.9; \geq 25$ kg/m^2; unknown), smoking status (never; former; current) and alcohol intake ($<1; 1-7; 8-15; \geq 16$ g/d). Energy intake (continuous) was included in the model for intakes only, while time since last food or drink at blood collection ($<1.5; 1.5-2.9; 3.0-4.4; \geq 4.5$ hours; unknown) and time between blood collection and processing (fourths of the distribution corresponding to $<25; 25-41; 41-72; \geq 72$ hours; unknown) were included in the model for plasma concentrations.

Spearman's rank correlation coefficients (r) were calculated: (i) between pairs of plasma amino acids, (ii) between plasma amino acids and percentage energy from protein from food types (excluding diet groups which do not consume the food type), and (iii) between amino acids measured in diet and plasma.

4.3.5.1 Multiple testing

When performing multiple statistical tests, risk of type I error increases; for p independent tests, the expected number of false positive conclusions is p times the per test significance level (α) [140]. To account for this, the probability of getting one or more false positive conclusion (i.e. the family wise error rate; FWER) can be controlled by calculating a corrected significance level α' such that the FWER $\leq \alpha$. The Bonferroni method is one approach to controlling the FWER; α' is calculated as $\alpha' = \alpha/$number of tests. This assures that, on average, the risk of drawing one or more false positive conclusions is α or less [140].

The Bonferroni method was used in all analyses of amino acid intakes and concentrations and α' was set to 0.05, resulting in $\alpha' = 0.0028$ for intake (18 tests) and $\alpha' = 0.0024$ for concentrations (21 tests).

4.3.5.2 Statistical power

With 98 men in each diet group, we had more than 80% power to detect a difference in geometric mean concentration of approximately 10% between the diet group with the highest and the lowest concentration, when using the Bonferroni corrected significance level.

4.3.5.3 Sensitivity analysis

To investigate whether the method of recruitment (general practitioner versus postal recruitment), which could be related to aspects of blood handling, affected the results of the amino acid concentrations by diet group, a sensitivity analysis was carried out excluding participants who

were recruited via their general practitioner (meat-eaters $n = 37$; fish-eaters $n = 3$; vegetarians $n = 1$; vegans $n = 0$).

The agreement between protein intakes estimated using the UK and USDA nutrient tables was assessed using a Bland–Altman plot.

4.3.5.4 Statistical software

All analyses were performed using Stata Statistical Package version 13 (Stata Corp., Texas, USA).

4.4 Results

Participants had followed their diet long term; the fish-eaters, vegetarians and vegans had on average followed their diet for 11, 12 and 9 years, respectively.

4.4.1 Covariates by diet group

Body mass index differed by up to 2.3 kg/m^2 between diet groups; it was highest in meat-eaters and lowest in vegans (Table 4.2 on page 47). Similarly, energy intake and percent energy from protein were highest in meat-eaters and lowest in vegans. Separating the protein intake by food types revealed differences in percent energy from protein from all plant products, non-soya and soya plant products; meat-eaters had the lowest intake while vegans had the highest. Protein from all animal products also differed by diet group; it was highest in meat-eaters and lowest in vegetarians (zero in vegans by definition). Although the median intake of protein from mixed animal products was the same in all three relevant diet groups (vegans had per definition no intake), a statistically significant difference between diet groups was seen ($p = 0.02$). In contrast, no differences between diet groups were observed for the other three types of animal products for which comparison was possible (fish, dairy and egg products).

There was some variation by diet group in median duration since last meal or drink at blood collection ($\leq$ 30 minutes) and in time between blood collection and processing ($\leq$ 19 hours). No statistical test was performed for recruitment method as few ($\leq 3.1\%$) from each of the non-meat-eating diet groups were recruited via general practitioners. Of note, a higher proportion (37.8%) of meat-eaters were recruited via this method.

4.4.2 Amino acid intakes and concentrations by covariates

All intakes of amino acids were strongly, positively related to energy and protein intakes, after correction for multiple testing (Table 4.3 on page 48). No significant association was seen

between amino acid intakes and age, BMI, smoking, physical activity, alcohol intake and use of supplements or medications at blood collection.

Several ($n = 6$) of the amino acid concentrations were associated with time since last food or drink at blood collection, after taking multiple testing into account (Table 4.3 on page 48). For four of these, the highest concentration was seen in men who had had blood taken between one to three hours after a meal, while the lowest was seen at more than four hours after a meal. Two amino acid concentrations were lower if samples were taken between 10 am and 1 pm than at other times of the day. Many amino acid concentrations ($n = 12$) were positively associated with time from blood collection to processing, while an additional few ($n = 3$) were also associated but with a different shape. Lipemic samples had lower concentration of threonine than non-lipemic samples, and some ($n = 5$) amino acid concentrations were associated with hemolysis, four of which were higher in hemolytic than non-hemolytic samples. Finally, two amino acid concentrations were related to postal recruitment, each in different directions. In contrast, personal characteristics and intake of medication or supplements at blood collection were not associated with amino acid concentrations, except an association between protein intake and glutamate.

4.4.3 Intake of amino acids

The estimated protein intake from the UK and US nutrient tables showed good agreement ($r = 0.98$) although estimates based on the USDA database were slightly higher (3.2 g/d; 95%CI: 2.8 – 3.6 g/d; Figure 4.1 on page 53).

4.4.3.1 Estimated amino acid intakes by diet group

Estimated mean intakes of all 18 amino acids varied by diet group ($p < 0.0001$ for all; Table 4.4 on page 49). Meat-eaters had the highest intakes, followed by fish-eaters, vegetarians, and vegans, except for arginine, aspartate, glycine and proline. For these amino acids, meat-eaters also had the highest intakes, but were followed by vegans for arginine and glycine, while other deviations from the above pattern were seen for aspartate and proline. The largest percentage differences between diet groups were for intakes of lysine and methionine, which in vegans were 56% and 53% of those in meat-eaters, respectively.

4.4.4 Plasma concentrations of amino acids

4.4.4.1 Plasma amino acids concentrations by diet group

The concentrations of six of the 21 plasma amino acids varied significantly by diet group after allowing for multiple testing (Table 4.5 on page 50). For lysine, methionine, tryptophan and

tyrosine, fish-eaters and vegetarians generally had the highest concentrations, followed by meat-eaters, while vegans had the lowest concentrations for all four. For alanine, the highest concentration was seen in fish-eaters, followed by vegetarians and vegans, and lowest in meat-eaters. The largest difference was seen for glycine; mean concentration in vegans was 16% higher than in meat-eaters, who had the lowest concentration, with intermediate concentrations observed in fish-eaters and vegetarians. Differences between diet groups in concentrations of leucine, valine, histidine, aspartate, glutamate, glutamine, citrulline and ornithine were significant at a conventional significance level ($p < 0.05$) but not after correction for multiple testing.

When restricting the analysis to men recruited via the postal method, results were not materially changed, except for alanine, which no longer differed statistically significantly between diet groups after allowing for multiple testing (% difference to meat-eaters: +11, +9, +8% in fish-eaters, vegetarians and vegans, respectively; $p = 0.03$).

4.4.4.2 Correlations between pairs of plasma amino acids

Concentrations of many of the plasma amino acids were correlated (Figure 4.2 on page 54); 62% of the correlations were statistically significant after allowance for multiple testing ($-0.23 \geq r \geq 0.23$; $p < 0.0002$). The strongest positive correlations were observed between the essential amino acids ($r \leq 0.91$). The essential amino acids were also positively correlated with several other amino acids; especially strongly with tyrosine. For all amino acids, the vast majority of correlations were positive, but some negative correlations were found, i.e. arginine with alanine and glutamate with ornithine and glutamine ($-0.34 \leq r \leq -0.23$).

4.4.5 Correlations between dietary factors and plasma concentrations of amino acids

4.4.5.1 Correlations between amino acid intakes and concentrations

Modest but statistically significant positive correlations were found between intake and plasma concentration of leucine, lysine, methionine, tryptophan and tyrosine ($0.16 \leq r \leq 0.21$), after allowing for multiple testing (Table 4.6 on page 51).

4.4.5.2 Correlations between protein intake from food types and amino acid concentrations

Concentrations of some plasma amino acids were correlated with intake of protein from plant products, after allowing for multiple testing (Table 4.7 on page 52). For soya products, inverse correlations were seen with valine, lysine, methionine and tyrosine concentrations. A similar pattern was seen for non-soya plant products, although only correlations for plasma lysine and

tyrosine were significant. In contrast, glycine concentration was positively correlated with protein from soya products.

There were no statistically significant correlations between protein from animal food types and plasma amino acids, after allowing for multiple testing (Table 4.7 on page 52). However, positive correlations of similar magnitude to the ones for plant protein were seen for dairy and egg protein, but these analyses had lower power because vegans did not contribute.

4.5 Discussion

4.5.1 Main findings

This analysis showed that plasma concentrations of six amino acids, namely lysine, methionine, tryptophan, alanine, glycine and tyrosine, differed by habitual diet group. These differences were less pronounced and did not mirror those observed for amino acid intakes, the majority of which were highest in meat-eaters followed by fish-eaters, then vegetarians and finally vegans. Concentrations of the essential amino acids and tyrosine were generally highest in fish-eaters and vegetarians, followed by meat-eaters and all lowest in vegans. Observing the lowest concentrations in vegans was not unexpected, as essential amino acids are not synthesised endogenously and are less abundant in plant than animal protein [127, 128]. For alanine, the highest concentrations were observed in fish-eaters, followed by vegetarians and vegans, and were lowest in meat-eaters. Vegans had the highest concentration of glycine followed by fish-eaters and vegetarians, while meat-eaters had the lowest concentrations. Whereas most plasma amino acids were correlated with each other, only a few significant correlations were seen between plasma concentrations and dietary intakes of amino acids, and between plasma amino acids and intakes of protein from different food sources.

4.5.2 Intake of amino acids

The observed differences between diet groups in amino acid intakes may be explained by the protein intake being lower the more animal products were excluded from the diet. The observed trend in protein intake across diet groups may partly reflect an increasing degree of underestimation of intake as more animal foods are omitted. Some plant foods eaten by vegetarians and vegans may not be in the FFQ, and vegetarians and vegans may eat larger portions of some plant foods than the standard portion sizes assigned. Also, the validation of the FFQ showed that protein intake was particularly difficult to estimate [107, 108].

Lysine or methionine plus cysteine are the limiting amino acids in many plant proteins [128], meaning that they are present in the lowest amount relative to the requirement. This is in line

with our findings of lysine and methionine accounting for the largest differences between intakes in vegans and meat-eaters (nearly 50% lower in vegans).

The results on amino acid intakes are further supported by a previous study which found similar percent differences in intakes of lysine and methionine between omnivores, lacto-ovo-vegetarians and vegans [80] to those in the current study. Somewhat contrasting, another study found that intakes of several amino acids, including lysine, were highest in lacto-vegetarians, followed by meat-eaters and lowest in vegans, while vegans had the highest intake of other amino acids, including methionine plus cysteine [132, 133]. However, in this latter study, the number of participants was small ($n = 12$ men) and no statistical tests were performed.

4.5.3 Amino acid concentrations

The plasma concentration of amino acids is a result of a complex interplay between dietary intake, tissue breakdown and *de novo* synthesis [129], and other factors such as gut microbial synthesis increasing the concentration [141], and tissue uptake and excretion reducing the circulating concentration [129]. This might, in addition to the recognised limitations of estimating nutrient intakes using a FFQ [142], at least partly explain the differences in our results for dietary and plasma amino acids by diet group, as well as the limited strength of correlation between intakes and plasma concentrations. Furthermore, as some diet groups were excluded from the analysis of intakes of protein from animal sources and amino acid concentrations because of zero intake, lower power may also have contributed to the lack of significant correlations in this analysis.

The largest percentage difference between meat-eaters and vegans in plasma concentration was found for glycine (16%), for which vegans had the highest concentration. Plasma glycine concentration was also positively correlated with intake of protein from plant products – especially soya products, which have high glycine content [120] and are consumed in greater quantities by the vegans than the other groups. This could at least in part explain why vegans had the highest glycine concentration.

There are a few previous studies available on plasma amino acid concentrations in individuals who exclude some or all animal products from their diet. The differences by diet group we observed for lysine and tyrosine were supported by similar findings from previous studies, as were the null results for isoleucine, phenylalanine, asparagine, proline, serine and ornithine [77, 134]. For leucine, valine, histidine, threonine, arginine and citrulline an association was found in previous studies but not in ours, while the opposite was the case for methionine, tryptophan, alanine and glycine [77, 134, 135]. However, the previous studies were all limited by small numbers of participants ($n < 36$).

4.5.4 Strengths and limitations

To our knowledge, this is the largest study to date of amino acids in the diet or in the circulation by habitual diet group, and on average participants had followed their diet for several years. Moreover, it provides data on both dietary intakes and plasma concentrations of amino acids as well as information on a wide range of covariates. These strengths facilitated robust comparisons between diet groups while limiting the likelihood of confounding. Moreover, it enabled us to investigate the correlations between dietary intake and circulating concentration of amino acids.

The main limitations are related to the measurement of amino acid intake and concentration.

Firstly, although care was taken to match the food items in the FFQ with appropriate food items in the US database and the comparison of the protein intake estimated using the US and the UK nutrient tables only showed minor differences, the amino acid content of the selected foods may differ somewhat between the US and the UK.

Secondly, the measurements of plasma amino acids could be affected by pre-analytic conditions such as fasting status, process delay and freeze-thaw cycles. In line with our results, some amino acids (namely leucine/isoleucine, tryptophan and tyrosine) have been suggested to reflect time since last meal [143]. Many amino acid concentrations were also positively associated with time from blood collection to processing, during which they were kept at room temperature. Methodological papers have reported similar findings [144, 145]. Furthermore, the process delay was shorter for meat-eaters than the other diet groups. Most amino acids have been found to be stable during up to four freeze-thaw cycles [145], although isoleucine, tryptophan and valine have been reported to decrease in concentration [144]. To minimise the risk of these pre-analytical factors affecting our results, we adjusted for fasting status and process delay, and the number of freeze-thaw cycles was distributed equally among diet groups. However, residual confounding especially from process delay cannot be ruled out.

Thirdly, the analytical variation in this study was somewhat higher than in previous publications, in which mean overall CV was 7% compared to 12% in the current study [126, 146]. Any such random analytical variation might attenuate the association between diet group and amino acid concentration.

4.5.5 Potential impact of findings

It is not known whether the observed differences by diet group in dietary intakes or plasma concentrations of amino acids might translate to differences in risk of non-communicable diseases. One study of type 2 diabetes found serum glycine and phenylalanine concentrations to be inversely and positively associated, respectively, with risk of this disease [147], while a second study reported a higher risk of diabetes in individuals with higher plasma concentrations of

several amino acids including tyrosine [148]. Taken together with the current results, this may imply that vegans might potentially have a lower risk of diabetes than meat-eaters, although further research is needed. Associations have also been found between high intake of methionine and lower risk of breast [149] and pancreatic [150] cancers. Future studies might help develop our understanding of the interplay between diet, alterations in amino acid metabolism and disease development, including that of prostate cancer.

4.6 Conclusion

British men belonging to different habitual diet groups have significantly different plasma concentrations of lysine, methionine, tryptophan, alanine, glycine and tyrosine. The differences in plasma concentrations were less marked than, and did not closely mirror, those seen for amino acid intakes; intakes were all highest in meat-eaters and mostly lowest in vegans.

4.7 Tables

Table 4.1: Coefficients of variation of blinded pooled plasma samples

	Mean concentration,	Coefficient of variation, %	
	µmol/L	Overall[1]	Within-batch[2]
Branched-chain essential amino acids			
Isoleucine	97	8.7	7.5
Leucine	218	8.3	7.9
Valine	229	15.8	15.0
Other essential amino acids			
Histidine	109	11.5	9.8
Lysine	221	15.2	12.5
Methionine	28	11.8	11.1
Phenylalanine[3]	113	16.2	15.5
Threonine	171	10.7	7.6
Tryptophan	63	8.7	8.5
Non-essential amino acids			
Alanine	602	5.3	4.1
Arginine	234	17.3	11.5
Asparagine	60	12.1	9.1
Aspartate	52	11.3	11.0
Glutamate	302	13.4	10.3
Glutamine	348	13.0	12.0
Glycine	412	11.0	9.9
Proline	259	12.9	13.0
Serine	209	15.8	11.6
Tyrosine	76	8.3	7.8
Non-standard amino acids			
Citrulline	37	11.3	10.0
Ornithine	79	12.6	10.1

[1] Based on 26 blinded pooled plasma samples

[2] Based on four to six blinded pooled plasma samples in each of six batches

[3] Due to signal saturation affecting all pooled samples in batch 3, the within-batch CV was calculated without taking batch 3 into account

Table 4.2: Characteristics, nutrient intakes and blood sample characteristics by habitual diet group

	Median (inter-quartile range) or n (%)[1]				p_{dif}[2]
	Meat-eaters (n=98)	Fish-eaters (n=98)	Vegetarians (n=98)	Vegans (n=98)	
Characteristics					
Age at blood collection, years	44 (37, 44)	41 (36, 45)	43 (36, 44)	40 (35, 44)	>0.9
BMI[3], kg/m^2	24.4 (22.1, 25.9)	22.7 (21.2, 24.2)	22.7 (21.5, 24.9)	22.1 (20.4, 24.0)	0.0001
Current smoking, n (%)	14 (14.3)	9 (9.2)	7 (7.1)	7 (7.1)	0.1
Nutrient intake					
Energy, kJ	9137 (7921, 10957)	8734 (7511, 10127)	9011 (7597, 10933)	7652 (6091, 8932)	0.0001
Protein, % of energy	15.0 (13.6, 16.9)	13.6 (12.3, 15.4)	13.3 (11.8, 14.2)	12.6 (11.6, 13.9)	0.0001
Protein from all plant products[4], % of energy	6.0 (5.4, 7.1)	7.4 (6.5, 8.8)	7.8 (6.8, 9.3)	12.3 (11.3, 13.7)	0.0001
Protein from non-soya plant products[4], % of energy	5.9 (5.2, 6.9)	7.0 (6.2, 7.9)	7.2 (6.4, 8.3)	9.6 (8.6, 10.7)	0.0001
Protein from soya products[4], % of energy	0.0 (0.0, 0.0)	0.3 (0.2, 1.0)	0.4 (0.2, 1.1)	2.8 (1.8, 3.4)	0.0001
Protein from all animal products[4], % of energy	8.4 (6.7, 10.5)	5.6 (4.7, 7.5)	4.6 (3.8, 6.0)	–	0.0001
Protein from meat products[4], % of energy	3.3 (1.6, 4.7)	–	–	–	–
Protein from fish products[4], % of energy	1.0 (0.6, 1.6)	1.2 (0.6, 1.7)	–	–	0.5
Protein from dairy products[4], % of energy	3.4 (2.4, 4.3)	3.6 (2.5, 4.7)	3.7 (2.9, 4.9)	–	0.1
Protein from egg products[4], % of energy	0.2 (0.1, 0.3)	0.2 (0.1, 0.5)	0.2 (0.1, 0.4)	–	0.1
Protein from mixed animal products[4], % of energy	0.5 (0.2, 0.7)	0.5 (0.4, 0.8)	0.5 (0.4, 0.8)	–	0.02
Carbohydrates, % of energy	51.2 (47.1, 55.6)	52.2 (48.3, 57.4)	54.3 (48.7, 58.2)	55.6 (52.1, 60.9)	0.0001
Fat, % of energy	31.9 (28.9, 34.9)	32.2 (27.2, 35.0)	31.3 (27.4, 35.3)	30.4 (25.2, 34.4)	0.2
Alcohol, g/d	9.8 (2.5, 17.2)	9.7 (2.9, 29.7)	10.2 (5.1, 23.0)	5.1 (1.0, 14.0)	0.0007
Variables related to blood collection and handling					
Postal recruitment	61 (62.2)	95 (96.9)	97 (99.0)	98 (100.0)	–
Medication or supplement at blood collection[3], n (%)	60 (61.2)	61 (63.5)	65 (66.3)	66 (67.4)	0.8
Time since last food or drink at blood collection[3], h	2.0 (1.0, 3.5)	2.3 (1.3, 4.5)	2.4 (1.5, 4.5)	2.5 (1.3, 4.3)	0.03
Time of blood collection[3], h:min	11:20 (10:00, 15:27)	10:45 (09:28, 15:30)	10:10 (09:30, 14:45)	10:40 (09:45, 16:08)	0.2
Time from blood collection to processing[3], h	25.8 (23.4, 67.1)	44.4 (26.1, 72.1)	42.9 (25.0, 71.4)	41.7 (24.7, 71.9)	0.004
Lipemic, n (%)	23 (23.5)	13 (13.3)	24 (24.5)	17 (17.4)	0.2
Hemolytic, n (%)	1 (1.0)	7 (7.1)	7 (7.1)	3 (3.1)	0.8

[1] Values are median (inter-quartile range) if not otherwise specified as n (%)

[2] Differences between diet groups were tested using the Kruskal-Wallis ANOVA and χ^2 test for continuous and categorical variables, respectively

[3] Unknown for some participants: BMI n = 20; at blood collection: medication or supplement n = 2; time since last food or drink n = 9; time n = 10; time to processing n = 10

[4] Foods were categorised into food types based on the main protein source of the food. Medians and inter-quartile ranges do not necessarily add up because mixed foods for which there are non-vegetarian as well as vegetarian and vegan versions were categorised as one food type only, e.g. mixed animal products

Table 4.3: Amino acid intakes and concentrations significantly associated with participant characteristics and factors related to blood sampling and processing[1]

	Amino acid intake[2]	Amino acid concentration[2]
Age		
BMI		
Smoking status		
Physical activity level		
Energy intake	All ↑	
Protein intake	All ↑	Glutamate
Alcohol intake		
Medication or supplement at blood collection		
Time since last food or drink at blood collection	N/A	Isoleucine Leucine ↓ Methionine ↓ Tryptophan Proline Tyrosine
Time of day at blood collection	N/A	Isoleucine Phenylalanine
Time from blood collection to processing	N/A	Isoleucine ↑ Leucine ↑ Valine ↑ Histidine ↑ Lysine ↑ Methionine Phenylalanine ↑ Threonine ↑ Alanine ↑ Arginine Glutamate ↑ Glycine ↑ Serine ↑ Tyrosine Ornithine ↑
Lipemic	N/A	Threonine ↓
Hemolysis	N/A	Histidine ↑ Alanine ↑ Glutamate ↑ Glutamine ↓ Glycine ↑
Postal recruitment	N/A	Alanine ↑ Arginine ↓

[1] Statistical significance was assessed after Bonferroni correction ($p < 0.0028$ for intakes and $p < 0.0024$ for concentrations). The analyses were adjusted for age ($30-34$; $35-39$; $40-44$; $45-49$ years) and BMI (< 22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), where these were not the variables of interest

[2] ↑ indicates a positive association or higher concentration in the 'yes'-category than the 'no'-category. ↓ indicates an inverse association or lower concentrations in the 'yes'-category than the 'no'-category. Metabolites not marked had a differently shaped association

Table 4.4: Intakes of amino acids by habitual diet group[1]

	Geometric mean (95% confidence interval), g/d **% mean difference compared to meat-eaters**				p_{diff}[2]
	Meat-eaters (n = 98)	**Fish-eaters (n = 98)**	**Vegetarians (n = 98)**	**Vegans (n = 98)**	
Branched-chain essential amino acids					
Isoleucine	3.54 (3.41, 3.68)	3.12 (3.00, 3.24)	2.95 (2.85, 3.06)	2.47 (2.38, 2.57)	**<0.0001**
	Ref.	**-12**	**-17**	**-30**	
Leucine	6.13 (5.90, 6.35)	5.51 (5.30, 5.71)	5.21 (5.03, 5.39)	4.33 (4.17, 4.49)	**<0.0001**
	Ref.	**-10**	**-15**	**-29**	
Valine	4.14 (3.99, 4.30)	3.78 (3.63, 3.91)	3.60 (3.47, 3.73)	2.95 (2.85, 3.07)	**<0.0001**
	Ref.	**-9**	**-13**	**-29**	
Other essential amino acids					
Histidine	2.12 (2.04, 2.20)	1.83 (1.77, 1.90)	1.72 (1.66, 1.78)	1.52 (1.46, 1.57)	**<0.0001**
	Ref.	**-13**	**-19**	**-28**	
Lysine	5.01 (4.78, 5.24)	4.14 (3.97, 4.33)	3.76 (3.60, 3.93)	2.82 (2.69, 2.95)	**<0.0001**
	Ref.	**-17**	**-25**	**-44**	
Methionine	1.67 (1.60, 1.74)	1.38 (1.33, 1.43)	1.24 (1.20, 1.29)	0.88 (0.84, 0.92)	**<0.0001**
	Ref.	**-18**	**-26**	**-47**	
Phenylalanine	3.55 (3.43, 3.68)	3.34 (3.24, 3.46)	3.21 (3.11, 3.32)	2.93 (2.82, 3.03)	**<0.0001**
	Ref.	**-6**	**-10**	**-18**	
Threonine	2.99 (2.88, 3.10)	2.61 (2.52, 2.71)	2.43 (2.34, 2.52)	2.19 (2.11, 2.27)	**<0.0001**
	Ref.	**-13**	**-19**	**-27**	
Tryptophan	0.93 (0.90, 0.96)	0.86 (0.84, 0.89)	0.82 (0.79, 0.85)	0.77 (0.74, 0.79)	**<0.0001**
	Ref.	**-7**	**-12**	**-18**	
Non-essential amino acids					
Alanine	3.56 (3.42, 3.69)	2.97 (2.87, 3.08)	2.67 (2.57, 2.77)	2.63 (2.53, 2.73)	**<0.0001**
	Ref.	**-16**	**-25**	**-26**	
Arginine	4.13 (3.96, 4.31)	3.69 (3.54, 3.85)	3.36 (3.23, 3.51)	3.92 (3.75, 4.09)	**<0.0001**
	Ref.	**-11**	**-19**	**-5**	
Aspartate	7.01 (6.75, 7.28)	6.44 (6.21, 6.67)	6.00 (5.79, 6.22)	6.33 (6.10, 6.58)	**<0.0001**
	Ref.	**-8**	**-14**	**-10**	
Cystine	1.04 (1.00, 1.08)	0.94 (0.91, 0.97)	0.88 (0.85, 0.91)	0.84 (0.81, 0.87)	**<0.0001**
	Ref.	**-9**	**-15**	**-19**	
Glutamate	16.10 (15.59, 16.63)	15.56 (15.09, 16.05)	15.10 (14.64, 15.57)	14.06 (13.61, 14.52)	**<0.0001**
	Ref.	**-3**	**-6**	**-13**	
Glycine	3.12 (3.00, 3.25)	2.56 (2.46, 2.66)	2.28 (2.20, 2.37)	2.61 (2.50, 2.71)	**<0.0001**
	Ref.	**-18**	**-27**	**-16**	
Proline	5.61 (5.42, 5.80)	5.47 (5.30, 5.66)	5.47 (5.30, 5.66)	4.32 (4.17, 4.47)	**<0.0001**
	Ref.	**-2**	**-2**	**-23**	
Serine	3.70 (3.58, 3.84)	3.56 (3.44, 3.68)	3.43 (3.31, 3.54)	3.08 (2.97, 3.19)	**<0.0001**
	Ref.	**-4**	**-8**	**-17**	
Tyrosine	2.71 (2.61, 2.82)	2.48 (2.39, 2.57)	2.36 (2.28, 2.45)	1.86 (1.79, 1.94)	**<0.0001**
	Ref.	**-9**	**-13**	**-31**	

[1] Adjusted for age ($30-34$; $35-39$; $40-44$; $45-49$ years), BMI (<22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), smoking status (never; former; current), alcohol intake (<1; $1-7$; $8-15$; ≥ 16 g/d) and energy intake (continuously)

[2] p-values refer to p_{diff} across all four diet groups calculated by ANOVA. Conventional p-values are shown and those marked in bold were significant after Bonferroni correction ($p < 0.0028$)

Table 4.5: Plasma concentrations of amino acids by habitual diet group[1]

	Geometric mean (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				p_{diff}[2]
	Meat-eaters (n = 98)	**Fish-eaters (n = 98)**	**Vegetarians (n = 98)**	**Vegans (n = 98)**	
Branched-chain essential amino acids					
Isoleucine	99 (94, 103)	99 (95, 103)	102 (97, 106)	96 (92, 100)	0.3
	Ref.	**0**	**+3**	**-3**	
Leucine	205 (197, 214)	208 (200, 216)	210 (202, 218)	191 (184, 199)	0.005
	Ref.	**+1**	**+2**	**-7**	
Valine	230 (221, 239)	233 (225, 242)	233 (225, 242)	217 (209, 225)	0.02
	Ref.	**+1**	**+1**	**-6**	
Other essential amino acids					
Histidine	114 (110, 118)	122 (118, 126)	119 (115, 122)	117 (113, 120)	0.04
	Ref.	**+7**	**+4**	**+2**	
Lysine[3]	241 (230, 253)	242 (232, 253)	234 (224, 245)	210 (201, 219)	**<0.0001**
	Ref.	**+1**	**-3**	**-13**	
Methionine	29 (28, 31)	30 (29, 31)	31 (30, 32)	27 (26, 28)	**0.0001**
	Ref.	**+2**	**+5**	**-8**	
Phenylalanine[3]	95 (92, 99)	101 (97, 105)	100 (97, 104)	97 (93, 101)	0.1
	Ref.	**+6**	**+5**	**+2**	
Threonine	164 (158, 170)	168 (163, 174)	167 (162, 173)	165 (159, 171)	0.7
	Ref.	**+3**	**+2**	**+1**	
Tryptophan[3,4]	69 (67, 72)	71 (68, 73)	72 (70, 74)	65 (63, 68)	**0.001**
	Ref.	**+2**	**+4**	**-6**	
Non-essential amino acids					
Alanine	564 (540, 590)	644 (618, 671)	627 (602, 654)	621 (595, 648)	**0.0004**
	Ref.	**+14**	**+11**	**+10**	
Arginine	51 (46, 57)	44 (40, 49)	44 (40, 49)	44 (39, 48)	0.1
	Ref.	**-14**	**-14**	**-15**	
Asparagine	92 (89, 95)	97 (94, 101)	96 (92, 99)	98 (95, 102)	0.07
	Ref.	**+6**	**+4**	**+7**	
Aspartate[4]	66 (62, 69)	64 (61, 67)	69 (66, 72)	69 (66, 72)	0.04
	Ref.	**-3**	**+5**	**+5**	
Glutamate	261 (247, 277)	296 (281, 312)	277 (263, 292)	262 (248, 277)	0.004
	Ref.	**+13**	**+6**	**0**	
Glutamine[3,4]	517 (497, 536)	506 (488, 525)	525 (507, 544)	547 (529, 566)	0.02
	Ref.	**-2**	**+2**	**+6**	
Glycine	390 (375, 407)	422 (406, 439)	411 (395, 427)	452 (434, 470)	**<0.0001**
	Ref.	**+8**	**+5**	**+16**	
Proline[3]	245 (233, 258)	255 (243, 268)	265 (253, 278)	244 (233, 256)	0.06
	Ref.	**+4**	**+8**	**-1**	
Serine	191 (183, 198)	201 (194, 208)	197 (190, 204)	197 (190, 205)	0.3
	Ref.	**+5**	**+3**	**+3**	
Tyrosine	77 (74, 81)	83 (80, 86)	83 (80, 86)	73 (70, 76)	**<0.0001**
	Ref.	**+7**	**+7**	**-5**	
Non-standard amino acids					
Citrulline	37 (35, 39)	37 (36, 39)	41 (39, 43)	40 (38, 42)	0.02
	Ref.	**+1**	**+10**	**+8**	
Ornithine[3]	182 (174, 191)	195 (187, 203)	191 (184, 200)	205 (197, 215)	0.005
	Ref.	**+7**	**+5**	**+13**	

[1] Adjusted for age, BMI, smoking status, alcohol intake, time since last food or drink at blood collection and time between blood collection and processing

[2] p-values refer to p_{diff} across all four diet groups calculated by ANOVA. Conventional p-values are shown and those marked in bold were significant after Bonferroni correction ($p < 0.0024$)

[3] Missing for some participants: lysine n =11; Phenylalanine n = 93; tryptophan n = 2; glutamine n = 2; proline n = 45; and ornithine n = 1

[4] Non-transformed data were used and arithmetic means are displayed

Table 4.6: Spearman's rank correlations between dietary intakes and plasma concentrations of amino acids

	r	p[1]
Branched-chain essential amino acids		
Isoleucine	0.10	0.05
Leucine	0.16	**0.001**
Valine	0.14	0.01
Other essential amino acids		
Histidine	-0.02	0.7
Lysine[2]	0.21	**<0.0001**
Methionine	0.19	**0.0002**
Phenylalanine[2]	0.05	0.4
Threonine	0.06	0.2
Tryptophan[2]	0.16	**0.001**
Non-essential amino acids		
Alanine	-0.05	0.4
Arginine	0.09	0.07
Aspartate	0.004	0.9
Glutamate	0.03	0.6
Glycine	-0.15	0.003
Proline[2]	0.12	0.02
Serine	0.04	0.4
Tyrosine	0.17	**0.0007**

[1] Conventional p-values are shown and those marked in bold were significant after Bonferroni correction ($p < 0.0029$)

[2] Missing information for some participants: lysine n = 11; phenylalanine n = 93; tryptophan n = 2; and proline n = 45

Table 4.7: Spearman's rank correlations between intake of protein (% of energy) from food types and plasma amino acid concentrations[1]

	Non-soya plant		Soya		Meat[2]		Fish[3]		Dairy[4]		Eggs[4]		Mixed animal[4]	
	r	p	r	p[5]	r	p[5]	r	p[5]	r	p[5]	r	p[5]	r	p[5]
Branched-chain essential amino acids														
Isoleucine	-0.10	0.04	-0.13	0.01	-0.03	0.8	-0.13	0.07	0.18	0.002	0.06	0.3	-0.01	0.9
Leucine	-0.15	0.003	-0.18	0.0004	-0.03	0.8	-0.11	0.1	0.20	0.0006	0.09	0.1	0.01	0.9
Valine	-0.14	0.005	-0.19	**0.0001**	0.13	0.2	-0.06	0.4	0.15	0.01	0.18	0.002	-0.02	0.8
Other essential amino acids														
Histidine	0.01	0.8	0.02	0.7	-0.13	0.2	-0.14	0.05	-0.01	0.9	0.06	0.3	-0.05	0.4
Lysine	-0.21	**<0.0001**	-0.21	**<0.0001**	0.06	0.5	0.00	1.0	0.15	0.01	0.16	0.008	-0.06	0.3
Methionine	-0.13	0.008	-0.19	**0.0002**	0.02	0.8	-0.02	0.8	0.18	0.002	0.04	0.5	-0.08	0.2
Phenylalanine	-0.01	0.8	-0.03	0.6	-0.06	0.6	-0.08	0.3	-0.01	0.9	0.15	0.03	-0.03	0.7
Threonine	-0.02	0.7	-0.03	0.5	-0.04	0.7	-0.04	0.6	0.06	0.3	0.12	0.04	0.06	0.3
Tryptophan	-0.10	0.05	-0.16	0.001	-0.03	0.8	0.03	0.7	0.09	0.1	0.11	0.06	-0.04	0.5
Non-essential amino acids														
Alanine	0.04	0.4	0.10	0.04	-0.09	0.4	-0.06	0.4	0.04	0.5	0.09	0.1	-0.02	0.7
Arginine	-0.06	0.3	-0.11	0.04	0.05	0.6	0.04	0.6	0.04	0.5	-0.12	0.03	-0.10	0.1
Asparagine	0.08	0.1	0.07	0.2	-0.15	0.1	-0.03	0.7	-0.02	0.8	0.11	0.07	0.08	0.1
Aspartate	0.06	0.2	0.03	0.6	0.03	0.8	0.10	0.1	0.10	0.08	0.14	0.02	0.04	0.5
Glutamate	-0.06	0.2	0.02	0.7	-0.12	0.2	0.02	0.8	0.05	0.4	0.20	0.0005	0.10	0.09
Glutamine	0.09	0.07	0.07	0.2	0.06	0.6	-0.01	0.9	-0.04	0.6	-0.09	0.1	-0.12	0.04
Glycine	0.14	0.005	0.26	**<0.0001**	-0.15	0.1	0.01	0.9	-0.02	0.8	0.18	0.003	0.09	0.1
Proline	0.01	0.9	-0.09	0.09	-0.04	0.7	-0.20	0.01	0.03	0.6	0.08	0.2	0.05	0.4
Serine	0.05	0.4	0.04	0.5	0.05	0.6	0.03	0.7	0.06	0.3	0.05	0.4	0.06	0.3
Tyrosine	-0.19	**0.0002**	-0.19	**0.0002**	0.02	0.9	-0.04	0.6	0.14	0.02	0.11	0.07	-0.07	0.3
Non-standard amino acids														
Citrulline	0.00	0.9	0.06	0.2	-0.08	0.4	-0.08	0.3	0.16	0.005	-0.14	0.02	0.04	0.5
Ornithine	0.12	0.02	0.15	0.004	-0.15	0.1	-0.13	0.07	-0.04	0.4	0.14	0.02	0.01	0.9

[1] Foods were categorised into food types based on the main protein source in the food

[2] In meat-eaters only

[3] In meat-eaters and fish-eaters only

[4] Excluding vegans

[5] Conventional p-values are shown and those marked in bold were significant after Bonferroni correction ($p < 0.00034$)

4.8 FIGURES

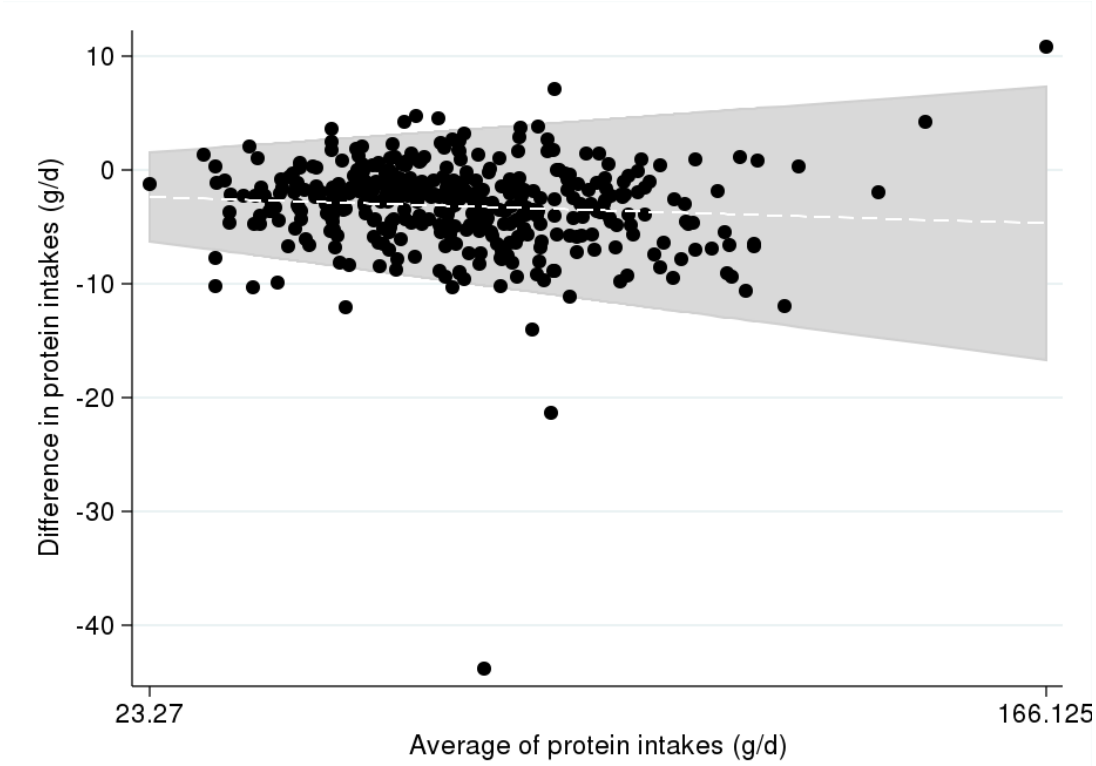


Figure 4.1: Agreement between estimated protein intakes using the UK and US nutrient tables
The difference in protein intake (shown on the y-axis) was calculated as protein intake estimated using the UK nutrient tables minus intake estimated using the US nutrient tables

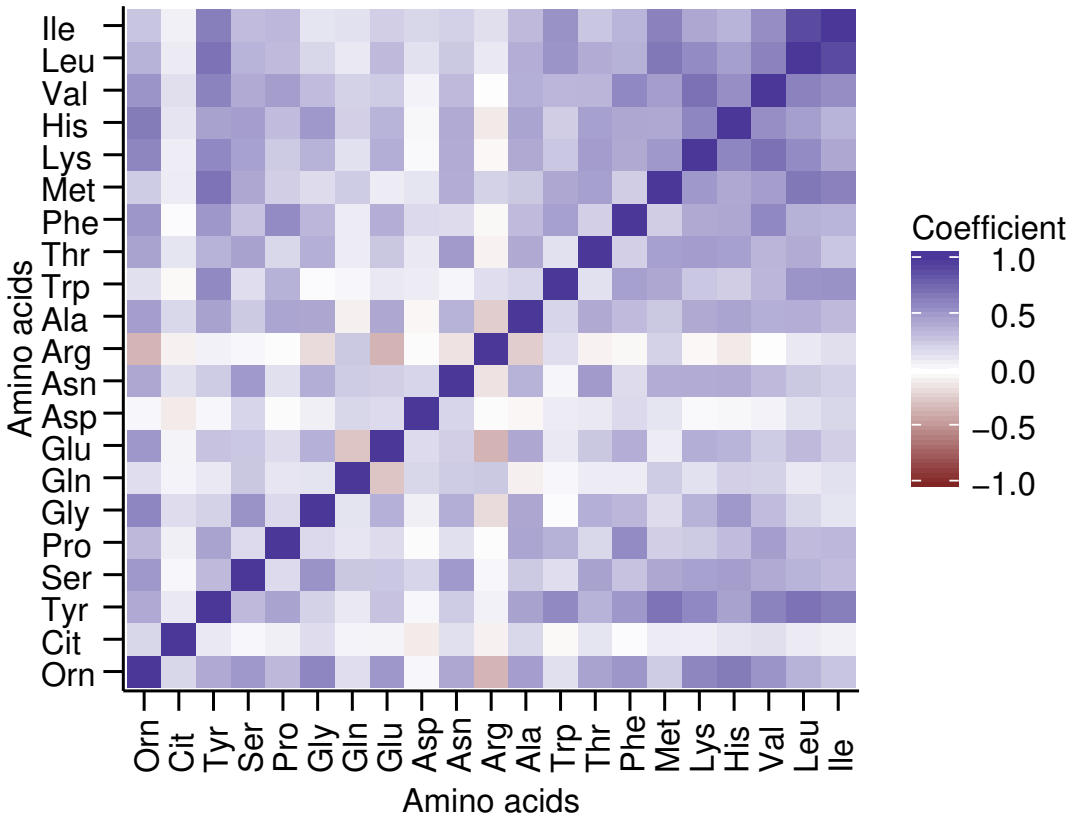


Figure 4.2: Correlation matrix for plasma amino acid concentrations

The colours indicate the strength and direction of the Spearman's rank correlation coefficients between pairs of plasma amino acid concentrations. Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartate; Cit, citrulline; Gln, glutamine; Glu, glutamate; Gly, glycine; His, histidine; Ile, isoleucine; leu, leucine; Lys, lysine; Met, methionine; Orn, ornithine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.

5. Metabolite concentrations by diet group

5.1 Introduction

Human metabolism is influenced by dietary, lifestyle, environmental and genetic factors [13, 14], and individuals with different dietary habits, e.g. meat-eaters and vegans, might therefore have different metabolite profiles in blood.

Metabolomics is the identification and quantification of metabolites, i.e. low molecular weight reactants, intermediates or products of biochemical reactions, in a biological system [12, 13]. In recent years metabolomics has been introduced into epidemiological research as technological advances have made assays of large number of samples feasible, in terms of laboratory time and costs. Epidemiological research using metabolomics aims to identify novel risk factors for disease and biomarkers of exposures, e.g. diet, as well as improving our understanding of disease mechanisms [10].

5.1.1 Previous studies

Besides the paper published on the work described in this chapter [18], only two smaller ($n \leq 42$ in each group) studies have been published on metabolite profiles in individuals from different habitual diet groups with respect to intake of food from animal sources, i.e. meat-eaters and vegetarians or vegetarians [151, 152]. Xu et al. found a clear separation between omnivores and lacto-vegetarians in urinary metabolite profiles, which in men was explained by lower concentrations of N-acetyl glycoprotein, trimethylamine-N-oxide, taurine, glucose, phenylalanine, methylhistidine and formate, and by higher concentrations of citrate, glycine and hippurate in vegetarians than in omnivores [151]. Wu et al. reported that 92 of 361 plasma metabolites differed between vegans and omnivores after correction for multiple testing, with higher concentrations of several lipids and amino acids in omnivores [152].

In addition to being a relevant research topic in its own right, studying metabolite profiles of men from different diet groups is of relevance to this thesis as there is some evidence to suggest that vegans have a lower risk of prostate cancer than meat-eaters [130, 131].

5.2 Aim

The aim of the current study was to investigate differences in metabolite profiles between male meat-eaters, fish-eaters, vegetarians and vegans from the Oxford component of the European Prospective Investigation into Cancer and Nutrition (EPIC) by analysing circulating concentra-

tions of individual metabolites and patterns in metabolite profiles. The metabolites investigated were acylcarnitines, amino acids, biogenic amines, glycerophospholipids, hexose and sphingolipids, related to lipid, protein and glucose metabolism. The present study will add to this existing literature by investigating the metabolite profile in plasma of four distinct habitual diet groups.

5.3 Subjects and methods

5.3.1 Study population and data collection

From 1993 to 2001, 65,000 men and women older than 20 years of age were recruited from across the UK into the EPIC-Oxford cohort [100]. This cohort aims to investigate associations between diet, lifestyle and cancer risk in individuals with different long-term dietary habits; thus a large number of vegetarians and vegans were recruited. Participants were mostly recruited via post (89%) or via their general practitioner's surgery. All participants gave written informed consent and the protocol for EPIC-Oxford was approved by a multi-centre research ethics committee (MREC/02/0/90).

5.3.1.1 Eligibility criteria

The study population for the current chapter is the same as in the previous Chapter 4 (see Section 4.3.1.1 on page 34). In brief, healthy men aged 30 to 49 years, who filled in the recruitment questionnaire with sufficient detail were eligible for the analysis; 98 eligible men in each diet group were selected.

5.3.2 Blood collection and laboratory analysis

At recruitment, participants were asked to have a blood sample taken at their local general practitioner's surgery; participants were not required to fast and time since last food or drink was recorded. Blood samples were sent at ambient temperature to the laboratory, where they were processed for long-term storage in liquid nitrogen (−196°C) until 2011 and subsequently in electric freezers (−80°C).

5.3.2.1 Metabolomics assay

Plasma samples (with citrate as an anticoagulant) were transported on dry ice to the International Agency for Research on Cancer, Lyon, France. Here the targeted metabolomic assay Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbrcuk, Austria) was used to quantify a total of 145 metabolites; amino acids ($n = 21$) and biogenic amines ($n = 9$) were separated by high performance liquid chromatography (HPLC) before injection into the mass spectrometer

(QTRAP® 5500, AB Sciex, Framingham, Massachusetts, USA), while flow injection analysis was used for acylcarnitines ($n = 18$), glycerophospholipids ($n = 82$), hexose ($n = 1$) and sphingolipids ($n = 14$).

Samples from the four diet groups were randomly distributed between analytic batches, and quality control samples were included in each batch (see section 5.3.2.2 for details on the quality control samples).

Please refer to Section 4.3.3 on page 36 for further laboratory details and to Section 3.5.2 on page 28 for metabolite nomenclature.

5.3.2.2 Quality of metabolomics data

Analytical reproducibility: In each assay batch, samples of pooled plasma were included for quality control (four to six in each batch). To assess the inter-assay reproducibility, overall coefficients of variation (CV) across all pooled plasma samples ($n = 26$) were calculated for each metabolite. The vast majority of metabolites (87.7%) had a CV between 5 and 15%, while a few (3.6%) had a CV larger than 20% (Table 5.1 on page 74). The median CV across all metabolites was 8.2%; it was 6.9% for acylcarnitines, 12.0% for amino acids, 9.8% for biogenic amines, 7.8% for glycerophospholipids, 5.9% for hexose and 7.7% for sphingolipid.

Non-quantifiable results: For 104 of the 145 metabolites (71.7%) the concentration could be quantified in samples from all participants by the assay. However, for 41 metabolites some participants had concentrations outside the measurable range ($n = 36$) and/or missing data ($n = 8$); for most of these metabolites, only a minor percentage of the samples were affected (Table 5.2 on page 74). However, for 10 metabolites more than 50% of samples were outside the measurable range.

5.3.2.3 Measurement of cholesterol and apolipoprotein concentrations

Serum concentrations of cholesterol and apolipoproteins have been measured for a previous analysis [153]. In brief, apolipoprotein A-I and B were measured by immunoturbimetric assay, high density lipoprotein (HDL) cholesterol was measured directly and total cholesterol was measured using the enzymatic assay (Beckman Synchron CX autoanalyzers, Beckman Coulter, High Wycombe, UK for all). The coefficients of variation calculated from 196 pooled serum samples ranged between 1.9 and 2.8% for the analytes.

5.3.3 Exclusions

Metabolites with more than 10% of samples outside the measurable range ($n = 21$; Table 5.2 on page 74) and with CV higher than 20% ($n = 5$; Table 5.1 on page 74) were excluded, to limit the risk of results being affected by unknown values and analytical variation.

Participants for whom data on one or more metabolites were missing (203 men) would have dropped out of the principal component analysis, resulting in loss of power. On the other hand, excluding metabolites with any missing data (8 metabolites) might have led to loss of important metabolites. By balancing the number of excluded men and excluded metabolites (Table 5.3 on page 74), it was decided to exclude metabolites with missing data for more than 5% of participants (4 metabolites and 13 men).

In total, the exclusions resulted in 27 metabolites (Figure 5.1 on page 105) and 13 men being excluded. This left complete data on 118 metabolites concentrations in 379 men, i.e. 95 meat-eaters, 97 fish-eaters, 91 vegetarians and 96 vegans, for further analyses.

5.3.4 Dietary and anthropometric data

At recruitment, participants filled in a validated semi-quantitative food frequency questionnaire (FFQ) [107, 108], including questions on 130 foods and drinks, 113 of which were relevant to vegetarians and vegans. Mean daily intakes were estimated using specified portion sizes [109] and mean daily nutrient intakes were estimated mostly using the fifth edition of 'McCance and Widdowson's The Composition of Foods' and its supplements [110, 111, 112, 113, 114, 115, 116, 117, 118, 119].

Participants were further asked if they ate meat, fish, dairy or eggs and the answers were used to categorise participants as meat-eaters (who eat meat), fish-eaters (who eat fish but not meat), vegetarians (who do not eat meat or fish) and vegans (who do not eat any of the foods).

Height and weight were used to calculate body mass index (BMI; kg/m^2). In addition to self-reported measurement, height and weight were also measured in a sub-sample of the cohort. Self-report and measured values showed good agreement (correlation coefficient (r) > 0.9) [122], and measured BMI was used if available.

5.3.5 Statistical analysis

5.3.5.1 Preparation of metabolomics data

For 13 of the included metabolites, some men ($< 10\%$ as defined by the exclusion criteria) had measurements outside the measurable range. Measurements below the limit of detection were set to half the lowest measured concentrations (ranging from 0.015 to 7.753 μmol/L; applicable to seven acylcarnitines for one to 15 men). Measurements below the limit of quantification (0.1

μmol/L for serotonin only, applicable to 10 men) were set to half the limit of quantification. Finally, measurements above the highest concentration calibration standards (400 μmol/L for leucine, histidine, threonine and ornithine, and 800 μmol/L for glutamate; applicable to one to two men) were set to the highest standards.

All metabolite concentrations were logarithmically transformed for all analyses to approximate the normal distribution.

5.3.5.2 Covariates by diet group

Participant characteristics, nutrient intakes and blood sample characteristics were compared across diet groups; Kruskal-Wallis one-way analysis of variance (ANOVA) and χ^2 tests were used for continuous and categorical variables, respectively (except for hemolysis for which Fisher's exact test was used due to low numbers).

5.3.5.3 Metabolite concentration by covariates

Geometric means and 95% confidence intervals (CI) of metabolite concentrations were computed for categories of personal characteristics and factors related to the blood samples. Differences were tested using ANOVA adjusted for age ($30-34; 35-39; 40-44; 45-49$ years) and BMI ($< 22.5; 22.5-24.9; \geq 25$ kg/m^2; unknown).

Multiple testing: To account for multiple comparisons in this analysis and other analyses of individual metabolite concentrations, a false discovery rate (FDR)-controlling procedure – the Benjamini–Hochberg method [154] – was used. First, the p-values were sorted and ranked from the lowest $p(1)$ to the highest $p(m)$. Let k be the largest rank $p(i)$ for which $p(i) < (i/m) \times \alpha$ is true; α is the significance level and set to 0.05 in all analyses. Then all the null hypotheses for p-values from $p(1)$ to $p(k)$ were rejected [154]. This method allows 5% of positive findings to be false, on average [140].

5.3.5.4 Analysis of individual metabolite concentrations by diet group

Differences in geometric mean concentrations of individual metabolites across diet groups were tested using ANOVA. To facilitate easier comparison between results for the metabolites, which vary hugely in concentration (ranging from hundredths of a μmol/L to thousands of μmol/L for C14:1 and hexose, respectively), results were additionally displayed as percent mean difference in metabolite concentration with meat-eaters as the reference group. The main model was adjusted for *a priori* selected personal and blood sample characteristic, i.e. age (categorised as above), BMI (categorised as above), smoking status (never; former; current), alcohol intake

($< 1; 1-7; 8-15; \geq 16$ g/d), time since last food or drink at blood collection ($< 1.5; 1.5-2.9; 3.0-4.4; \geq 4.5$ hours; unknown), and time between blood collection and processing (fourths of the distribution corresponding to $< 25; 25-41; 41-72; \geq 72$ hours; unknown).

Further adjustments were made for other covariates associated with metabolite concentrations, i.e. energy intake (fourths of the distribution, corresponding to $< 7,205; 7,206-8,632; 8,633-10,335; \geq 10,336$ kJ/d), time of day at blood collection (00:00 − 09:59; 10:00 − 12:59; 13:00 − 23:59; unknown), hemolysis (no; yes) and lipemic status (no; yes).

Statistical power: To provide an estimate of the power in the current study, two power calculations were performed; the significance level was set to 0.05 and 0.0004 (i.e. $118/118 \times 0.05$ and $1/118 \times 0.05$), respectively. This approach was taken to illustrate the power in the extreme cut-offs for significance when applying the FDR-controlling procedure. With 91 to 97 men in each diet group, we had more than 80% power to detect a difference in geometric mean concentration of approximately 6 to 20% between the diet groups with the highest and the lowest concentration.

Concentrations of metabolites, cholesterol and apolipoproteins: Phosphatidylcholines (i.e. the type of glycerophospholipid investigated in this thesis) are involved in lipoprotein synthesis [155], and cholesterol and apolipoprotein concentrations could therefore be associated with these metabolites. To explore this, Spearman's rank correlation coefficients were calculated for metabolite concentrations in relation to total, HDL and non-HDL cholesterol; apolipoproteins A1 and B; and the ratios total:HDL cholesterol and apolipoprotein B:A-I. Partial correlations adjusted for age and BMI (both categorised as above) showed similar results, and thus, the crude results from Spearman's correlations are presented.

The results of this analysis led to further adjustments of the main analysis of metabolite concentration by diet group for concentrations of total cholesterol (fifths of the distribution, corresponding to $< 4.0; 4.0-4.5; 4.5-5.0; 5.0-5.7; \geq 5.7$ mmol/L; unknown).

Correlation between metabolites: Spearman's rank correlation coefficients were calculated between pairs of metabolites.

5.3.5.5 Analysis of metabolite patterns using principal component analysis

Dimension reduction methods are often applied to metabolomics data in order to overcome the limitations of univariate analysis, i.e. multiple testing, to overlook correlations between metabolites and not to evaluate associations involving several metabolites in combination [140]. Principal component analysis (PCA) is one such method, commonly used in case-control studies

of metabolite profile and prostate cancer, for example [11]. This method summarises the original data by identifying fewer new variables, which explain as much of the total variance in the original data as possible [156]. These new variables are called principal components and describe latent patterns in the original data. The first principal component explains the most variance in the data; the next explains the most variance not accounted for by the first principal component and so forth; the principal components are uncorrelated [156].

The number of principal components to retain for further analysis can be chosen based on several criteria. These include (but are not limited to): a cut-off for the cumulative amount of variance explained and a visual examination of the scree plot [156]. A cut-off of 70% of the total variance has been suggested, but a lower cut-off may be reasonable when the numbers of variables and observations are large [156], like in the current analysis. In a scree plot, the eigenvalues (i.e. the amount of variance each principal component explains [157]) are plotted against the number of principal components. The number of factors to retain is chosen at the point where the graph flattens [156], because little extra variance is accounted for by retaining more components.

The influence of the original variables on each principal component is given by the loading [158], which is the correlation between the metabolite and the principal component [157]. Scores are assigned to each participant for each principal component; the score is a linear combination of participant's concentration of the metabolites weighted by the loadings [157]. Score plots (scatter plots of participants' scores) of the retained principal components are used to visualise clustering within and separation between groups [158], e.g. diet group. Loading plots (scatter plots of the loadings) show the influence of the original metabolites on principal components. The direction of the score and loading plots corresponds, and the loading plots can therefore be used to identify metabolites responsible for the separation between groups seen in the score plots [158].

A PCA model can be constructed in different ways and there is currently no consensus on which is more appropriate. For this chapter two models, previously used in EPIC [147, 159], were computed. Firstly, the main model was computed using the covariance matrix of logarithmically transformed metabolite concentrations [159]. The choice of numbers of components to retain was based both on a visual examination of the scree plot and on the percentage of variation explained; It was decided to capture approximately 60% of the total variation. Secondly, the alternative model was based on the correlation matrix of logarithmically transformed metabolite concentrations which were z-score standardised (mean of 0, standard deviation of 1) [147]. The choice of numbers of components to retain was done as for the main model, although it was decided to capture approximately 75% of the total variation. After selecting the number of components for the alternative model, varimax rotation was applied [147].

The main and alternative models differ with regards to use of standardisation of metabolite concentrations, choice of matrix and use of rotation. The model using standardised metabolite concentrations and the correlation matrix (the alternative model) gives equal weight to all metabolites, as the variance of all metabolites are set to 1 and because a PCA based on the correlation matrix does not depend on the scale on which the original variables are measured [156]. In contrast, using the covariance matrix, which does take the scale into account, on untransformed data would give more weight to metabolites with the greatest absolute variance. Such metabolites would therefore completely dominate the first number of principal components. Using the covariance matrix of log-transformed metabolite concentrations, as in the main model, will to some extent even out the difference in variance between metabolites. Rotation can be applied to simplify interpretation of the components, because rotation draws the loadings closer to -1, 0 or 1 [156]. However, rotation has little theoretical support [160], and the specific method and the number of components retained prior to rotation can greatly affect the characteristics of the resulting components. Based on these considerations the covariance matrix of log-transformed data without rotation was chosen as the main model.

5.3.5.6 Principal component partial R-square method

The principal component partial R-square (PC-PR2) method quantifies how much of the variability in metabolite data is explained by a predefined list of characteristic, e.g. personal, lifestyle and analytic factors [161]. This method combines PCA, multivariate linear regression and $R_{partial}{}^2$ statistics. First, PCA is run on the metabolite concentrations and principal components, which cumulatively explain 80% of the total variance, are retained. Second, multiple linear regression models are fitted with the variance of the principal components as the dependent variables and the characteristics as the independent variables (e.g. diet group and covariates). Third, the $R_{partial}{}^2$ is calculated for all independent variables separately for each principal component. Finally, for each independent variable, a weighted average of the $R_{partial}{}^2$ is calculated using the eigenvalue of the principal components as the weight. These overall $R_{partial}{}^2$ values represent the amount of variability in the metabolite concentrations that is explained by each independent variable (e.g. diet group) when adjusting for the other independent variables (e.g. covariates) in the model.

In the current analysis, the PC-PR2 model included diet group, age, BMI, smoking status, alcohol intake, time since last food or drink at blood collection and time between blood sampling and processing (all categorised as above) as the independent variables.

5.3.5.7 Statistical software

All analyses were performed in Stata Statistical Software version 13 (StataCorp LP, Texas, United States), except the PC-PR2 analysis which was done in R: A language and environment for statistical computing (Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/)

5.4 Results

5.4.1 Participant characteristics, nutrient intake and blood sample characteristics by diet group

Body mass index varied by diet group, with the highest BMI being observed in meat-eaters, followed by fish-eaters and vegetarians while vegans had the lowest BMI (Table 5.4 on page 75).

The intake of energy and nutrients also differed by diet group. Meat-eaters had the highest energy intake, followed by vegetarians, then fish-eaters and finally vegans. The intake of energy from protein was also highest in meat-eaters, followed by fish-eaters, then vegetarians and lowest in vegans, whereas the opposite was seen for carbohydrate intake; vegans had the highest intake, followed by vegetarians, fish-eaters and meat-eaters with the lowest intake. Overall, intake of energy from fat did not vary by diet group, but differences were observed for sub-types of fat. The intake of energy from saturated fat was highest in meat-eaters, followed by vegetarians and fish-eaters, and was lowest in vegans. For monounsaturated fat, meat-eaters also had the highest intake but were followed by fish-eaters, vegans and then vegetarians, whereas intake of energy from polyunsaturated fat was higher the more animal products were excluded from the diet. The intake of alcohol was highest in fish-eaters and vegetarians, followed by meat-eaters and lowest in vegans.

Of the factors related to blood collection and handling, an association was seen only for time between blood collection and processing, with the process delay being shorter in meat-eaters than in the other groups. Furthermore, a borderline statistically significant association was seen for time since last meal or drink at blood collection, with vegetarians and vegans having longer time than fish-eaters and especially than meat-eaters.

5.4.2 Metabolite concentrations by covariates

Metabolites significantly associated with participant and blood sample characteristics after allowing for multiple testing are seen in Table 5.5 on page 76. BMI, energy and alcohol intake were each related to concentrations of several glycerophospholipids (BMI $n = 28$; others $n = 26$), a few sphingolipids (BMI $n = 3$; energy $n = 6$; alcohol $n = 6$), certain amino acids (BMI $n = 7$;

others $n = 1$) and few acylcarnitines (BMI $n = 2$; others $n = 1$). Additionally, BMI and energy intake were associated with kynurenine concentration and BMI was associated with concentrations of creatinine and hexose. Men with higher BMI had higher concentrations of 31 of these metabolites and lower concentrations of 10, while half the metabolites concentrations increased with increasing energy or alcohol intakes.

Concentrations of histidine, serine, creatinine, 12 glycerophospholipids and SM (OH) C14:1 varied according to age-group, with few showing a consistent direction in the association.

Time since last food or drink, and time of day at blood collection were associated with some amino acids ($n = 5$ for each) and one biogenic amine (kynurenine). Time of day at blood collection was additionally associated with one glycerophospholipid. As expected, concentrations of most of these metabolites were lower the longer the participants had fasted before blood collection. The time between blood collection and processing was associated with some acylcarnitines ($n = 4$), several amino acids ($n = 16$), a few glycerophospholipids ($n = 7$) and hexose; concentrations tended to increase with longer process delay. Lipemic samples had different concentrations of some amino acids ($n = 3$), glycerophospholipids ($n = 23$) and sphingolipids ($n = 2$) than non-lipemic samples, while hemolysis was related to some amino acids ($n = 6$), two glycerophospholipid, hexose and two sphingolipids. For most of these metabolites, concentrations were higher in lipemic and hemolytic samples, respectively, than samples which were not affected.

Smoking status, physical activity and use of medication or supplements on the day of blood collection were not associated with metabolite concentrations.

5.4.3 Individual metabolite concentration by diet group

5.4.3.1 Main model

Overall, mean concentrations of 93 out of 118 metabolites (78.8%) differed significantly by diet group after allowing for multiple testing (Figure 5.2 on page 106 and Table 5.6 on page 81). The strongest and the highest proportions of differences were seen for the sphingolipids (85.7% differed) and glycerophospholipids (84.7% differed).

Acylcarnitines: Six of nine acylcarnitines varied by diet group; C0, C4 and C5 concentrations were highest in meat-eaters, followed by fish-eaters and vegetarians, and lowest in vegans. Similarly, C3 and C16 were highest in meat-eaters and lowest in vegans. In contrast, vegans had the highest concentration of C18:2, followed by fish-eaters and vegetarians, and lowest in meat-eaters. This difference in C18:2, with concentrations being 45% higher in vegans than meat-eaters, represents the largest difference between vegans and meat-eaters for those metabolites which were highest in vegans.

Amino acids: Of the 19 amino acids, 12 varied varied by diet group. Concentrations of leucine, valine, lysine, methionine, tryptophan, and tyrosine were in general highest in fish-eaters or vegetarians, followed by meat-eaters and lowest in vegans, while concentration of citrulline was higher in vegetarians and vegans than in meat-eaters or fish-eaters and concentrations of glycine and ornithine were highest in vegans. Alanine concentration was lower in meat-eaters than in the other three diet-groups, while glutamate was highest in fish-eaters, followed by vegetarians and lowest in the meat-eaters, and glutamine was lowest in the fish-eaters.

Biogenic amines: The biogenic amine creatinine concentration was highest in meat-eaters, followed by fish-eaters and vegetarians and lowest in vegans, while kynurenine was highest in vegetarians, followed by fish-eaters, then meat-eaters and lowest in vegans.

Glycerophospholipids: Of the 72 glycerophospholipids, 61 differed across diet groups; most often, vegans had the lowest concentrations. For 23 glycerophospholipids, meat-eaters had the highest concentrations, followed by fish-eaters, then vegetarians while the vegans had the lowest concentrations. This pattern was found for PC aa C36:6 and the difference of 64% between meat-eaters and vegans was the largest observed among all the metabolites. For five glycerophospholipids, the lowest concentration was observed in vegans, while concentrations were similarly higher in the three other diet groups. The remaining 33 glycerophospholipids showed different patterns; the concentrations of 28 of these were also lower in vegans than in meat-eaters while for the remaining five, vegans had higher concentrations than meat-eaters.

Sphingolipids: Concentrations of 12 of the 14 sphingolipids differed across diet group. For all except three, meat-eaters had the highest concentrations followed by fish-eaters, then vegetarians and finally vegans. Similar patterns to the above were seen for two additional sphingolipids (SM C20:2 and SM C24:1), although concentrations were similar in fish-eaters and vegetarians for the former, and similar in vegetarians and vegans in the latter. For SM (OH) C24:1 the mean concentration was similar in meat-eaters, fish-eaters and vegetarians while vegans had the lowest concentration.

Hexose: No difference was seen in hexose by diet group.

5.4.3.2 Further models

Analysis of vegans and non-vegans: With the results from analyses of metabolite concentrations across the four diet groups indicating that metabolite concentrations in vegans in particular may differ from those in the other diet groups ('non-vegans'), we conducted additional

analyses, comparing (i) metabolite concentrations in vegans versus the other three diet groups combined, and (ii) metabolite concentrations across the three non-vegan diet groups.

Mean concentrations of 86 of the 118 metabolites (72.9%) differed significantly between the non-vegans and vegans (Table 5.7 on page 89) and, as for the analysis across all four diet groups, the strongest and the highest proportions of differences were seen for the sphingolipids (85.7% differed) and glycerophospholipids (79.1% differed).

Fewer differences were observed between the three non-vegan diet groups; 39 of the 118 metabolite concentrations (33.1%) differed significantly between meat-eaters, fish-eaters and vegetarians (Table 5.8 on page 95). The majority of these metabolites were glycerophospholipids and sphingolipids (29 glycerophospholipids, 6 sphingolipids, 3 amino acids and 1 biogenic amine), and the most commonly observed pattern was for meat-eaters to have the highest concentrations, followed by fish-eaters and then vegetarians. However, compared to meat-eaters, fish-eaters and vegetarians had higher concentrations of alanine and glutamine, while citrulline concentration was higher in vegetarians than in meat-eaters and fish-eaters.

Further adjustments: Results for the main model further adjusted for energy intake, time of day at blood collection, hemolysis and lipemic status were similar to the main results (Table 5.6 on page 81; only p-values are shown – see the column denoted $p_{diff}{}^{1,3}$). The only differences were that glutamine and PC aa C36:4 were no longer significant after further adjustments.

Cholesterols and apolipoproteins: Total cholesterol was associated with 97 metabolites (82.2%) after controlling the FDR across all 826 tests (7 cholesterol species $\times$ 118 metabolites), the majority of which were positive; the strongest correlations were seen with glycerophospholipids ($r \leq 0.70$) and sphingomyelins ($r \leq 0.78$; Figure 5.3 on page 107). Results for non-HDL cholesterol ($r \leq 0.64$ for glycerophospholipids and ≤ 0.73 sphingolipids) and apolipoprotein B ($r \leq 0.57$ for glycerophospholipids and ≤ 0.70 for sphingolipid, respectively) were similar, although slightly weaker, to those for total cholesterol. More modest, positive correlations were seen for the anti-atherosclerotic particles HDL cholesterol and apolipoprotein A-I, with glycerophospholipids ($r \leq 0.49$ and ≤ 0.47, respectively) and sphingolipids ($r \leq 0.36$ and ≤ 0.35, respectively). The results for the ratios between total and HDL cholesterol, and between apolipoprotein B and A-I were similar, with modest positive correlations with some glycerophospholipids ($r \leq 0.48$ and ≤ 0.39, respectively) and sphingolipid ($r \leq 0.39$ and ≤ 0.46, respectively) as well as negative associations with a few amino acids, lysophosphatidylcholines and acyl-alkyl phosphatidylcholines (PC ae C$x{:}y$) with long fatty acids ($r \geq -0.37$ for all).

As total cholesterol showed the strongest correlation with metabolites, the main model of metabolite concentration by diet group was further adjusted for this variable. This did not

materially change the results for acylcarnitines, amino acids, biogenic amines and hexose, although results for the amino acids valine and citrulline were no longer significant while results for histidine and aspartate became significant after further adjustments (Table 5.6 on page 81; only p-values are shown – see the column denoted $p_{diff}{}^{1,4}$). As expected, taking total cholesterol into account had a bigger impact on the results for glycerophospholipids and sphingolipid, although major changes were not seen. However, the differences across diet groups were no longer statistically significant after adjustment for total cholesterol for six glycerophospholipids and sphingolipids (lysoPC a C16:1, PC aa C32:3, PC aa C36:1, PC aa C38:3, PC aa C42:5, PC ae C44:5 and SM C16:0). In contrast, results for three glycerophospholipids and two sphingolipids (lysoPC a C18:1, PC aa C40:5, PC ae C42:1, SM C16:1 and SM C24:0) became significant only after this adjustment. Overall, 92 (78.0%) metabolites showed significant differences across the diet groups after adjustment for total cholesterol compared to 93 (78.8%) in the main model, after taking multiple testing into account.

5.4.4 Correlations between metabolites

The metabolites were highly correlated; 70.4% of the correlations were significant after allowing for multiple testing (Figure 5.4 on page 108). Most correlations were positive, and the strongest correlations were seen within metabolite classes, especially within the glycerophospholipids and sphingolipids ($r \leq 0.97$). Strong, positive correlations were also seen between metabolites from these two classes. Only a few negative correlations were observed ($r \geq -0.57$); these were mainly for hexose and arginine but also for some acylcarnitines and amino acids.

5.4.5 Patterns in metabolite profile by diet group

5.4.5.1 PCA based on covariance matrix of log-transformed metabolite concentrations – the main model

Four principal components were retained for further inspection based on visual examination of the scree plot (Figure 5.5 on page 109) and because together they explained 60.1% of the total variation. The first principal component explained 37.7% of the total variation and was characterised by positive loadings on some glycerophospholipids (including PC aa C32:1, PC aa C36:5, PC aa C36:6 and PC ae C34:0) and a few sphingolipids (including SM (OH) C14:1) (Figure 5.6 on page 110). Principal component 2 explained 10.6% of the total variance and it was primarily characterised by positive loadings on three acylcarnitines (C16, C18:1 and C18:2) and negative loadings on arginine and hexose. The third principal component explained 6.3% of the total variance and was characterised by positive loadings on some glycerophospholipids (including PC aa C32:1, PC aa C34:4 and PC aa C40:5) and negative loadings on a few sphingolipids (includ-

ing SM (OH) C16:1). Finally, principal component 4 explained 5.5% of the total variation and was mainly characterised by positive loadings on some glycerophospholipids (including PC ae C44:6).

The comparison of mean scores showed significant differences across the diet groups for all four principal components (Table 5.9 on page 103). The score plot of principal component 1 against 3 showed the best visual separation between diet groups (Figure 5.7 on page 111); the separation was mainly seen on principal component 1. For principal component 1, meat-eaters had the highest score followed by fish-eaters, then vegetarians, and then vegans who had the lowest score (31% lower than that of meat-eaters; Table 5.9 on page 103). In contrast, for principal components 2, 3 and 4, the mean score was in general highest in vegans and lowest in meat-eaters.

Adjustment for potential confounders did not materially change the results for principal component 1, 3 and 4, while differences across diet groups in scores for principal component 2 were attenuated and no longer significant (Table 5.9 on page 103).

The loading plot of principals component 1 and 3 revealed that the separation between the diet groups – especially the vegans from the other groups – was primarily explained by lower concentrations of some glycerophospholipids and two sphingolipids (especially PC ae C34:0, SM (OH) C14:1 and SM (OH) C16:1 but also PC aa C32:1, PC aa C36:5 and PC aa C36:6; Figure 5.8 on page 112).

5.4.5.2 PCA based on correlation matrix of standardised data with rotation – the alternative model

Ten principal components were retained for further inspection and they explained 73.1% of the total variation after rotation. As the first three principal components, which explained 33.6% of the total variation, showed the best separation between diet groups, only these are described here. The first principal component explained 13.8% of the total variation and was characterised by positive loadings on some glycerophospholipids (including lysoPC a C17:0, PC ae C34:2, PC ae C36:2, PC ae C38:3, PC ae C38:4 and PC ae C40:4) and sphingolipids (including SM C20:2; Figure 5.9 on page 113). Principal component 2 explained 10.6% of the total variance and it was characterised by positive loadings on some glycerophospholipids (including PC aa C36:5, PC aa C36:6, PC aa C38:0, PC aa C38:6, PC aa C40:6, PC ae C38:0, PC ae C38:6, PC ae C40:1 and PC ae C40:6). Finally, the third principal component explained 9.2% of the total variance and was characterised by positive loadings on some glycerophospholipids (including PC aa C30:0, PC aa C32:1, PC aa C32:2, PC aa C34:1, PC aa C34:3, PC aa C34:4, PC aa C36:1, PC aa C36:3 and PC aa C38:3).

The score plots showed separation between diet groups and differences in mean scores were

significantly different across diet groups (Figure 5.10 on page 114; Table 5.10 on page 104). Vegans had lower scores than the other three diet groups on all three principal components. Less separation was observed between the meat-eaters, fish-eaters and vegetarians, although fish-eaters and vegetarians had intermediate scores on principal components 1 and 2, and vegetarians had higher scores on the third principal component. Adjusting the ANOVA for potential confounders did not materially change the results from the unadjusted analysis (Table 5.10 on page 104).

The loading plots revealed that the separation between the diet groups was primarily explained by lower concentrations of some glycerophospholipids and a sphingolipid in vegans, and to a lesser extent in fish-eaters and vegetarians, than in meat-eaters (Figure 5.9 on page 113). Additionally, the vegetarians' slight separation from the other three diet groups was explained by higher concentrations of metabolites with positive loadings on components 1 and 3, and inverse loadings on component 2 (especially PC aa C36:3).

5.4.6 Variability in metabolomics data explained by diet group and other variables

Diet group, and personal and blood sample characteristics, i.e. age, BMI, smoking status, alcohol intake, time since last food or drink at blood collection, and time between blood collection and processing, explained 29.4% of the total variability in the metabolomics data (Figure 5.11 on page 115). The major contributor to the variability was diet group, explaining 18.6% of the total. Other variables explained minor proportions only; time between blood collection and processing explained 4.4%, followed by time since last food or drink, BMI, age, alcohol intake and smoking status (all $< 1.8\%$).

5.5 Discussion

5.5.1 Main findings

In this analysis of plasma metabolites by habitual diet group, significant differences in metabolite profiles were observed between meat-eaters, fish-eaters, vegetarians and especially vegans. This difference was primarily driven by lower concentrations of some glycerophospholipids and sphingolipids in vegans than in the other diet groups, although there were also differences in levels of other metabolites; concentrations of 12 sphingolipids (86%), 61 glycerophospholipids (79%), six acylcarnitines (67%), two biogenic amines (67%) and twelve amino acids (63%) varied across the four diet groups, while no association was seen for hexose. For the vast majority of these metabolites, vegans had the lowest concentrations while the highest was most often observed in meat-eaters.

5.5.1.1 Biological plausibility of findings

Our results of lower concentrations of glycerophospholipids and sphingolipids in vegans compared to the other diet groups, as well as the strong correlations between sphingolipids and glycerophospholipids, are consistent with synthesis pathways and dietary sources of these metabolites. Both phosphatidylcholine and sphingomyelin (i.e. the type of glycerophospholipids and sphingolipids in this study) are provided by the diet, mainly from animal products such as eggs, poultry and red meat, and by *de novo* synthesis [162, 163]. Of the two main pathways for *de novo* synthesis of phosphatidylcholines, one requires choline [155], mainly found in animal products but also in plant foods like wholegrains [163], and the other involves methylation of phosphatidylethanolamine, which is inhibited when methionine availability is low [155]. Vegans had the lowest methionine intake and concentration (Sections 4.4.3.1 on page 40 and 4.4.4.1 on page 40). Sphingomyelins are synthesised from phosphatidylcholines [164].

5.5.2 Other studies

Only two previous studies have investigated metabolite profiles of meat-eaters compared to those excluding some or all animal products from their diet. Both studies found clear differences between omnivores and lacto-vegetarians [151] and vegans [152], respectively. Of the ten urinary metabolites, which explained the differences between male omnivores and lacto-vegetarians [151] only glycine and glucose (as part of hexose) were also investigated in the current study, and differences in these metabolites did not explain differences in metabolite profiles between the diet groups in our study. However, glycine concentration did vary by diet group in the present analysis of individual metabolites, and in line with the previous study [151], meat-eaters had lower concentrations than the other diet groups. Consistent with the current study, the other previous study reported lower plasma concentrations of several lipids and amino acids (including isoleucine, leucine, and tryptophan) in vegans than omnivores [152].

Comparison of our and previous results, especially in the study by Xu et al. [151], is not straightforward due to the use of different populations, biological samples and analytical platforms. Xu et al. [151] studied a Chinese population consisting of military personnel (omnivores) and members of a Buddhist College and temple (lacto-vegetarians). These men are likely to have a very different diet, lifestyle and genetic makeup than the British men in the current study, and thereby different metabolite profiles [14, 165]. Moreover, different samples (urine versus plasma) [166] and assay methods (untargeted nuclear magnetic resonance (NMR) versus targeted mass spectrometry) were used.

Further studies support an association between habitual dietary intake and the metabolite concentrations. Metabolites measured in serum using untargeted mass spectrometry were asso-

ciated with intakes of several foods, including citrus fruits, meat, fish, butter, peanuts, coffee and alcoholic drinks, and the Healthy Eating Index [167], while metabolite concentrations measured in urine using ^{1}H NMR were associated with intakes of vegetables or red meat, and with dietary patterns [168].

Our results of higher circulating concentrations of creatinine (a biogenic amine) and five of the six acylcarnitines in meat-eaters than in vegans are broadly consistent with findings from feeding studies that aimed to identify biomarkers of meat intake [169, 170]. These studies have found higher concentrations of creatinine and acylcarnitines in urine, using a non-metabolomics assay or NMR, after experimental diets high in meat compared to a vegetarian diet (the experimental diets contained up to 420 g/d of meat compared to the mean intake of 73 g/d in meat-eaters in the current study).

5.5.3 Strengths and limitations

The main strengths of this study are the well-characterised cohort and the investigation of four distinct diet groups, the importance of which was highlighted by the marked separation of metabolite profile in vegans from that of the other diet groups. Moreover, the availability of information on a large number of covariates and traditional biomarkers (i.e. cholesterol and apolipoproteins) enabled us to reduce the risk of confounding and thereby to investigate the independent associations between diet group and metabolite concentrations.

5.5.3.1 Analytical considerations

Pre-analytical: Pre-analytic factors like fasting status, processing delay and freeze-thaw cycles could potentially impact the concentrations of metabolites and thereby the study results.

Time since last food or drink at blood collection contributed little to the total variability in our data (explaining 1.8%) and only four acylcarnitines (C2, C3, C5 and C14:1), five amino acids (isoleucine, leucine, methionine, tryptophan and tyrosine) and a biogenic amine (kynurenine) were associated with time since last food or drink in our analysis. Similarly, methodological work suggests that fasting status only explains a minor amount of the variability of metabolite concentrations [161, 171, 172]. Thus, this variable is likely to have at most a small effect on our results, although results for some amino acids (namely leucine/isoleucine, tryptophan and tyrosine) and lysophosphatidylcholines may potentially be affected by residual confounding as these metabolites have been suggested to reflect time since last meal [143]. The limited impact of fasting status on study results is in line with previous findings in EPIC [147, 161] and may be partly due to the relatively narrow time range since the last meal with the majority (72%) of samples being collected within four hours of the last food or drink.

Time from blood collection to processing did explain some but not a lot (4.4%) of the variability in the metabolite concentrations, and 28 of the metabolites as well as diet group were related to this variable. Although adjusted for, residual confounding by processing delay cannot be completely ruled out. Methodological work also using a Biocrates assay on plasma or serum, found that 72% and 81% of metabolites were stable after samples had been left at room temperature for 24 hours [144, 145]. The least stable metabolites were mainly amino acids and glycerophospholipids.

Finally, the thawing for previous assays is unlikely to have materially affected our results because the vast majority of metabolites (93 to 100%) have been reported to be stable after two to four freeze-thaw cycles [144, 145] and the number of freeze-thaw cycles in the current study was independent of diet group.

Analytical precision: The analytical precision in the current study, quantified by a mean overall CV of 8.2% across all measured metabolites and 3.6% of CVs being higher that 20%, was somewhat lower than reported elsewhere. A median overall CV of 11.4% using the Absolute*IDQ*® p150 Kit (Biocrates Life Sciences AG, Innsbruck, Austria) was found in serum from EPIC-Potsdam [126], while 8% of metabolites were reported to have a CV above 20% in a study using a different liquid chromatography tandem mass spectrometry platform for plasma [146]. The relatively low CVs in the current study reduced the likelihood of random measurement error attending our results.

To avoid results being affected by differences in measurements between analytical batches, samples from the four diet groups were randomly distributed between batches. This limits the risk of any batch effect and thereby reduces the risk of systematic error which could bias our results.

5.5.3.2 Statistical considerations

As for other metabolomics studies, multiple testing, and thereby increased risk of finding statistically significant associations purely by chance, is a limitation of the current analysis. To minimise this problem we controlled the FDR (Benjamini–Hochberg method). This method was chosen over the Bonferroni method, which was used in the previous Chapter 4. The Bonferroni method is very conservative as it ensures that the risk of getting one or more false positive findings is on average $< \alpha$ [140, 173]; thus this method is more conservative the more tests are performed. In contrast, controlling the FDR ensures that on average the proportion of false positive findings is α. Both methods assumes independence of tests which might not be the case here as metabolites are correlated, so the FDR might still be a relatively conservative method.

A further limitation is the moderate number of participants, although it is the largest study to date to investigate metabolite profiles in individuals from different habitual diet groups. Regardless of this limitation and the application of correction for multiple testing [140], many strong associations were seen between metabolites and diet group, indicating that individuals who exclude more or less animal products from their diet could have different metabolite profiles.

As described, PCA can be constructed in several ways and two different approaches were applied in this chapter. Although there were differences in which specific glycerophospholipids and sphingolipids that loaded most strongly on the principal components, and that the alternative model had higher loadings for more metabolites (due to the use of rotations), the conclusions drawn from the two different models were similar. This suggests that the PCA model used was robust.

5.5.4 Further research

The findings in this chapter inform our understanding of correlates of metabolite profile with implications for non-cancer and cancer research. The differences in phospholipid concentrations are particularly noteworthy given the essential role of phosphatidylcholines in lipoprotein synthesis [155], the strong positive correlations between phospholipids and atherosclerotic particles, and previous findings from the EPIC-Oxford cohort of both a more favourable lipid profile and a lower risk for ischemic heart disease in vegetarians and vegans than the other diet groups [153, 174]. Findings of associations between pre-diagnostic plasma concentrations of specific phosphatidylcholines and sphingomyelins and risk of cardiovascular disease also suggest a potential role of phospholipids in the development of cardiovascular disease [175, 176]. While further investigations into this might reveal novel risk factors for cardiovascular disease and help us understand the reason for the lower risk previously seen in vegetarians and vegans, such research is beyond the scope of the current thesis.

In contrast, investigations of the association between metabolite profile and risk of prostate cancer are included in subsequent chapters of this thesis (Chapters 6 and 7).

5.6 Conclusion

In conclusion, metabolite profiles in plasma could effectively differentiate between men from different habitual diet groups; especially vegan men compared to those who consume animal products. The difference in metabolite profile was mainly explained by the vegans having lower concentrations of glycerophospholipids and sphingolipids.

5.7 Tables

Table 5.1: Summary of overall coefficients of variation[1]

Overall coefficient of variation, %	Metabolites, n	Metabolites, %	Cumulative, %
≤ 5.0	5	3.6	3.6
5.1 – 10.0	99	71.7	75.4
10.1 – 15.0	22	15.9	91.3
15.1 – 20.0	7	5.1	96.4
> 20.0	5	3.6	100.0

[1] The coefficient of variation was based on 26 pooled plasma samples. It could not be calculated for seven metabolites, as there were too many non-quantifiable measurements.

Table 5.2: Summary of non-quantifiable measurements

Percent of samples affected	Number of metabolites, n	
	Outside measurable range	Missing data
> 0.0 – 5.0	14	4
5.1 – 10.0	1	0
10.1 – 20.0	3	3
20.1 – 30.0	1	1
30.1 – 40.0	1	0
40.1 – 50.0	6	0
> 50	10	0
Total	36	8

Table 5.3: The effect of excluding metabolites and men with missing metabolite concentration on the number of men with complete information for the included metabolites

Cut-off for samples with missing information, %	Excluded metabolites, n	Excluded men, n	Men with complete data available for analyses, n
> 25	0	203	189
> 20	1	177	215
> 15	3	51	341
> 10	4	13	379
> 5	4	13	379
> 1	5	5	387
> 0	8	0	392

Table 5.4: Characteristics and intakes of selected nutrients in 379 men in EPIC-Oxford by diet group

	Median (inter-quartile range) or n (%)[1]				$p_{difference}$[2]
	Meat-eaters (n=95)	Fish-eaters (n=97)	Vegetarians (n=91)	Vegans (n=96)	
Participant characteristics					
Age at blood collection, years	44 (37, 44)	40 (36, 45)	43 (36, 44)	40 (35, 44)	0.9
Body mass index[3], kg/m^2	24.4 (22.4, 26.0)	22.7 (21.1, 24.2)	22.7 (21.8, 25.1)	22.1 (20.5, 23.8)	0.0001
Current smoker, n (%)	14 (14.7)	9 (9.3)	6 (6.6)	7 (7.3)	0.1
Very physically active[3,4], n (%)	57 (60.0)	60 (61.9)	61 (67.0)	65 (67.7)	0.3
Nutrient intake					
Energy, kJ	9,198 (7,997, 11,045)	8,751 (7,518, 10,127)	9,012 (7,597, 10,971)	7,652 (6,084, 8,866)	0.0001
Protein, % of energy	14.98 (13.61, 16.82)	13.64 (12.29, 15.42)	13.26 (11.86, 14.14)	12.64 (11.68, 13.90)	0.0001
Carbohydrates, % of energy	51.30 (47.19, 55.56)	52.14 (48.32, 57.42)	54.28 (48.56, 58.49)	55.60 (52.27, 60.56)	0.0001
Fat, % of energy	31.96 (28.61, 35.00)	32.19 (27.22, 35.02)	31.46 (27.43, 35.39)	30.40 (25.39, 34.30)	0.2
Saturated fat, % of energy	11.18 (9.37, 13.15)	10.38 (8.31, 12.42)	10.47 (8.58, 12.60)	6.21 (5.12, 7.66)	0.0001
Monounsaturated fat, % of energy	10.85 (9.45, 12.22)	10.19 (8.63, 11.47)	9.86 (8.64, 11.38)	10.07 (7.90, 11.79)	0.02
Polyunsaturated fat, % of energy	6.15 (4.77, 7.65)	6.68 (5.52, 8.38)	7.27 (5.43, 8.68)	9.71 (7.63, 11.98)	0.0001
Alcohol, g/d	2.68 (0.87, 5.40)	3.40 (1.11, 6.88)	3.35 (1.34, 7.66)	1.92 (0.35, 6.04)	0.04
Blood sample characteristics					
Medication or supplement taken[3], n (%)	57 (60.0)	60 (61.9)	61 (67.0)	65 (67.7)	0.3
Time since last food or drink[3], h	2.0 (1.0, 3.5)	2.3 (1.3, 4.5)	2.5 (1.5, 4.5)	2.5 (1.3, 4.2)	0.05
Time of blood collection[3], h:min	11:50 (10:10, 15:27)	10:45 (09:28, 15:30)	10:30 (09:40, 15:05)	10:40 (09:45, 16:04)	0.3
Time from collection to processing[3], h	25.8 (23.4, 67.1)	44.1 (25.8, 71.9)	43.7 (24.6, 71.4)	41.9 (24.7, 72.3)	0.006
Lipemic, n (%)	23 (24.2)	13 (13.4)	23 (25.3)	17 (17.7)	0.1
Hemolytic, n (%)	1 (1.1)	6 (6.2)	6 (6.6)	3 (3.1)	0.2

[1] Values are median (interquartile range) if not otherwise specified as n (%)

[2] Differences between diet groups were tested using the Kruskal-Wallis one-way analysis of variance and χ^2 test for continuous and categorical variables, respectively (except for hemolysis for which Fisher's exact test was used)

[3] Information was missing for some participants: BMI n = 20; physical activity level n = 2; medicines or supplements taken at blood collection n = 2; time since last meal n = 11; time of blood collection n = 10; time from blood collection to processing n = 10

[4] Defined according to the Cambridge Physical Activity Index [121]

Table 5.5: Metabolites significantly associated with participant characteristics and factors related to blood sampling and processing in EPIC-Oxford[1,2,3]

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Body mass index (<22.5; 22.5-24.9; ≥25 kg/m^2)	C0 ↑ C5 ↑	Isoleucine ↑ Leucine ↑ Tryptophan ↑ Asparagine ↓ Glycine ↓ Serine ↓ Tyrosine ↑	Creatinine ↑ Kynurenine ↑	lysoPC a C18:1 ↓ lysoPC a C18:2 ↓ PC aa C28:1 ↑ PC aa C30:0 ↑ PC aa C32:0 ↑ PC aa C32:1 ↑ PC aa C32:2 ↑ PC aa C32:3 ↑ PC aa C34:1 ↑ PC aa C34:3 ↑ PC aa C34:4 ↑ PC aa C36:5 ↑ PC aa C36:6 ↑ PC aa C38:3 ↑ PC aa C38:6 ↑ PC aa C40:4 PC aa C40:5 ↑ PC aa C40:6 ↑ PC aa C42:2 ↓ PC ae C30:2 ↑ PC ae C36:4 ↑ PC ae C36:5 ↑ PC ae C38:0 PC ae C38:6 ↑ PC ae C42:4 ↓ PC ae C44:4 ↓ PC ae C44:5 ↓ PC ae C44:6 ↓	Hexose ↑	SM C18:0 ↑ SM C18:1 ↑ SM C20:2 ↑	43	36.4

Continues...

Table 5.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Energy intake (Fourths of the distribution: <7205; 7,206-8,632; 8,633-10,335; ≥10,336 kJ/d)	C5	Tryptophan ↑	Kynurenine ↑	lysoPC a C17:0 PC aa C28:1 ↑ PC aa C30:0 ↑ PC aa C32:0 PC aa C32:1 ↑ PC aa C32:2 ↑ PC aa C36:6 ↑ PC aa C38:6 PC aa C40:6 PC aa C42:0 PC aa C42:1 PC ae C30:2 ↑ PC ae C32:1 PC ae C32:2 PC ae C34:0 ↑ PC ae C34:1 ↑ PC ae C34:2 ↑ PC ae C36:1 ↑ PC ae C36:2 ↑ PC ae C38:0 PC ae C38:2 ↑ PC ae C38:3 ↑ PC ae C40:2 ↑ PC ae C40:3 ↑ PC ae C40:6 PC ae C42:2		SM (OH) C14:1 SM (OH) C16:1 SM (OH) C22:1 SM (OH) C22:2 SM (OH) C24:1 SM C16:0	35	29.7
Age (30-34; 35-39; 40-44; 45-49 years)		Histidine ↓ Serine	Creatinine	PC aa C30:0 PC aa C32:2 ↑ PC aa C36:1 ↑ PC aa C36:2 PC aa C36:3 ↑ PC ae C36:2 ↑ PC ae C36:3 PC ae C36:4 PC ae C38:2 PC ae C38:3 PC ae C38:4 PC ae C40:4		SM (OH) C14:1	16	13.6

Continues...

Table 5.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Alcohol intake (<1; 1-7; 8-15; ≥16 g/d)	C16 ↑	Tyrosine ↑		lysoPC a C16:0 ↑ PC aa C28:1 ↑ PC aa C30:0 ↑ PC aa C32:0 PC aa C32:1 PC aa C32:2 PC aa C34:1 PC aa C34:3 PC aa C34:4 PC aa C36:1 PC aa C36:3 PC aa C36:4 PC aa C36:5 ↑ PC aa C36:6 ↑ PC aa C38:5 PC aa C38:6 ↑ PC aa C40:6 ↑ PC ae C32:1 PC ae C34:0 ↑ PC ae C34:1 PC ae C36:1 ↑ PC ae C36:5 ↑ PC ae C38:0 ↑ PC ae C40:1 ↑ PC ae C40:2 ↑ PC ae C42:3 ↑		SM (OH) C14:1 SM (OH) C16:1 SM (OH) C24:1 ↑ SM C18:0 SM C24:1 SM C26:1	34	28.8
Smoking status							0	0.0
Physical activity level							0	0.0

Continues...

Table 5.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Time since last food or drink at blood collection (<1.5; 1.5-2.9; 3-4.4; ≥4.5 hours)	C14:1 C2 C3 C5 ↓	Isoleucine ↓ Leucine ↓ Methionine ↓ Tryptophan ↓ Tyrosine ↓	Kynurenine ↓				10	8.5
Time of day at blood collection (00:00-09:59; 10:00-12:59; 13:00-23:59)		Isoleucine Leucine Methionine ↓ Tryptophan Glutamine ↓	Kynurenine ↓	lysoPC a C18:2 lysoPC a C20:3 lysoPC a C20:4			9	7.6
Medication or supplement use at blood collection							0	0.0
Time from blood collection to processing (<25; 25-41; 41-72; ≥72 hours)	C0 C16 ↑ C18:1 ↑ C18:2 ↑	Isoleucine ↑ Leucine ↑ Valine ↑ Histidine ↑ Lysine ↑ Methionine Threonine ↑ Alanine ↑ Arginine Asparagine ↑ Aspartate Glutamate ↑ Glycine ↑ Serine ↑ Tyrosine ↑ Ornithine ↑		lysoPC a C16:0 ↑ lysoPC a C16:1 lysoPC a C17:0 lysoPC a C18:0 ↑ lysoPC a C18:1 ↑ lysoPC a C20:4 PC ae C44:3 ↑	Hexose ↓		28	23.7

Continues...

Table 5.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Lipemic (No; yes)		Isoleucine ↑ Threonine ↓ Tryptophan ↑		lysoPC a C16:0 ↑ lysoPC a C16:1 ↑ lysoPC a C18:0 ↑ lysoPC a C20:3 ↑ PC aa C30:0 ↑ PC aa C32:0 ↑ PC aa C32:1 ↑ PC aa C32:2 ↑ PC aa C32:3 ↑ PC aa C34:1 ↑ PC aa C34:3 ↑ PC aa C34:4 ↑ PC aa C36:1 ↑ PC aa C36:2 ↑ PC aa C36:3 ↑ PC aa C36:4 ↑ PC aa C38:3 ↑ PC aa C38:4 ↑ PC aa C38:5 ↑ PC aa C40:4 ↑ PC aa C40:5 ↑ PC aa C42:5 ↑ PC ae C44:5 ↓		SM C24:1 ↓ SM C26:1 ↓	28	23.7
Hemolysis (No; yes)		Histidine ↑ Alanine ↑ Glutamate ↑ Glutamine ↓ Glycine ↑ Ornithine ↑		lysoPC a C16:0 ↑ lysoPC a C18:0 ↑	Hexose ↓	SM C26:0 ↑ SM C26:1 ↑	11	9.3

[1] Associations were estimated using ANOVA of differences in geometric mean metabolite concentrations by categories of characteristics, after age ($30-34$; $35-39$; $40-44$; $45-49$ years) and BMI (<22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), where these were not the variables of interest

[2] Metabolites were included in the table if the association remained significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

[3] Upward arrows indicate that the mean metabolite concentration in each subsequent category of the characteristics variable was higher than in the preceding category. Downward arrows indicate that the mean metabolite concentration in each subsequent category of the characteristics variable was lower than in the preceding category

Table 5.6: Geometric mean metabolite concentrations by diet group and percentage difference compared to meat-eaters in EPIC-Oxford

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1, 2}$	$p_{diff}^{1, 3}$	$p_{diff}^{1, 4}$
	Meat-eaters n = 95	**Fish-eaters n = 97**	**Vegetarians n = 91**	**Vegans n = 96**			
Acylcarnitines							
C0	34.556 (32.809, 36.397) **Ref.**	33.206 (31.649, 34.841) **-4**	32.387 (30.829, 34.024) **-6**	29.042 (27.639, 30.516) **-16**	**$3x10^{-5}$**	**$3x10^{-5}$**	**$9x10^{-5}$**
C14:1	0.062 (0.056, 0.068) **Ref.**	0.060 (0.054, 0.065) **-4**	0.061 (0.056, 0.067) **-1**	0.059 (0.054, 0.065) **-4**	0.9	0.9	0.9
C16	0.152 (0.137, 0.167) **Ref.**	0.164 (0.150, 0.180) **+8**	0.156 (0.142, 0.172) **+3**	0.114 (0.103, 0.125) **-25**	**$3x10^{-7}$**	**$4x10^{-6}$**	**$1x10^{-5}$**
C18:1	0.225 (0.206, 0.246) **Ref.**	0.240 (0.221, 0.260) **+7**	0.233 (0.215, 0.254) **+4**	0.236 (0.217, 0.257) **+5**	0.8	0.8	0.8
C18:2	0.138 (0.124, 0.152) **Ref.**	0.163 (0.149, 0.180) **+19**	0.161 (0.146, 0.178) **+17**	0.199 (0.180, 0.219) **+45**	**$3x10^{-5}$**	**$1x10^{-5}$**	**$1x10^{-4}$**
C2	4.825 (4.494, 5.181) **Ref.**	4.510 (4.223, 4.817) **-7**	4.613 (4.312, 4.936) **-4**	4.419 (4.129, 4.729) **-8**	0.4	0.5	0.6
C3	0.314 (0.292, 0.337) **Ref.**	0.292 (0.274, 0.312) **-7**	0.314 (0.294, 0.336) **0**	0.255 (0.239, 0.273) **-19**	**$1x10^{-4}$**	**$9x10^{-9}$**	**$5x10^{-4}$**
C4	0.160 (0.149, 0.172) **Ref.**	0.149 (0.140, 0.160) **-7**	0.148 (0.138, 0.159) **-8**	0.127 (0.119, 0.136) **-21**	**$1x10^{-4}$**	**$1x10^{-4}$**	**$2x10^{-4}$**
C5	0.131 (0.121, 0.141) **Ref.**	0.122 (0.114, 0.130) **-7**	0.122 (0.114, 0.131) **-6**	0.092 (0.086, 0.099) **-30**	**$2x10^{-10}$**	**$2x10^{-9}$**	**$9x10^{-9}$**
Amino acids							
Isoleucine	98 (94, 103) **Ref.**	99 (94, 103) **0**	101 (96, 105) **+3**	95 (91, 100) **-3**	0.4	0.4	0.5
Leucine	204 (196, 213) **Ref.**	207 (199, 215) **+1**	209 (201, 217) **+2**	191 (183, 198) **-7**	**0.006**	**0.01**	**0.01**
Valine	231 (222, 240) **Ref.**	233 (224, 241) **+1**	235 (226, 244) **+2**	217 (209, 226) **-6**	**0.02**	**0.01**	0.05
Histidine	114 (110, 118) **Ref.**	122 (118, 126) **+7**	119 (116, 124) **+5**	117 (113, 120) **+2**	0.04	0.1	**0.03**
Lysine	241 (230, 252) **Ref.**	243 (233, 253) **+1**	236 (226, 247) **-2**	210 (201, 219) **-13**	**$1x10^{-5}$**	**$1x10^{-5}$**	**$7x10^{-5}$**
Methionine	29 (28, 31) **Ref.**	30 (29, 31) **+1**	31 (30, 32) **+5**	27 (26, 28) **-8**	**$2x10^{-4}$**	**$6x10^{-4}$**	**$9x10^{-4}$**

Continues...

Table 5.6 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1,2}$	$p_{diff}^{1,3}$	$p_{diff}^{1,4}$
	Meat-eaters n = 95	**Fish-eaters n = 97**	**Vegetarians n = 91**	**Vegans n = 96**			
Threonine	163 (157, 170) **Ref.**	168 (162, 174) **+3**	168 (162, 174) **+3**	164 (158, 170) **0**	0.6	0.7	0.5
Tryptophan	67 (65, 70) **Ref.**	70 (68, 73) **+4**	71 (68, 73) **+5**	64 (62, 67) **-4**	**0.001**	**0.003**	**0.004**
Alanine	566 (541, 593) **Ref.**	644 (618, 672) **+14**	627 (601, 655) **+11**	624 (598, 652) **+10**	**$9x10^{-4}$**	**$3x10^{-3}$**	**$7x10^{-4}$**
Arginine	50 (45, 56) **Ref.**	44 (40, 49) **-12**	43 (39, 48) **-14**	44 (39, 48) **-13**	0.2	0.3	0.1
Asparagine	93 (89, 96) **Ref.**	97 (94, 100) **+5**	96 (93, 99) **+4**	98 (95, 102) **+6**	0.2	0.4	0.2
Aspartate	64 (60, 67) **Ref.**	61 (58, 65) **-3**	67 (64, 71) **+6**	67 (64, 71) **+6**	0.04	0.08	**0.008**
Glutamate	259 (245, 274) **Ref.**	297 (282, 313) **+15**	282 (267, 298) **+9**	264 (250, 279) **+2**	**0.002**	**0.002**	**0.005**
Glutamine	511 (491, 531) **Ref.**	496 (478, 514) **-3**	522 (503, 541) **+2**	540 (521, 560) **+6**	**0.009**	0.05	**0.02**
Glycine	392 (376, 409) **Ref.**	422 (406, 439) **+8**	413 (397, 430) **+5**	451 (433, 470) **+15**	**$1x10^{-4}$**	**$1x10^{-4}$**	**$2x10^{-4}$**
Serine	190 (182, 198) **Ref.**	201 (193, 208) **+6**	199 (191, 206) **+5**	196 (189, 204) **+3**	0.3	0.4	0.3
Tyrosine	77 (74, 80) **Ref.**	82 (79, 86) **+7**	82 (79, 85) **+6**	73 (70, 76) **-5**	**$2x10^{-5}$**	**$2x10^{-4}$**	**$2x10^{-4}$**
Citrulline	37 (35, 39) **Ref.**	37 (36, 39) **0**	41 (39, 43) **+10**	40 (38, 42) **+7**	**0.02**	**0.01**	0.04
Ornithine	182 (174, 191) **Ref.**	195 (187, 203) **+7**	194 (186, 203) **+6**	206 (197, 216) **+13**	**0.005**	**0.01**	**0.004**
Biogenic amines							
Creatinine	80.26 (77.97, 82.63) **Ref.**	76.34 (74.32, 78.42) **-5**	76.10 (74.03, 78.23) **-5**	73.38 (71.38, 75.44) **-9**	**$6x10^{-4}$**	**$8x\ 10^{-4}$**	**$3x10^{-3}$**
Kynurenine	2.70 (2.57, 2.83) **Ref.**	2.83 (2.71, 2.96) **+5**	2.97 (2.84, 3.11) **+10**	2.56 (2.45, 2.68) **-5**	**$7x10^{-5}$**	**$4x10^{-3}$**	**$4x10^{-4}$**

Continues...

Table 5.6 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1, 2}$	$p_{diff}^{1, 3}$	$p_{diff}^{1, 4}$
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	**Vegans** **n = 96**			
Serotonin	0.38 (0.33, 0.42) **Ref.**	0.37 (0.33, 0.41) **-3**	0.36 (0.33, 0.40) **-3**	0.34 (0.30, 0.38) **-10**	0.6	0.7	0.6
Glycerophospholipids							
lysoPC a C16:0	209.15 (199.63, 219.14) **Ref.**	221.26 (211.92, 231.02) **6**	211.96 (202.77, 221.56) **1**	192.34 (183.98, 201.09) **-8**	**$2x10^{-4}$**	**$8x10^{-5}$**	**$1x10^{-2}$**
lysoPC a C16:1	5.817 (5.462, 6.194) **Ref.**	5.968 (5.630, 6.325) **3**	6.093 (5.739, 6.468) **5**	5.358 (5.046, 5.689) **-8**	**0.02**	**0.01**	0.05
lysoPC a C17:0	4.509 (4.206, 4.833) **Ref.**	4.503 (4.222, 4.802) **0**	4.116 (3.853, 4.397) **-9**	2.085 (1.951, 2.228) **-54**	**$3x10^{-53}$**	**$2x10^{-49}$**	**$5x10^{-46}$**
lysoPC a C18:0	75.531 (71.512, 79.776) **Ref.**	80.735 (76.750, 84.926) **7**	76.385 (72.518, 80.458) **1**	74.351 (70.574, 78.331) **-2**	0.1	0.08	0.3
lysoPC a C18:1	44.472 (41.962, 47.131) **Ref.**	45.935 (43.531, 48.472) **3**	45.250 (42.820, 47.817) **2**	49.351 (46.692, 52.162) **11**	0.07	0.2	**0.01**
lysoPC a C18:2	59.594 (55.999, 63.419) **Ref.**	58.224 (54.966, 61.676) **-2**	60.116 (56.666, 63.776) **1**	64.339 (60.633, 68.272) **8**	0.1	0.2	0.05
lysoPC a C20:3	3.603 (3.382, 3.837) **Ref.**	3.671 (3.463, 3.892) **2**	3.903 (3.676, 4.144) **8**	3.207 (3.019, 3.405) **-11**	**$1x10^{-4}$**	**$7x10^{-5}$**	**$1x10^{-3}$**
lysoPC a C20:4	9.903 (9.254, 10.599) **Ref.**	9.635 (9.048, 10.259) **-3**	10.255 (9.615, 10.938) **4**	10.498 (9.840, 11.199) **6**	0.2	0.4	0.1
PC aa C28:1	2.702 (2.531, 2.885) **Ref.**	2.744 (2.583, 2.915) **2**	2.645 (2.485, 2.814) **-2**	1.236 (1.161, 1.316) **-54**	**$8x10^{-62}$**	**$1x10^{-56}$**	**$3x10^{-60}$**
PC aa C30:0	3.600 (3.368, 3.849) **Ref.**	3.519 (3.308, 3.744) **-2**	3.427 (3.216, 3.651) **-5**	2.147 (2.014, 2.288) **-40**	**$1x10^{-29}$**	**$2x10^{-27}$**	**$3x10^{-23}$**
PC aa C32:0	18.120 (17.323, 18.955) **Ref.**	17.706 (16.983, 18.459) **-2**	16.881 (16.174, 17.618) **-7**	14.636 (14.021, 15.278) **-19**	**$4x10^{-11}$**	**$3x10^{-10}$**	**$5x10^{-5}$**
PC aa C32:1	14.127 (12.640, 15.789) **Ref.**	13.410 (12.098, 14.864) **-5**	14.553 (13.094, 16.175) **3**	7.970 (7.168, 8.862) **-44**	**$2x10^{-15}$**	**$2x10^{-15}$**	**$5x10^{-11}$**
PC aa C32:2	5.164 (4.741, 5.625) **Ref.**	5.423 (5.010, 5.870) **5**	5.957 (5.492, 6.461) **15**	3.634 (3.349, 3.943) **-30**	**$2x10^{-15}$**	**$4x10^{-14}$**	**$2x10^{-11}$**
PC aa C32:3	0.841 (0.797, 0.888) **Ref.**	0.829 (0.788, 0.871) **-2**	0.826 (0.784, 0.870) **-2**	0.731 (0.694, 0.77) **-13**	**$7x10^{-4}$**	**$4x10^{-3}$**	$3x10^{-1}$

Continues...

Table 5.6 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1,2}$	$p_{diff}^{1,3}$	$p_{diff}^{1,4}$
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	**Vegans** **n = 96**			
PC aa C34:1	189.47 (181.82, 197.44) **Ref.**	184.40 (177.50, 191.56) **-3**	189.33 (182.06, 196.88) **0**	164.13 (157.81, 170.71) **-13**	**$7x10^{-7}$**	**$1x10^{-6}$**	**$5x10^{-3}$**
PC aa C34:2	228.29 (219.61, 237.31) **Ref.**	229.30 (221.22, 237.67) **0**	227.07 (218.87, 235.59) **-1**	227.48 (219.23, 236.04) **0**	>0.9	0.9	>0.9
PC aa C34:3	20.712 (19.434, 22.074) **Ref.**	20.507 (19.333, 21.753) **-1**	22.374 (21.061, 23.770) **8**	18.689 (17.587, 19.859) **-10**	**0.001**	**0.002**	**0.01**
PC aa C34:4	1.910 (1.764, 2.069) **Ref.**	1.892 (1.758, 2.038) **-1**	2.061 (1.911, 2.224) **8**	1.400 (1.297, 1.511) **-27**	**$3x10^{-11}$**	**$6x10^{-11}$**	**$1x10^{-7}$**
PC aa C36:1	51.100 (48.050, 54.344) **Ref.**	49.756 (47.000, 52.672) **-3**	49.853 (47.022, 52.855) **-2**	42.689 (40.256, 45.270) **-16**	**$1x10^{-4}$**	**$7x10^{-5}$**	$4x10^{-1}$
PC aa C36:2	201.65 (195.35, 208.16) **Ref.**	200.87 (195.05, 206.86) **0**	204.53 (198.45, 210.79) **1**	199.69 (193.73, 205.83) **-1**	0.7	0.9	0.09
PC aa C36:3	137.32 (132.08, 142.77) **Ref.**	138.61 (133.70, 143.69) **1**	151.01 (145.53, 156.70) **10**	147.61 (142.24, 153.20) **7**	**$4x10^{-4}$**	**$1x10^{-3}$**	**$2x10^{-10}$**
PC aa C36:4	158.83 (151.93, 166.04) **Ref.**	153.04 (146.87, 159.46) **-4**	167.81 (160.88, 175.04) **6**	161.67 (154.97, 168.66) **2**	**0.02**	0.05	**$4x10^{-5}$**
PC aa C36:5	27.457 (24.681, 30.545) **Ref.**	25.252 (22.879, 27.870) **-8**	18.466 (16.687, 20.433) **-33**	11.278 (10.188, 12.485) **-59**	**$6x10^{-31}$**	**$1x10^{-30}$**	**$3x10^{-24}$**
PC aa C36:6	1.130 (1.031, 1.239) **Ref.**	1.059 (0.973, 1.153) **-6**	0.727 (0.667, 0.794) **-36**	0.402 (0.368, 0.439) **-64**	**$2x10^{-50}$**	**$6x10^{-49}$**	**$4x10^{-43}$**
PC aa C38:0	2.760 (2.610, 2.918) **Ref.**	2.431 (2.308, 2.559) **-12**	1.645 (1.560, 1.734) **-40**	1.413 (1.340, 1.490) **-49**	**$7x10^{-58}$**	**$3x10^{-55}$**	**$3x10^{-53}$**
PC aa C38:3	40.017 (37.859, 42.298) **Ref.**	40.483 (38.458, 42.615) **1**	41.774 (39.631, 44.034) **4**	36.847 (34.950, 38.848) **-8**	**0.01**	**0.01**	0.2
PC aa C38:4	78.260 (73.829, 82.956) **Ref.**	74.974 (71.037, 79.130) **-4**	84.997 (80.419, 89.836) **9**	89.060 (84.245, 94.151) **14**	**$4x10^{-5}$**	**$2x10^{-4}$**	**$2x10^{-9}$**
PC aa C38:5	50.612 (47.703, 53.698) **Ref.**	49.022 (46.408, 51.783) **-3**	50.665 (47.894, 53.596) **0**	47.867 (45.239, 50.647) **-5**	0.5	0.5	0.4
PC aa C38:6	80.145 (74.727, 85.957) **Ref.**	73.884 (69.248, 78.831) **-8**	46.618 (43.618, 49.824) **-42**	33.465 (31.303, 35.775) **-58**	**$8x10^{-61}$**	**$6x10^{-59}$**	**$4x10^{-54}$**
PC aa C40:2	0.329 (0.311, 0.348) **Ref.**	0.316 (0.300, 0.333) **-4**	0.287 (0.272, 0.303) **-13**	0.304 (0.288, 0.320) **-8**	**0.004**	**0.003**	**0.002**

Continues...

Table 5.6 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1, 2}$	$p_{diff}^{1, 3}$	$p_{diff}^{1, 4}$
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	**Vegans** **n = 96**			
PC aa C40:3	0.584 (0.554, 0.616) **Ref.**	0.545 (0.520, 0.573) **-7**	0.511 (0.486, 0.538) **-12**	0.488 (0.464, 0.513) **-16**	**$1x10^{-5}$**	**$2x10^{-5}$**	**$1x10^{-3}$**
PC aa C40:4	2.439 (2.300, 2.586) **Ref.**	2.387 (2.261, 2.520) **-2**	2.918 (2.760, 3.085) **20**	2.745 (2.595, 2.903) **13**	**$2x10^{-7}$**	**$6x10^{-7}$**	**$5x10^{-12}$**
PC aa C40:5	7.020 (6.575, 7.495) **Ref.**	6.903 (6.497, 7.334) **-2**	7.543 (7.088, 8.027) **7**	6.826 (6.413, 7.266) **-3**	0.1	0.2	**0.03**
PC aa C40:6	23.510 (21.819, 25.333) **Ref.**	21.998 (20.529, 23.572) **-6**	13.703 (12.764, 14.710) **-42**	10.622 (9.892, 11.406) **-55**	**$2x10^{-50}$**	**$9x10^{-49}$**	**$7x10^{-44}$**
PC aa C42:0	0.598 (0.568, 0.629) **Ref.**	0.582 (0.555, 0.610) **-3**	0.479 (0.457, 0.503) **-20**	0.446 (0.425, 0.469) **-25**	**$9x10^{-19}$**	**$7x10^{-17}$**	**$2x10^{-14}$**
PC aa C42:1	0.314 (0.300, 0.329) **Ref.**	0.314 (0.301, 0.328) **0**	0.260 (0.249, 0.272) **-17**	0.235 (0.224, 0.245) **-25**	**$5x10^{-23}$**	**$7x10^{-21}$**	**$1x10^{-17}$**
PC aa C42:2	0.229 (0.218, 0.240) **Ref.**	0.236 (0.226, 0.247) **3**	0.193 (0.184, 0.202) **-16**	0.216 (0.206, 0.226) **-6**	**$2x10^{-9}$**	**$3x10^{-3}$**	**$2x10^{-9}$**
PC aa C42:4	0.145 (0.139, 0.152) **Ref.**	0.145 (0.139, 0.151) **0**	0.147 (0.141, 0.153) **1**	0.161 (0.154, 0.168) **11**	**0.002**	**0.003**	**$9x10^{-7}$**
PC aa C42:5	0.316 (0.299, 0.333) **Ref.**	0.296 (0.282, 0.311) **-6**	0.291 (0.276, 0.306) **-8**	0.267 (0.254, 0.281) **-15**	**$4x10^{-4}$**	**$6x10^{-4}$**	$6x10^{-2}$
PC ae C30:2	0.101 (0.096, 0.107) **Ref.**	0.101 (0.096, 0.106) **0**	0.101 (0.096, 0.106) **0**	0.067 (0.063, 0.070) **-34**	**$1x10^{-30}$**	**$4x10^{-27}$**	**$2x10^{-23}$**
PC ae C32:1	3.487 (3.324, 3.658) **Ref.**	3.216 (3.076, 3.361) **-8**	3.079 (2.942, 3.222) **-12**	2.433 (2.324, 2.547) **-30**	**$6x10^{-23}$**	**$4x10^{-20}$**	**$6x10^{-16}$**
PC ae C32:2	0.986 (0.941, 1.034) **Ref.**	0.897 (0.859, 0.937) **-9**	0.813 (0.778, 0.850) **-18**	0.629 (0.601, 0.657) **-36**	**$3x10^{-35}$**	**$2x10^{-32}$**	**$2x10^{-28}$**
PC ae C34:0	1.848 (1.730, 1.974) **Ref.**	1.720 (1.618, 1.828) **-7**	1.568 (1.473, 1.669) **-15**	0.728 (0.684, 0.775) **-61**	**$1x10^{-69}$**	**$1x10^{-63}$**	**$2x10^{-67}$**
PC ae C34:1	11.322 (10.798, 11.871) **Ref.**	10.830 (10.366, 11.316) **-4**	11.007 (10.523, 11.514) **-3**	8.178 (7.817, 8.556) **-28**	**$8x10^{-23}$**	**$2x10^{-21}$**	**$3x10^{-16}$**
PC ae C34:2	14.311 (13.639, 15.016) **Ref.**	12.301 (11.766, 12.861) **-14**	12.590 (12.028, 13.179) **-12**	9.700 (9.265, 10.155) **-32**	**$7x10^{-24}$**	**$1x10^{-20}$**	**$7x10^{-18}$**
PC ae C34:3	10.716 (10.161, 11.301) **Ref.**	9.264 (8.819, 9.731) **-14**	9.116 (8.667, 9.588) **-15**	8.799 (8.364, 9.256) **-18**	**$3x10^{-6}$**	**$3x10^{-5}$**	**$1x10^{-4}$**

Continues...

Table 5.6 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1,2}$	$p_{diff}^{1,3}$	$p_{diff}^{1,4}$
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	**Vegans** **n = 96**			
PC ae C36:1	8.601 (8.109, 9.123) **Ref.**	8.135 (7.704, 8.591) **-5**	7.937 (7.505, 8.393) **-8**	4.100 (3.876, 4.337) **-52**	**$9x10^{-61}$**	**$1x10^{-56}$**	**$2x10^{-58}$**
PC ae C36:2	17.227 (16.290, 18.217) **Ref.**	15.774 (14.978, 16.611) **-8**	16.291 (15.448, 17.179) **-5**	9.782 (9.274, 10.318) **-43**	**$2x10^{-42}$**	**$1x10^{-38}$**	**$3x10^{-36}$**
PC ae C36:3	8.716 (8.337, 9.113) **Ref.**	7.709 (7.398, 8.033) **-12**	8.056 (7.722, 8.404) **-8**	7.469 (7.159, 7.793) **-14**	**$3x10^{-5}$**	**$2x10^{-4}$**	**$6x10^{-4}$**
PC ae C36:4	15.507 (14.822, 16.224) **Ref.**	12.772 (12.249, 13.317) **-18**	13.732 (13.155, 14.335) **-11**	11.956 (11.452, 12.482) **-23**	**$5x10^{-13}$**	**$5x10^{-12}$**	**$1x10^{-10}$**
PC ae C36:5	10.795 (10.235, 11.386) **Ref.**	8.737 (8.316, 9.178) **-19**	8.427 (8.011, 8.864) **-22**	7.693 (7.312, 8.094) **-29**	**$1x10^{-15}$**	**$5x10^{-15}$**	**4x10-^{12}**
PC ae C38:0	2.628 (2.466, 2.800) **Ref.**	2.505 (2.362, 2.656) **-5**	1.831 (1.724, 1.945) **-30**	1.331 (1.253, 1.414) **-49**	**$4x10^{-48}$**	**$2x10^{-47}$**	**$3x10^{-41}$**
PC ae C38:2	2.302 (2.190, 2.420) **Ref.**	2.19 (2.090, 2.294) **-5**	2.207 (2.104, 2.314) **-4**	1.723 (1.643, 1.807) **-25**	**$6x10^{-16}$**	**$8x10^{-14}$**	**$7x10^{-10}$**
PC ae C38:3	4.029 (3.826, 4.243) **Ref.**	3.798 (3.620, 3.984) **-6**	3.965 (3.775, 4.165) **-2**	2.359 (2.246, 2.479) **-41**	**$1x10^{-46}$**	**$7x10^{-43}$**	**$1x10^{-41}$**
PC ae C38:4	11.083 (10.558, 11.635) **Ref.**	9.839 (9.407, 10.292) **-11**	11.046 (10.549, 11.568) **0**	8.045 (7.681, 8.426) **-27**	**$8x10^{-21}$**	**$5x10^{-19}$**	**$4x10^{-16}$**
PC ae C38:5	14.746 (14.091, 15.431) **Ref.**	13.208 (12.664, 13.774) **-10**	13.667 (13.090, 14.269) **-7**	14.144 (13.545, 14.770) **-4**	**0.006**	**0.01**	**0.0001**
PC ae C38:6	7.132 (6.759, 7.525) **Ref.**	6.013 (5.722, 6.319) **-16**	4.719 (4.485, 4.966) **-34**	4.221 (4.011, 4.443) **-41**	**$9x10^{-39}$**	**$2x10^{-37}$**	**$4x10^{-34}$**
PC ae C40:1	1.397 (1.324, 1.473) **Ref.**	1.349 (1.284, 1.418) **-3**	1.065 (1.013, 1.121) **-24**	0.888 (0.844, 0.935) **-36**	**$3x10^{-34}$**	**$9x10^{-34}$**	**$8x10^{-28}$**
PC ae C40:2	1.740 (1.648, 1.838) **Ref.**	1.625 (1.545, 1.710) **-7**	1.422 (1.350, 1.498) **-18**	0.815 (0.774, 0.859) **-53**	**$2x10^{-66}$**	**$3x10^{-61}$**	**$2x10^{-62}$**
PC ae C40:3	1.012 (0.973, 1.054) **Ref.**	0.990 (0.954, 1.027) **-2**	0.981 (0.944, 1.019) **-3**	0.740 (0.712, 0.769) **-27**	**$1x10^{-29}$**	**$5x10^{-26}$**	**$7x10^{-23}$**
PC ae C40:4	1.932 (1.852, 2.016) **Ref.**	1.831 (1.761, 1.905) **-5**	1.979 (1.900, 2.060) **2**	1.650 (1.585, 1.719) **-15**	**$1x10^{-8}$**	**$1x10^{-7}$**	**$1x10^{-5}$**
PC ae C40:5	3.135 (2.998, 3.278) **Ref.**	2.941 (2.822, 3.065) **-6**	2.993 (2.868, 3.122) **-5**	2.525 (2.419, 2.635) **-19**	**$2x10^{-10}$**	**$2x10^{-9}$**	**4x10-^{6}**

Continues...

Table 5.6 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1, 2}$	$p_{diff}^{1, 3}$	$p_{diff}^{1, 4}$
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	**Vegans** **n = 96**			
PC ae C40:6	4.297 (4.071, 4.535) **Ref.**	3.894 (3.704, 4.093) **-9**	2.978 (2.829, 3.135) **-31**	2.312 (2.196, 2.434) **-46**	**$2x10^{-50}$**	**$7x10^{-47}$**	**$1x10^{-43}$**
PC ae C42:1	0.325 (0.310, 0.342) **Ref.**	0.323 (0.309, 0.338) **-1**	0.336 (0.321, 0.352) **3**	0.343 (0.327, 0.359) **5**	0.2	0.3	**0.01**
PC ae C42:2	0.514 (0.488, 0.541) **Ref.**	0.500 (0.477, 0.525) **-3**	0.473 (0.450, 0.497) **-8**	0.377 (0.359, 0.396) **-27**	**$2x10^{-17}$**	**$8x10^{-16}$**	**$3x10^{-11}$**
PC ae C42:3	0.831 (0.800, 0.864) **Ref.**	0.830 (0.801, 0.86) **0**	0.729 (0.702, 0.756) **-12**	0.628 (0.606, 0.652) **-24**	**$5x10^{-27}$**	**$6x10^{-25}$**	**$2x10^{-20}$**
PC ae C42:4	0.745 (0.711, 0.780) **Ref.**	0.735 (0.704, 0.768) **-1**	0.759 (0.726, 0.793) **2**	0.707 (0.676, 0.739) **-5**	0.2	0.2	0.5
PC ae C42:5	1.700 (1.629, 1.774) **Ref.**	1.630 (1.567, 1.696) **-4**	1.645 (1.580, 1.714) **-3**	1.480 (1.421, 1.542) **-13**	**$7x10^{-5}$**	**$9x10^{-5}$**	**$5x10^{-3}$**
PC ae C44:3	0.161 (0.154, 0.168) **Ref.**	0.169 (0.163, 0.176) **5**	0.158 (0.152, 0.165) **-1**	0.149 (0.143, 0.155) **-7**	**$2x10^{-4}$**	**$1x10^{-3}$**	**$2x10^{-2}$**
PC ae C44:4	0.324 (0.311, 0.338) **Ref.**	0.326 (0.314, 0.339) **0**	0.339 (0.326, 0.353) **4**	0.293 (0.282, 0.305) **-10**	**$1x10^{-5}$**	**$6x10^{-5}$**	**$4x10^{-4}$**
PC ae C44:5	1.239 (1.173, 1.308) **Ref.**	1.195 (1.136, 1.257) **-4**	1.203 (1.142, 1.267) **-3**	1.083 (1.028, 1.140) **-13**	**0.004**	**0.002**	0.05
PC ae C44:6	1.002 (0.944, 1.064) **Ref.**	1.009 (0.955, 1.067) **1**	0.981 (0.927, 1.039) **-2**	1.068 (1.009, 1.131) **7**	0.2	0.2	0.1
Hexose							
Hexose	2310 (2126, 2510) **Ref.**	2131 (1973, 2301) **-8**	2154 (1991, 2331) **-7**	2073 (1915, 2244) **-10**	0.3	0.3	0.5
Sphingolipids							
SM (OH) C14:1	5.979 (5.606, 6.378) **Ref.**	5.694 (5.364, 6.044) **-5**	5.496 (5.169, 5.843) **-8**	2.250 (2.116, 2.393) **-62**	**$2x10^{-79}$**	**$2x10^{-72}$**	**$1x10^{-78}$**
SM (OH) C16:1	3.064 (2.885, 3.254) **Ref.**	2.798 (2.646, 2.958) **-9**	2.579 (2.436, 2.731) **-16**	1.314 (1.241, 1.392) **-57**	**$3x10^{-67}$**	**$1x10^{-61}$**	**$1x10^{-63}$**
SM (OH) C22:1	9.081 (8.602, 9.586) **Ref.**	9.029 (8.588, 9.493) **-1**	8.680 (8.245, 9.138) **-4**	6.683 (6.347, 7.037) **-26**	**$3x10^{-17}$**	**$2x10^{-14}$**	**$2x10^{-10}$**
SM (OH) C22:2	7.139 (6.771, 7.527) **Ref.**	6.935 (6.604, 7.283) **-3**	6.365 (6.053, 6.693) **-11**	4.394 (4.177, 4.621) **-38**	**$5x10^{-37}$**	**$3x10^{-33}$**	**$5x10^{-32}$**

Continues...

Table 5.6 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% mean difference compared to meat-eaters**				$p_{diff}^{1,2}$	$p_{diff}^{1,3}$	$p_{diff}^{1,4}$
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	**Vegans** **n = 96**			
SM (OH) C24:1	0.744 (0.701, 0.788) **Ref.**	0.752 (0.713, 0.794) **1**	0.733 (0.693, 0.775) **-1**	0.518 (0.490, 0.547) **-30**	**$4x10^{-22}$**	**$1x10^{-19}$**	**$2x10^{-15}$**
SM C16:0	103.24 (99.54, 107.09) **Ref.**	99.84 (96.52, 103.27) **-3**	96.52 (93.23, 99.93) **-7**	90.59 (87.48, 93.80) **-12**	**$6x10^{-6}$**	**$1x10^{-4}$**	$7x10^{-2}$
SM C16:1	14.257 (13.670, 14.870) **Ref.**	14.048 (13.512, 14.606) **-1**	13.727 (13.190, 14.287) **-4**	13.591 (13.056, 14.147) **-5**	0.4	0.4	**0.008**
SM C18:0	18.534 (17.514, 19.613) **Ref.**	17.137 (16.262, 18.059) **-8**	15.808 (14.981, 16.682) **-15**	11.580 (10.972, 12.223) **-38**	**$2x10^{-28}$**	**$1x10^{-26}$**	**$6x10^{-22}$**
SM C18:1	8.873 (8.416, 9.355) **Ref.**	8.344 (7.945, 8.762) **-6**	7.924 (7.536, 8.332) **-11**	7.133 (6.782, 7.502) **-20**	**$1x10^{-7}$**	**$5x10^{-7}$**	**$2x10^{-3}$**
SM C20:2	0.602 (0.554, 0.655) **Ref.**	0.514 (0.476, 0.556) **-15**	0.524 (0.484, 0.568) **-13**	0.414 (0.382, 0.448) **-31**	**$6x10^{-8}$**	**$3x10^{-7}$**	**$4x10^{-5}$**
SM C24:0	12.884 (12.230, 13.572) **Ref.**	13.131 (12.513, 13.780) **2**	13.006 (12.378, 13.666) **1**	13.361 (12.714, 14.042) **4**	0.8	0.6	**$2x10^{-8}$**
SM C24:1	28.339 (27.059, 29.680) **Ref.**	26.935 (25.806, 28.112) **-5**	24.206 (23.166, 25.293) **-15**	24.214 (23.169, 25.305) **-15**	**$2x10^{-7}$**	**$1x10^{-7}$**	**$4x10^{-6}$**
SM C26:0	0.108 (0.102, 0.115) **Ref.**	0.107 (0.101, 0.113) **-1**	0.103 (0.098, 0.109) **-5**	0.088 (0.083, 0.093) **-19**	**$9x10^{-7}$**	**$1x10^{-5}$**	**$1x10^{-2}$**
SM C26:1	0.219 (0.206, 0.232) **Ref.**	0.206 (0.195, 0.217) **-6**	0.172 (0.162, 0.182) **-21**	0.165 (0.156, 0.175) **-24**	**$1x10^{-12}$**	**$3x10^{-12}$**	**$1x10^{-10}$**

[1] p-values refer to p_{diff} across all four diet groups calculated by ANOVA. Conventional p-values are shown and those marked in bold were significant after correcting for multiple testing using a false discovery rate procedure (Benjamini–Hochberg) at $\alpha = 0.05$

[2] The main model was adjusted for age ($30-34$; $35-39$; $40-44$; $45-49$ years), BMI (<22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), smoking status (never; former; current), alcohol intake (<1; $1-7$; $8-15$; ≥ 16 g/d), time since last food or drink at blood collection (<1.5; $1.5-2.9$; $3.0-4.4$; ≥ 4.5 hours; unknown), and time between blood collection and processing (fourths of the distribution; unknown)

[3] The main model was further adjusted for energy intake (fourths of the distribution, corresponding to $<7{,}205$; $7{,}206-8{,}632$; $8{,}633-10{,}335$; $\geq 10{,}336$ kJ/d), time of day at blood collection (00:00 − 9:59; 10:00 − 12:59; 13:00 − 23:59; unknown), hemolysis (no; yes) and lipemic status (no; yes)

[4] The main model was further adjusted for total cholesterol (fifths of the distribution, i.e. <4.0; $4.0-4.5$; $4.5-5.0$; $5.0-5.7$; ≥ 5.7 mmol/L; unknown)

Table 5.7: Geometric mean metabolite concentrations in vegans and non-vegans and percentage difference compared to non-vegans in EPIC-Oxford[1]

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to non-vegans**		$p_{difference}$[2]
	Non-Vegans **n = 283**	**Vegans** **n = 96**	
Acylcarnitines			
C0	33.363 (32.453, 34.298) **100**	29.097 (27.694, 30.571) **-13**	**5×10^{-6}**
C14:1	0.061 (0.058, 0.064) **100**	0.059 (0.054, 0.065) **-3**	0.7
C16	0.157 (0.149, 0.166) **100**	0.113 (0.103, 0.124) **-28**	**1×10^{-8}**
C18:1	0.233 (0.222, 0.244) **100**	0.235 (0.216, 0.256) **1**	0.9
C18:2	0.154 (0.146, 0.163) **100**	0.197 (0.179, 0.217) **28**	**3×10^{-5}**
C2	4.642 (4.469, 4.820) **100**	4.435 (4.145, 4.744) **-4**	0.3
C3	0.306 (0.295, 0.318) **100**	0.256 (0.239, 0.274) **-16**	**2×10^{-5}**
C4	0.152 (0.147, 0.158) **100**	0.128 (0.119, 0.137) **-16**	**2×10^{-5}**
C5	0.125 (0.120, 0.130) **100**	0.092 (0.086, 0.099) **-26**	**9×10^{-12}**
Amino acids			
Isoleucine	99 (97, 102) **100**	95 (91, 100) **-4**	0.2
Leucine	207 (202, 211) **100**	190 (183, 198) **-8**	**0.006**
Valine	233 (228, 238) **100**	217 (209, 225) **-7**	**0.002**
Histidine	118 (116, 121) **100**	116 (112, 120) **-2**	0.3
Lysine	240 (234, 246) **100**	210 (201, 219) **-13**	**6×10^{-7}**
Methionine	30 (29, 31) **100**	27 (26, 28) **-10**	**3×10^{-5}**
Threonine	166 (163, 170) **100**	164 (158, 170) **-1**	0.5
Tryptophan	69 (68, 71) **100**	64 (62, 67) **-7**	**5×10^{-4}**
Alanine	613 (598, 628) **100**	620 (593, 648) **1**	0.7
Arginine	46 (43, 48) **100**	44 (40, 49) **-4**	0.6
Asparagine	95 (93, 97) **100**	98 (95, 101) **3**	0.2

Continues...

Table 5.7 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to non-vegans**		$p_{difference}$ [2]
	Non-Vegans **n = 283**	**Vegans** **n = 96**	
Aspartate	64 (62, 66) **100**	68 (64, 71) **6**	0.08
Glutamate	280 (271, 289) **100**	262 (248, 277) **-6**	0.05
Glutamine	509 (498, 519) **100**	541 (522, 561) **6**	**0.005**
Glycine	410 (400, 419) **100**	449 (432, 468) **10**	**1×10^{-4}**
Serine	196 (192, 201) **100**	196 (189, 203) **0**	0.9
Tyrosine	81 (79, 82) **100**	73 (70, 76) **-10**	**2×10^{-5}**
Citrulline	39 (37, 40) **100**	40 (38, 42) **3**	0.3
Ornithine	191 (186, 195) **100**	206 (197, 215) **8**	**0.004**
Biogenic amines			
Creatinine	77.49 (76.30, 78.71) **100**	73.57 (71.55, 75.64) **-5**	**0.002**
Kynurenine	2.83 (2.76, 2.91) **100**	2.56 (2.44, 2.68) **-10**	**2×10^{-4}**
Serotonin	0.37 (0.35, 0.39) **100**	0.34 (0.30, 0.38) **-8**	0.2
Glycerophospholipds			
lysoPC a C16:0	214.366 (209.106, 219.759) **100**	191.750 (183.424, 200.454) **-11**	**4×10^{-5}**
lysoPC a C16:1	5.959 (5.763, 6.161) **100**	5.352 (5.042, 5.681) **-10**	**0.003**
lysoPC a C17:0	4.377 (4.218, 4.543) **100**	2.084 (1.950, 2.227) **-52**	**1×10^{-54}**
lysoPC a C18:0	77.653 (75.422, 79.950) **100**	74.078 (70.318, 78.038) **-5**	0.1
lysoPC a C18:1	45.246 (43.872, 46.663) **100**	49.266 (46.625, 52.057) **9**	**0.01**
lysoPC a C18:2	59.260 (57.334, 61.250) **100**	64.426 (60.734, 68.343) **9**	**0.02**
lysoPC a C20:3	3.721 (3.598, 3.849) **100**	3.205 (3.018, 3.403) **-14**	**4×10^{-5}**
lysoPC a C20:4	9.915 (9.564, 10.280) **100**	10.516 (9.859, 11.216) **6**	0.1
PC aa C28:1	2.699 (2.606, 2.794) **100**	1.235 (1.161, 1.314) **-54**	**2×10^{-64}**

Continues...

Table 5.7 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to non-vegans**		$p_{difference}$[2]
	Non-Vegans **n = 283**	**Vegans** **n = 96**	
PC aa C30:0	3.515 (3.392, 3.642) **100**	2.149 (2.017, 2.289) **-39**	**1×10^{-31}**
PC aa C32:0	17.566 (17.149, 17.994) **100**	14.649 (14.033, 15.293) **-17**	**1×10^{-13}**
PC aa C32:1	13.996 (13.192, 14.848) **100**	7.994 (7.193, 8.885) **-43**	**4×10^{-17}**
PC aa C32:2	5.502 (5.256, 5.759) **100**	3.626 (3.342, 3.935) **-34**	**5×10^{-16}**
PC aa C32:3	0.832 (0.808, 0.856) **100**	0.731 (0.695, 0.770) **-12**	**4×10^{-5}**
PC aa C34:1	187.572 (183.511, 191.724) **100**	164.379 (158.074, 170.936) **-12**	**4×10^{-8}**
PC aa C34:2	228.263 (223.617, 233.005) **100**	227.415 (219.212, 235.926) **0**	0.9
PC aa C34:3	21.154 (20.446, 21.886) **100**	18.707 (17.603, 19.879) **-12**	**8×10^{-4}**
PC aa C34:4	1.951 (1.869, 2.035) **100**	1.401 (1.299, 1.512) **-28**	**3×10^{-12}**
PC aa C36:1	50.211 (48.597, 51.879) **100**	42.748 (40.324, 45.317) **-15**	**6×10^{-6}**
PC aa C36:2	202.281 (198.898, 205.721) **100**	199.750 (193.820, 205.860) **-1**	0.5
PC aa C36:3	142.037 (139.072, 145.066) **100**	147.610 (142.149, 153.280) **4**	0.09
PC aa C36:4	159.489 (155.727, 163.341) **100**	162.053 (155.287, 169.113) **2**	0.02
PC aa C36:5	23.464 (22.118, 24.892) **100**	11.308 (10.175, 12.567) **-52**	**2×10^{-26}**
PC aa C36:6	0.959 (0.910, 1.011) **100**	0.403 (0.367, 0.442) **-58**	**2×10^{-41}**
PC aa C38:0	2.234 (2.153, 2.318) **100**	1.419 (1.329, 1.516) **-36**	**4×10^{-26}**
PC aa C38:3	40.741 (39.558, 41.960) **100**	36.832 (34.942, 38.823) **-10**	**0.002**
PC aa C38:4	79.117 (76.673, 81.639) **100**	89.314 (84.444, 94.464) **13**	**4×10^{-4}**
PC aa C38:5	50.047 (48.498, 51.646) **100**	47.953 (45.333, 50.724) **-4**	0.2
PC aa C38:6	65.440 (62.609, 68.398) **100**	33.524 (30.977, 36.280) **-49**	**3×10^{-36}**
PC aa C40:2	0.311 (0.301, 0.320) **100**	0.304 (0.288, 0.321) **-2**	0.5

Continues...

Table 5.7 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to non-vegans**		$p_{difference}$ [2]
	Non-Vegans **n = 283**	**Vegans** **n = 96**	
PC aa C40:3	0.546 (0.531, 0.562) **100**	0.490 (0.465, 0.515) **-10**	**4×10^{-4}**
PC aa C40:4	2.563 (2.481, 2.647) **100**	2.750 (2.596, 2.915) **7**	0.04
PC aa C40:5	7.139 (6.894, 7.393) **100**	6.835 (6.421, 7.276) **-4**	0.2
PC aa C40:6	19.312 (18.436, 20.230) **100**	10.632 (9.785, 11.551) **-45**	**4×10^{-28}**
PC aa C42:0	0.552 (0.536, 0.568) **100**	0.447 (0.424, 0.470) **-19**	**5×10^{-11}**
PC aa C42:1	0.296 (0.288, 0.304) **100**	0.234 (0.224, 0.246) **-21**	**3×10^{-15}**
PC aa C42:2	0.219 (0.213, 0.225) **100**	0.215 (0.205, 0.226) **-2**	0.6
PC aa C42:4	0.146 (0.142, 0.149) **100**	0.161 (0.154, 0.168) **10**	**1×10^{-4}**
PC aa C42:5	0.300 (0.292, 0.309) **100**	0.268 (0.255, 0.282) **-11**	**3×10^{-4}**
PC ae C30:2	0.100(0.097, 0.103) **100**	0.067 (0.063, 0.070) **-33**	**1×10^{-34}**
PC ae C32:1	3.254 (3.171, 3.339) **100**	2.442 (2.332, 2.558) **-25**	**4×10^{-22}**
PC ae C32:2	0.896 (0.873, 0.920) **100**	0.631 (0.603, 0.661) **-30**	**3×10^{-30}**
PC ae C34:0	1.708 (1.649, 1.770) **100**	0.731 (0.686, 0.778) **-57**	**1×10^{-69}**
PC ae C34:1	11.042 (10.767, 11.323) **100**	8.198 (7.837, 8.575) **-26**	**2×10^{-24}**
PC ae C34:2	13.005 (12.667, 13.351) **100**	9.776 (9.328, 10.246) **-25**	**2×10^{-21}**
PC ae C34:3	9.654 (9.377, 9.938) **100**	8.864 (8.416, 9.335) **-8**	**0.006**
PC ae C36:1	8.216 (7.961, 8.478) **100**	4.111 (3.887, 4.348) **-50**	**1×10^{-62}**
PC ae C36:2	16.391 (15.908, 16.887) **100**	9.828 (9.317, 10.366) **-40**	**9×10^{-44}**
PC ae C36:3	8.130 (7.936, 8.328) **100**	7.518 (7.201, 7.848) **-8**	**0.003**
PC ae C36:4	13.905 (13.559, 14.259) **100**	12.080 (11.549, 12.635) **-13**	**3×10^{-07}**
PC ae C36:5	9.238 (8.964, 9.521) **100**	7.775 (7.368, 8.205) **-16**	**2×10^{-07}**

Continues...

Table 5.7 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to non-vegans**		$p_{difference}$ [2]
	Non-Vegans **n = 283**	**Vegans** **n = 96**	
PC ae C38:0	2.301 (2.216, 2.388) **100**	1.332 (1.246, 1.424) **-42**	**2×10^{-34}**
PC ae C38:2	2.230 (2.172, 2.291) **100**	1.728 (1.647, 1.812) **-23**	**2×10^{-17}**
PC ae C38:3	3.924 (3.817, 4.033) **100**	2.367 (2.254, 2.486) **-40**	**2×10^{-48}**
PC ae C38:4	10.605 (10.329, 10.888) **100**	8.099 (7.726, 8.489) **-24**	**2×10^{-19}**
PC ae C38:5	13.829 (13.495, 14.172) **100**	14.226 (13.618, 14.862) **3**	0.3
PC ae C38:6	5.875 (5.685, 6.073) **100**	4.253 (4.009, 4.511) **-28**	**5×10^{-18}**
PC ae C40:1	1.265 (1.227, 1.305) **100**	0.889 (0.841, 0.939) **-30**	**2×10^{-23}**
PC ae C40:2	1.592 (1.545, 1.640) **100**	0.817 (0.775, 0.863) **-49**	**6×10^{-63}**
PC ae C40:3	0.994 (0.973, 1.015) **100**	0.741 (0.713, 0.769) **-25**	**2×10^{-31}**
PC ae C40:4	1.910 (1.867, 1.954) **100**	1.656 (1.590, 1.724) **-13**	**1×10^{-08}**
PC ae C40:5	3.018 (2.947, 3.091) **100**	2.533 (2.428, 2.643) **-16**	**3×10^{-11}**
PC ae C40:6	3.688 (3.570, 3.809) **100**	2.321 (2.190, 2.460) **-37**	**5×10^{-33}**
PC ae C42:1	0.328 (0.320, 0.337) **100**	0.343 (0.327, 0.359) **5**	0.1
PC ae C42:2	0.496 (0.482, 0.510) **100**	0.377 (0.359, 0.396) **-24**	**4×10^{-18}**
PC ae C42:3	0.796 (0.780, 0.814) **100**	0.628 (0.604, 0.652) **-21**	**3×10^{-22}**
PC ae C42:4	0.746 (0.727, 0.764) **100**	0.707 (0.676, 0.739) **-5**	0.05
PC ae C42:5	1.657 (1.620, 1.695) **100**	1.484 (1.425, 1.545) **-10**	**8×10^{-06}**
PC ae C44:3	0.163 (0.159, 0.167) **100**	0.148 (0.143, 0.155) **-9**	**2×10^{-04}**
PC ae C44:4	0.330 (0.322, 0.337) **100**	0.293 (0.282, 0.305) **-11**	**1×10^{-06}**
PC ae C44:5	1.221 (1.177, 1.247) **100**	1.085 (1.030, 1.142) **-11**	**4×10^{-04}**
PC ae C44:6	0.998 (0.967, 1.030) **100**	1.067 (1.009, 1.130) **7**	0.05

Continues...

Table 5.7 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to non-vegans**		$p_{difference}$ [2]
	Non-Vegans **n = 283**	**Vegans** **n = 96**	
Hexose			
Hexose	2194 (2099, 2293) **100**	2082 (1924, 2253) **-5**	0.3
Sphingolipids			
SM (OH) C14:1	5.718 (5.525, 5.918) **100**	2.255 (2.123, 2.398) **-61**	**2 x 10^{-81}**
SM (OH) C16:1	2.806 (2.716, 2.899) **100**	1.320 (1.245, 1.399) **-53**	**2 x 10^{-66}**
SM (OH) C22:1	8.932 (8.679, 9.193) **100**	6.684 (6.349, 7.036) **-25**	**7 x 10^{-19}**
SM (OH) C22:2	6.810 (6.619, 7.007) **100**	4.398 (4.180, 4.628) **-35**	**5 x 10^{-37}**
SM (OH) C24:1	0.743 (0.721, 0.767) **100**	0.517 (0.489, 0.547) **-30**	**4 x 10^{-24}**
SM C16:0	99.826 (97.891, 101.799) **100**	90.726 (87.608, 93.955) **-9**	**7 x 10^{-06}**
SM C16:1	14.011 (13.701, 14.328) **100**	13.600 (13.067, 14.154) **-3**	0.2
SM C18:0	17.122 (16.605, 17.656) **100**	11.622 (11.002, 12.277) **-32**	**3 x 10^{-27}**
SM C18:1	8.369 (8.135, 8.611) **100**	7.153 (6.799, 7.526) **-15**	**4 x 10^{-07}**
SM C20:2	0.544 (0.520, 0.569) **100**	0.417 (0.385, 0.452) **-23**	**8 x 10^{-08}**
SM C24:0	13.012 (12.657, 13.377) **100**	13.348 (12.705, 14.023) **3**	0.4
SM C24:1	26.455 (25.792, 27.135) **100**	24.263 (23.189, 25.389) **-8**	**0.002**
SM C26:0	0.106 (0.103, 0.110) **100**	0.088 (0.083, 0.093) **-17**	**5 x 10^{-08}**
SM C26:1	0.198 (0.191, 0.205) **100**	0.166 (0.156, 0.176) **-16**	**1 x 10^{-08}**

[1] Adjusted for age ($30-34$; $35-39$; $40-44$; $45-49$ years), BMI (<22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), smoking status (never; former; current), alcohol intake (<1; $1-7$; $8-15$; ≥ 16 g/d), time since last food or drink at blood collection (<1.5; $1.5-2.9$; $3.0-4.4$; ≥ 4.5 hours; unknown), and time between blood collection and processing (fourths of the distribution; unknown). Non-vegans are meat-eaters, fish-eaters and vegetarians

[2] p-values refer to p_{diff} across the two diet groups calculated by ANOVA. Conventional p-values are shown and those marked in bold were significant after correcting for multiple testing using a false discovery rate procedure (Benjamini–Hochberg) at $\alpha = 0.05$

Table 5.8: Geometric mean metabolite concentrations in non-vegan diet groups and percentage difference compared to meat-eaters in EPIC-Oxford

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$[2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
Acylcarnitines				
C0	34.612 (32.947, 36.360) **100**	33.500 (31.949, 35.127) **-3**	32.743 (31.228, 34.331) **-5**	0.3
C14:1	0.061 (0.056, 0.067) **100**	0.060 (0.055, 0.066) **-2**	0.061 (0.056, 0.066) **0**	>0.9
C16	0.152 (0.138, 0.167) **100**	0.167 (0.153, 0.183) **10**	0.157 (0.144, 0.172) **3**	0.4
C18:1	0.224 (0.205, 0.244) **100**	0.241 (0.222, 0.262) **8**	0.234 (0.216, 0.255) **4**	0.5
C18:2	0.138 (0.124, 0.152) **100**	0.163 (0.148, 0.179) **18**	0.161 (0.146, 0.177) **17**	0.05
C2	4.803 (4.482, 5.146) **100**	4.547 (4.255, 4.859) **-5**	4.576 (4.282, 4.890) **-5**	0.5
C3	0.316 (0.296, 0.339) **100**	0.292 (0.274, 0.312) **-8**	0.316 (0.296, 0.337) **0**	0.2
C4	0.160 (0.151, 0.173) **100**	0.150 (0.141, 0.160) **-6**	0.148 (0.139, 0.158) **-8**	0.2
C5	0.133 (0.125, 0.142) **100**	0.122 (0.115, 0.130) **-8**	0.122 (0.114, 0.130) **-8**	0.1
Amino acids				
Isoleucine	99 (95, 104) **100**	98 (94, 103) **-1**	101 (97, 106) **2**	0.6
Leucine	206 (197, 215) **100**	206 (198, 215) **0**	210 (202, 219) **2**	0.7
Valine	232 (222, 241) **100**	233 (224, 242) **0**	238 (229, 247) **3**	0.6
Histidine	114 (109, 118) **100**	121 (117, 126) **6**	120 (116, 124) **5**	0.04
Lysine	241 (230, 252) **100**	241 (231, 253) **0**	237 (227, 248) **-2**	0.9
Methionine	29 (28, 31) **100**	30 (28, 31) **3**	31 (30, 32) **7**	0.2

Continues...

Table 5.8 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$ [2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
Threonine	164 (158, 170) **100**	167 (161, 173) **2**	169 (163, 175) **3**	0.6
Tryptophan	67 (65, 70) **100**	70 (68, 73) **4**	71 (68, 73) **6**	0.2
Alanine	569 (544, 596) **100**	640 (613, 669) **12**	635 (608, 663) **12**	**8×10^{-04}**
Arginine	50 (45, 56) **100**	44 (40, 49) **-12**	42 (38, 46) **-16**	0.07
Asparagine	93 (90, 96) **100**	97 (93, 100) **4**	96 (93, 100) **3**	0.2
Aspartate	64 (60, 67) **100**	62 (58, 65) **-3**	67 (63, 71) **5**	0.1
Glutamate	260 (245, 275) **100**	297 (282, 314) **14**	286 (271, 302) **10**	**0.005**
Glutamine	508 (489, 528) **100**	498 (480, 516) **2**	521 (502, 540) **3**	0.2
Glycine	392 (376, 410) **100**	418 (401, 436) **7**	415 (398, 433) **6**	0.1
Serine	189 (182, 197) **100**	200 (192, 208) **6**	198 (191, 206) **5**	0.2
Tyrosine	78 (75, 81) **100**	82 (79, 86) **5**	83 (79, 86) **6**	0.06
Citrulline	37 (35, 39) **100**	37 (35, 39) **0**	41 (39, 44) **10**	**0.008**
Ornithine	183 (174, 191) **100**	194 (186, 203) **6**	196 (188, 205) **7**	0.09
Biogenic amines				
Creatinine	80.87 (78.48, 83.33) **100**	76.38 (74.21, 78.61) **-6**	76.17 (74.01, 78.40) **-6**	**0.01**
Kynurenine	2.73 (2.60, 2.87) **100**	2.85 (2.72, 2.98) **4**	2.99 (2.86, 3.14) **10**	0.03

Continues...

Table 5.8 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$ [2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
Serotonin	0.37 (0.33, 0.41) **100**	0.37 (0.33, 0.41) **0**	0.35 (0.32, 0.39) **-5**	0.8
Glycerophospholipids				
lysoPC a C16:0	209.186 (199.716, 219.105) **100**	222.488 (212.789, 232.629) **6**	213.900 (204.583, 223.642) **2**	0.2
lysoPC a C16:1	5.772 (5.426, 6.141) **100**	6.003 (5.657, 6.372) **4**	6.147 (5.792, 6.524) **6**	0.4
lysoPC a C17:0	4.467 (4.162, 4.795) **100**	4.473 (4.179, 4.789) **0**	4.094 (3.824, 4.382) **-8**	0.1
lysoPC a C18:0	75.224 (71.153, 79.528) **100**	80.767 (76.557, 85.208) **7**	77.023 (73.0.11, 81.254) **1**	0.2
lysoPC a C18:1	44.163 (41.729, 46.739) **100**	45.426 (43.015, 47.973) **3**	45.378 (42.969, 47.918) **3**	0.8
lysoPC a C18:2	59.829 (56.194, 63.700) **100**	57.329 (53.974, 60.893) **-4**	59.405 (55.931, 63.094) **-1**	0.6
lysoPC a C20:3	3.616 (3.398, 3.849) **100**	3.696 (3.481, 3.925) **2**	3.913 (3.685, 4.154) **8**	0.2
lysoPC a C20:4	9.879 (9.264, 10.535) **100**	9.664 (9.085, 10.281) **-2**	10.236 (9.623, 10.889) **4**	0.4
PC aa C28:1	2.708 (2.536, 2.892) **100**	2.791 (2.620, 2.972) **3**	2.666 (2.503, 2.839) **-2**	0.6
PC aa C30:0	3.665 (3.420, 3.928) **100**	3.575 (3.345, 3.822) **-2**	3.453 (3.231, 3.691) **-6**	0.5
PC aa C32:0	18.234 (17.421, 19.084) **100**	17.908 (17.140, 18.711) **-2**	16.996 (16.268, 17.758) **-7**	0.08
PC aa C32:1	14.358 (12.825, 16.075) **100**	13.930 (12.496, 15.529) **-3**	14.945 (13.407, 16.659) **4**	0.7
PC aa C32:2	5.269 (4.858, 5.713) **100**	5.527 (5.113, 5.975) **5**	6.010 (5.559, 6.496) **14**	0.07
PC aa C32:3	0.840 (0.799, 0.882) **100**	0.835 (0.797, 0.876) **-1**	0.824 (0.786, 0.864) **-2**	0.9

Continues...

Table 5.8 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$[2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
PC aa C34:1	190.468 (182.932, 198.315) **100**	186.625 (179.515, 194.016) **-2**	190.625 (183.369, 198.168) **0**	0.7
PC aa C34:2	228.491 (219.932, 237.384) **100**	228.900 (220.644, 237.465) **0**	225.860 (217.720, 234.304) **-1**	0.9
PC aa C34:3	20.778 (19.499, 22.139) **100**	20.810 (19.577, 22.121) **0**	22.442 (21.113, 23.854) **8**	0.1
PC aa C34:4	1.930 (1.786, 2.086) **100**	1.934 (1.794, 2.083) **0**	2.082 (1.932, 2.243) **8**	0.3
PC aa C36:1	51.267 (48.123, 54.617) **100**	50.398 (47.421, 53.562) **-2**	50.629 (47.640, 53.805) **-1**	>0.9
PC aa C36:2	202.144 (196.006, 208.476) **100**	201.773 (195.875, 207.850) **0**	204.753 (198.772, 210.914) **1**	0.8
PC aa C36:3	137.369 (132.183, 142.758) **100**	139.657 (134.581, 144.924) **2**	151.543 (146.039, 157.254) **10**	**7×10^{-04}**
PC aa C36:4	159.088 (152.454, 166.011) **100**	155.344 (149.107, 161.842) **-2**	168.056 (161.314, 175.081) **6**	0.02
PC aa C36:5	27.565 (24.698, 30.764) **100**	25.484 (22.928, 28.324) **-8**	18.777 (16.896, 20.868) **-32**	**2×10^{-06}**
PC aa C36:6	1.135 (1.036, 1.245) **100**	1.070 (0.980, 1.169) **-6**	0.734 (0.672, 0.801) **-35**	**2×10^{-11}**
PC aa C38:0	2.736 (2.584, 2.896) **100**	2.438 (2.308, 2.576) **-11**	1.641 (1.554, 1.734) **-40**	**5×10^{-30}**
PC aa C38:3	39.995 (37.841, 42.272) **100**	41.403 (39.255, 43.668) **4**	42.271 (40.080, 44.582) **6**	0.4
PC aa C38:4	77.935 (73.651, 82.468) **100**	76.215 (72.181, 80.476) **-2**	85.337 (80.823, 90.104) **9**	**0.01**
PC aa C38:5	50.244 (47.453, 53.200) **100**	49.288 (46.651, 50.914) **-2**	50.914 (48.191, 53.790) **1**	0.7
PC aa C38:6	79.829 (74.522, 85.514) **100**	75.064 (70.257, 80.200) **-6**	46.649 (43.663, 49.838) **-42**	**3×10^{-26}**
PC aa C40:2	0.329 (0.310, 0.348) **100**	0.316 (0.299, 0.334) **-4**	0.287 (0.271, 0.304) **-13**	**0.005**

Continues...

Table 5.8 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$ [2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
PC aa C40:3	0.580 (0.549, 0.613) **100**	0.548 (0.520, 0.578) **-6**	0.513 (0.487, 0.541) **-12**	**0.009**
PC aa C40:4	2.429 (2.293, 2.572) **100**	2.436 (2.306, 2.574) **0**	2.944 (2.786, 3.111) **21**	**6 x 10^{-07}**
PC aa C40:5	6.968 (6.535, 7.430) **100**	6.998 (6.579, 7.443) **0**	7.637 (7.180, 8.123) **10**	0.07
PC aa C40:6	23.319 (21.663, 25.100) **100**	22.272 (20.749, 23.907) **-4**	13.772 (12.831, 14.782) **-41**	**2 x 10^{-23}**
PC aa C42:0	0.595 (0.564, 0.627) **100**	0.583 (0.554, 0.614) **-2**	0.475 (0.452, 0.500) **-20**	**1 x 10^{-11}**
PC aa C42:1	0.312 (0.297, 0.327) **100**	0.315 (0.301, 0.330) **1**	0.259 (0.248, 0.271) **-17**	**2 x 10^{-09}**
PC aa C42:2	0.229 (0.218, 0.240) **100**	0.235 (0.225, 0.246) **3**	0.193 (0.184, 0.202) **-16**	**3 x 10^{-09}**
PC aa C42:4	0.144 (0.138, 0.151) **100**	0.145 (0.139, 0.152) **1**	0.147 (0.141, 0.154) **2**	0.8
PC aa C42:5	0.313 (0.297, 0.329) **100**	0.299 (0.284, 0.314) **-4**	0.293 (0.279, 0.308) **-6**	0.2
PC ae C30:2	0.101 (0.095, 0.106) **100**	0.101 (0.096, 0.107) **0**	0.097 (0.092, 0.103) **-4**	0.5
PC ae C32:1	3.513 (3.347, 3.686) **100**	3.227 (3.081, 3.381) **-8**	3.072 (2.933, 3.218) **-13**	**8 x 10^{-04}**
PC ae C32:2	0.986 (0.941, 1.033) **100**	0.898 (0.859, 0.940) **-9**	0.812 (0.776, 0.849) **-18**	**2 x 10^{-07}**
PC ae C34:0	1.862 (1.737, 1.997) **100**	1.727 (1.615, 1.847) **-7**	1.564 (1.463, 1.673) **-16**	**0.003**
PC ae C34:1	11.336 (10.789, 11.911) **100**	10.867 (10.362, 11.397) **-4**	10.991 (10.480, 11.526) **-3**	0.5
PC ae C34:2	14.274 (13.595, 14.986) **100**	12.276 (11.714, 12.865) **-14**	12.473 (11.902, 13.071) **-13**	**6 x 10^{-05}**
PC ae C34:3	10.700 (10.138, 11.293) **100**	9.319 (8.848, 9.815) **-13**	9.040 (8.583, 9.521) **-16**	**7 x 10^{-05}**

Continues...

Table 5.8 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$[2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
PC ae C36:1	8.613 (8.089, 9.170) **100**	8.188 (7.709, 8.698) **-5**	7.941 (7.476, 8.435) **-8**	0.2
PC ae C36:2	17.198 (16.247, 18.205) **100**	15.806 (14.964, 16.695) **-8**	16.132 (15.273, 17.039) **-6**	0.1
PC ae C36:3	8.662 (8.284, 9.058) **100**	7.727 (7.402, 8.066) **-11**	7.998 (7.662, 8.349) **-8**	**0.003**
PC ae C36:4	15.411 (14.720, 16.134) **100**	12.918 (12.361, 13.501) **-16**	13.758 (13.165, 14.378) **-11**	**3×10^{-06}**
PC ae C36:5	10.769 (10.222, 11.345) **100**	8.845 (8.412, 9.300) **-18**	8.478 (8.064, 8.914) **-21**	**2×10^{-09}**
PC ae C38:0	2.617 (2.456, 2.789) **100**	2.514 (2.365, 2.673) **-4**	1.829 (1.721, 1.944) **-30**	**2×10^{-15}**
PC ae C38:2	2.299 (2.188, 2.415) **100**	2.190 (2.089, 2.297) **-5**	2.198 (2.096, 2.304) **-4**	0.3
PC ae C38:3	4.015 (3.803, 4.239) **100**	3.830 (3.635, 4.035) **-5**	3.952 (3.751, 4.164) **-2**	0.5
PC ae C38:4	11.043 (10.504, 11.609) **100**	9.927 (9.461, 10.417) **-10**	10.990 (10.474, 11.531) **0**	**0.004**
PC ae C38:5	14.659 (14.016, 15.330) **100**	13.311 (12.749, 13.897) **-9**	13.701 (13.124, 14.305) **-7**	**0.01**
PC ae C38:6	7.103 (6.730, 7.496) **100**	6.042 (5.737, 6.363) **-15**	4.731 (4.493, 4.983) **-33**	**6×10^{-21}**
PC ae C40:1	1.389 (1.321, 1.460) **100**	1.355 (1.291, 1.422) **-2**	1.060 (1.011, 1.113) **-24**	**1×10^{-14}**
PC ae C40:2	1.745 (1.652, 1.844) **100**	1.634 (1.550, 1.723) **-6**	1.421 (1.348, 1.498) **-19**	**1×10^{-06}**
PC ae C40:3	1.010 (0.969, 1.052) **100**	0.992 (0.953, 1.032) **-2**	0.974 (0.937, 1.013) **-4**	0.5
PC ae C40:4	1.923 (1.843, 2.007) **100**	1.840 (1.766, 1.917) **-4**	1.962 (1.883, 2.044) **2**	0.09
PC ae C40:5	3.116 (2.978, 3.260) **100**	2.948 (2.822, 3.079) **-5**	2.978 (2.851, 3.110) **-4**	0.2

Continues...

Table 5.8 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$[2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
PC ae C40:6	4.265 (4.033, 4.510) **100**	3.892 (3.688, 4.107) **-9**	2.964 (2.809, 3.128) **-31**	**9×10^{-18}**
PC ae C42:1	0.324 (0.309, 0.339) **100**	0.325 (0.311, 0.340) **0**	0.336 (0.321, 0.352) **4**	0.5
PC ae C42:2	0.514 (0.487, 0.542) **100**	0.503 (0.477, 0.530) **-2**	0.472 (0.448, 0.497) **-8**	0.07
PC ae C42:3	0.828 (0.796, 0.860) **100**	0.829 (0.799, 0.860) **0**	0.724 (0.697, 0.751) **-13**	**1×10^{-07}**
PC ae C42:4	0.741 (0.708, 0.777) **100**	0.733 (0.701, 0.766) **-1**	0.749 (0.716, 0.783) **1**	0.8
PC ae C42:5	1.694 (1.622, 1.769) **100**	1.634 (1.567, 1.704) **-4**	1.635 (1.568, 1.705) **-3**	0.5
PC ae C44:3	0.160 (0.153, 0.167) **100**	0.169 (0.162, 0.177) **6**	0.159 (0.153, 0.166) **-1**	0.09
PC ae C44:4	0.323 (0.309, 0.338) **100**	0.325 (0.311, 0.339) **1**	0.336 (0.323, 0.351) **4**	0.4
PC ae C44:5	1.229 (1.166, 1.296) **100**	1.199 (1.140, 1.262) **-2**	1.193 (1.134, 1.255) **-3**	0.7
PC ae C44:6	0.996 (0.939, 1.056) **100**	1.010 (0.955, 1.069) **1**	0.969 (0.916, 1.026) **-3**	0.6
Hexose				
Hexose	2296 (2111, 2497) **100**	2153 (1986, 2333) **-6**	2151 (1984, 2331) **-6**	0.5
Sphingolipids				
SM (OH) C14:1	5.960 (5.577, 6.370) **100**	5.713 (5.360, 6.010) **-4**	5.470 (5.132, 5.831) **-8**	0.2
SM (OH) C16:1	3.066 (2.882, 3.262) **100**	2.803 (2.641, 2.975) **-9**	2.562 (2.414, 2.719) **-16**	**4×10^{-04}**
SM (OH) C22:1	9.115 (8.627, 9.630) **100**	9.065 (8.598, 9.558) **-1**	8.621 (8.177, 9.089) **-5**	0.3
SM (OH) C22:2	7.123 (6.746, 7.521) **100**	6.948 (6.594, 7.320) **-2**	6.336 (6.014, 6.676) **-11**	**0.006**

Continues...

Table 5.8 continued

	Geometric mean concentration (95% confidence interval), µmol/L **% change in concentration, compared to meat-eaters**			$p_{difference}$ [2]
	Meat-eaters **n = 95**	**Fish-eaters** **n = 97**	**Vegetarians** **n = 91**	
SM (OH) C24:1	0.747 (0.704, 0.792) **100**	0.758 (0.716, 0.802) **1**	0.730 (0.690, 0.772) **-2**	0.6
SM C16:0	103.642 (99.906, 107.518) **100**	99.929 (96.461, 103.522) **-4**	96.210 (92.874, 99.667) **-7**	0.02
SM C16:1	14.267 (13.678, 14.882) **100**	14.163 (13.601, 14.750) **-1**	13.781 (13.234, 14.351) **-3**	0.5
SM C18:0	18.617 (17.618, 19.672) **100**	17.327 (16.431, 18.271) **-7**	15.874 (15.054, 16.738) **-15**	**4×10^{-04}**
SM C18:1	8.866 (8.430, 9.325) **100**	8.417 (8.019, 8.836) **-5**	7.955 (7.578, 8.350) **-10**	**0.01**
SM C20:2	0.598 (0.550, 0.650) **100**	0.520 (0.480, 0.563) **-13**	0.525 (0.485, 0.569) **-12**	0.05
SM C24:0	12.980 (12.302, 13.696) **100**	13.273 (12.605, 13.976) **2**	12.976 (12.323, 13.663) **0**	0.8
SM C24:1	28.402 (27.106, 29.760) **100**	27.131 (25.939, 28.377) **-4**	24.207 (23.144, 25.318) **-15**	**8×10^{-06}**
SM C26:0	0.110 (0.103, 0.117) **100**	0.106 (0.100, 0.113) **-4**	0.104 (0.098, 0.110) **-5**	0.4
SM C26:1	0.220 (0.207, 0.234) **100**	0.204 (0.193, 0.216) **-7**	0.172 (0.162, 0.182) **-22**	**9×10^{-08}**

[1] Adjusted for age ($30-34$; $35-39$; $40-44$; $45-49$ years), BMI (<22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), smoking status (never; former; current), alcohol intake (<1; $1-7$; $8-15$; ≥ 16 g/d), time since last food or drink at blood collection (<1.5; $1.5-2.9$; $3.0-4.4$; ≥ 4.5 hours; unknown), and time between blood collection and processing (fourths of the distribution; unknown)

[2] p-values refer to p_{diff} across the three diet groups calculated by ANOVA. Conventional p-values are shown and those marked in bold were significant after correcting for multiple testing using a false discovery rate procedure (Benjamini–Hochberg) at $\alpha = 0.05$

Table 5.9: Mean principal component score by diet group and percentage difference compared to meat-eaters in EPIC-Oxford – main model[1]

Principal component	Mean principal component score (95% confidence interval) % mean difference compared to meat-eaters				p_{diff}[2]
	Meat-eaters (n = 95)	**Fish-eaters (n = 97)**	**Vegetarians (n = 91)**	**Vegans (n = 96)**	
1					
Unadjusted	12.3 (12.0, 12.6)	12.0 (11.7, 12.2)	11.2 (10.9, 11.5)	8.5 (8.2, 8.8)	$4x10^{-61}$
	Ref.	**-3**	**-8**	**-31**	
Adjusted[3]	12.3 (12.0, 12.6)	11.9 (11.6, 12.2)	11.2 (10.9, 11.5)	8.6 (8.3, 8.9)	$8x10^{-52}$
	Ref.	**-3**	**-9**	**-30**	
2					
Unadjusted	3.0 (2.8, 3.2)	3.5 (3.2, 3.7)	3.4 (3.1, 3.6)	3.4 (3.1, 3.6)	0.01
	Ref.	**+7**	**+23**	**+20**	
Adjusted[3]	3.1 (2.9, 3.3)	3.4 (3.2, 3.6)	3.3 (3.1, 3.5)	3.3 (3.2, 3.5)	0.4
	Ref.	**+8**	**+6**	**+7**	
3					
Unadjusted	6.6 (6.5, 6.8)	6.6 (6.5, 6.8)	7.0 (6.8, 7.1)	7.2 (7.0, 7.3)	$2x10^{-6}$
	Ref.	**0**	**+5**	**+8**	
Adjusted[3]	6.5 (6.3, 6.7)	6.6 (6.5, 6.8)	7.0 (6.8, 7.1)	7.3 (7.1, 7.4)	$5x10^{-10}$
	Ref.	**+2**	**+7**	**+12**	
4					
Unadjusted	7.8 (7.7, 8.0)	7.6 (7.4, 7.7)	7.9 (7.7, 8.0)	8.4 (8.2, 8.5)	$1x10^{-11}$
	Ref.	**-3**	**+1**	**+7**	
Adjusted[3]	7.8 (7.7, 8.0)	7.6 (7.4, 7.7)	7.9 (7.7, 8.0)	8.3 (8.2, 8.5)	$8x10^{-10}$
	Ref.	**-3**	**0**	**+6**	

[1] Principal component scores were derived using principal component analysis based on the covariance matrix of log-transformed metabolite concentrations

[2] p-values refer to test for difference in component score across the four diet groups calculated using analysis of variance

[3] Adjusted for age ($30-34$; $35-39$; $40-44$; $45-49$ years), BMI (<22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), smoking status (never; former; current), alcohol intake (<1; $1-7$; $8-15$; ≥ 16 g/d), time since last food or drink at blood collection (<1.5; $1.5-2.9$; $3.0-4.4$; ≥ 4.5 hours; unknown), and time between blood collection and processing (fourths of the distribution; unknown)

Table 5.10: Mean principal component score by diet group and percentage difference compared to meat-eaters in EPIC-Oxford – alternative model[1]

Principal Component	Mean principal component score (95% confidence interval) % mean difference compared to meat-eaters				p_{diff}[2]
	Meat-eaters (n = 95)	**Fish-eaters (n = 97)**	**Vegetarians (n = 91)**	**Vegans (n = 96)**	
1					
Unadjusted	10.36 (5.72, 18.77)	3.30 (1.84, 5.94)	3.46 (1.89, 6.35)	0.01 (0.01, 0.02)	$7x10^{-52}$
	Ref.	**-68**	**-67**	**-100**	
Adjusted[3]	12.96 (6.70, 25.07)	2.96 (1.60, 5.45)	3.27 (1.75, 6.12)	0.01 (0.00, 0.02)	$1x10^{-45}$
	Ref.	**-77**	**-75**	**-100**	
2					
Unadjusted	13.89 (8.66, 22.28)	8.68 (5.44, 13.85)	0.40 (0.25, 0.65)	0.02 (0.01, 0.03)	$3x10^{-67}$
	Ref.	**-38**	**-97**	**-100**	
Adjusted[3]	14.13 (8.42, 23.70)	7.79 (4.82, 12.57)	0.42 (0.25, 0.68)	0.02 (0.01, 0.03)	$1x10^{-59}$
	Ref.	**-45**	**-97**	**-100**	
3					
Unadjusted	1.90 (1.03, 3.48)	1.64 (0.90, 2.98)	3.48 (1.88, 6.47)	0.10 (0.05, 0.18)	$1x10^{-15}$
	Ref.	**-14**	**+84**	**-95**	
Adjusted[3]	1.48 (0.78, 2.79)	1.62 (0.90, 2.91)	3.23 (1.77, 5.92)	0.14 (0.07, 0.25)	$1x10^{-11}$
	Ref.	**+9**	**+119**	**-91**	

[1] Principal component scores were derived using principal component analysis based on the correlation matrix of log-transformed and z-score standardised metabolite concentrations and varimax rotation

[2] p-values refer to test for difference in component score across the four diet groups calculated using analysis of variance

[3] Adjusted for age ($30-34$; $35-39$; $40-44$; $45-49$ years), BMI (<22.5; $22.5-24.9$; ≥ 25 kg/m^2; unknown), smoking status (never; former; current), alcohol intake (<1; $1-7$; $8-15$; ≥ 16 g/d), time since last food or drink at blood collection (<1.5; $1.5-2.9$; $3.0-4.4$; ≥ 4.5 hours; unknown), and time between blood collection and processing (fourths of the distribution; unknown)

5.8 Figures

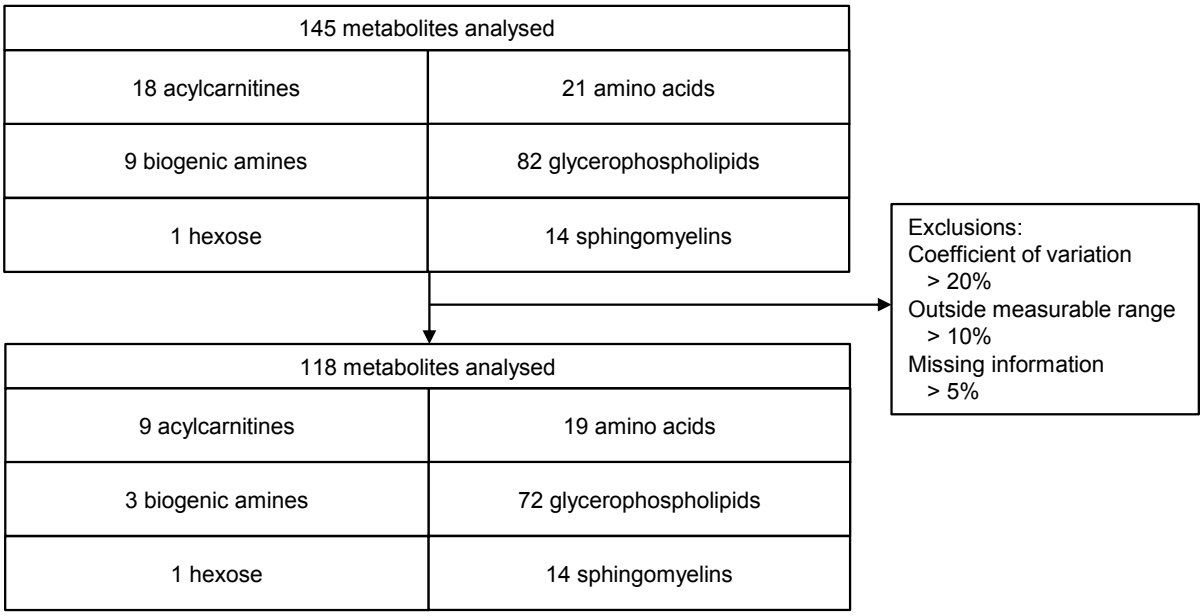


Figure 5.1: Exclusion of metabolites

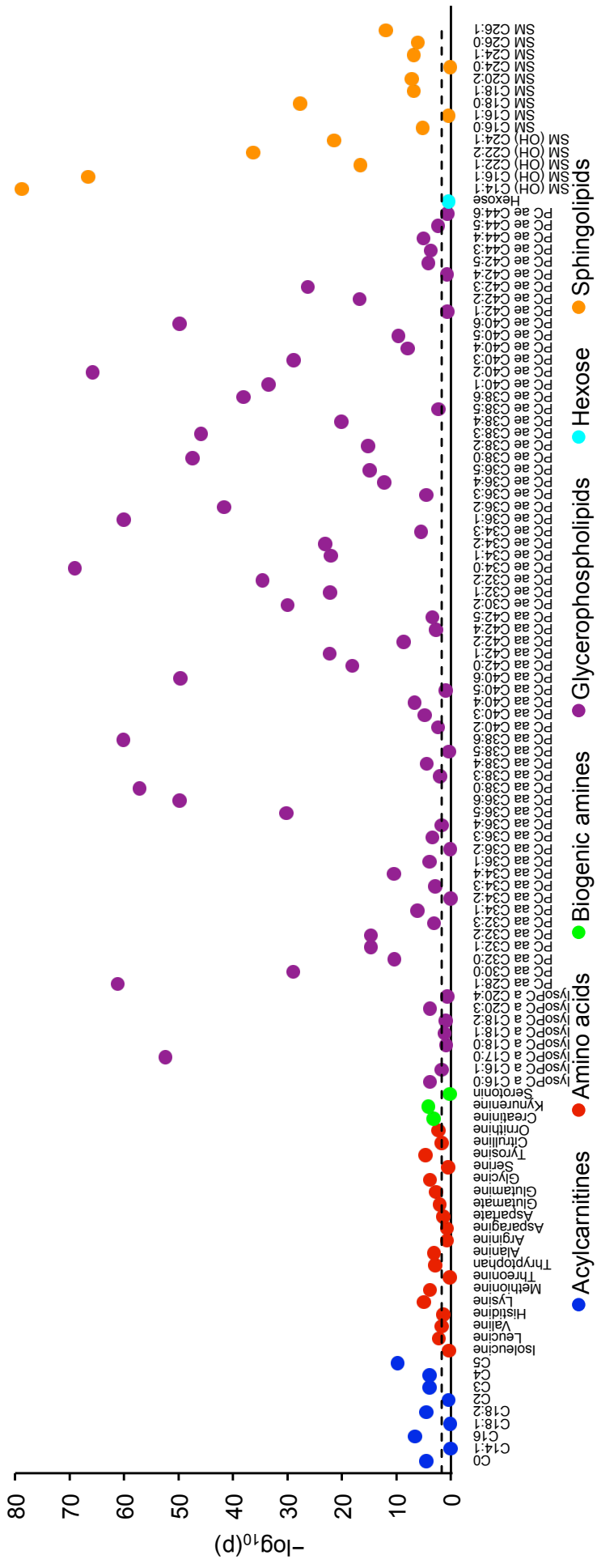


Figure 5.2: Statistical significance of the associations between diet group and metabolite concentrations in 379 men from the EPIC-Oxford

Statistical significance was plotted as $-\log10$(p-values). The dashed line represents statistical significance after controlling the false discovery rate (Benjamini–Hochberg) at $\alpha = 0.05$. The p-values were derived from an analysis of variance comparing geometric mean metabolite concentration across the four diet groups, adjusted for age ($30 - 34$; $35 - 39$; $40 - 44$; $45 - 49$ years), BMI ($<$ 22.5; $22.5 - 24.9$; ≥ 25 kg/m^2; unknown), smoking status (never; former; current), alcohol intake (< 1; $1 - 7$; $8 - 15$; ≥ 16 g/d), time since last food or drink at blood collection ($<$1.5; $1.5 - 2.9$; $3.0 - 4.4$; ≥ 4.5 h; unknown), and time between blood collection and processing (fourths of the distribution corresponding to < 25; $25 - 41$; $41 - 72$; ≥ 72 hours; unknown). The geometric mean and 95%CI of metabolite concentrations by diet group are shown in Table 5.6 on page 81.

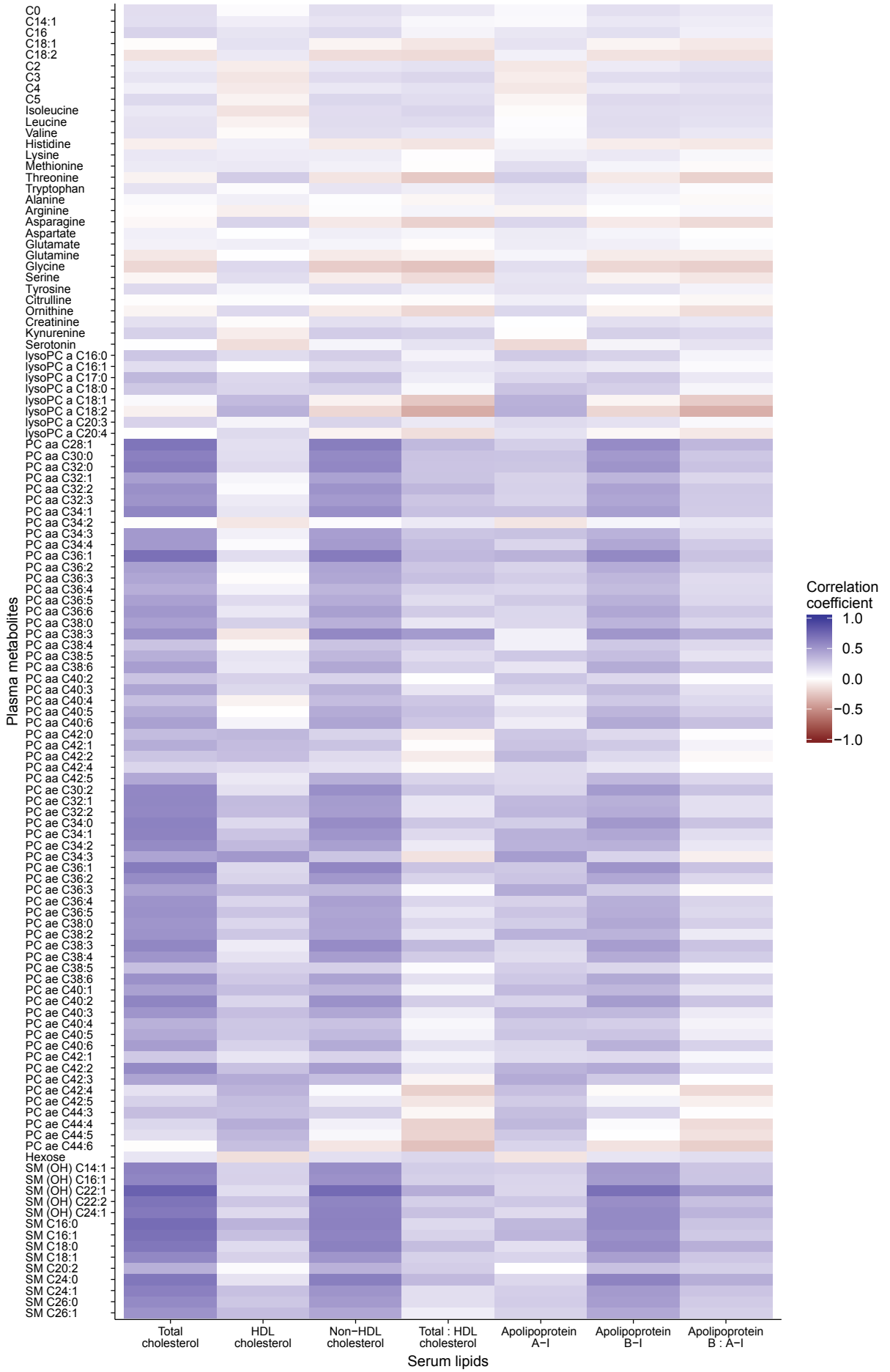


Figure 5.3: Correlation coefficients of metabolites with cholesterols and apolipoproteins

The colours indicate the strength and direction of the Spearman's rank correlation coefficients between concentrations of plasma metabolites and serum lipids.

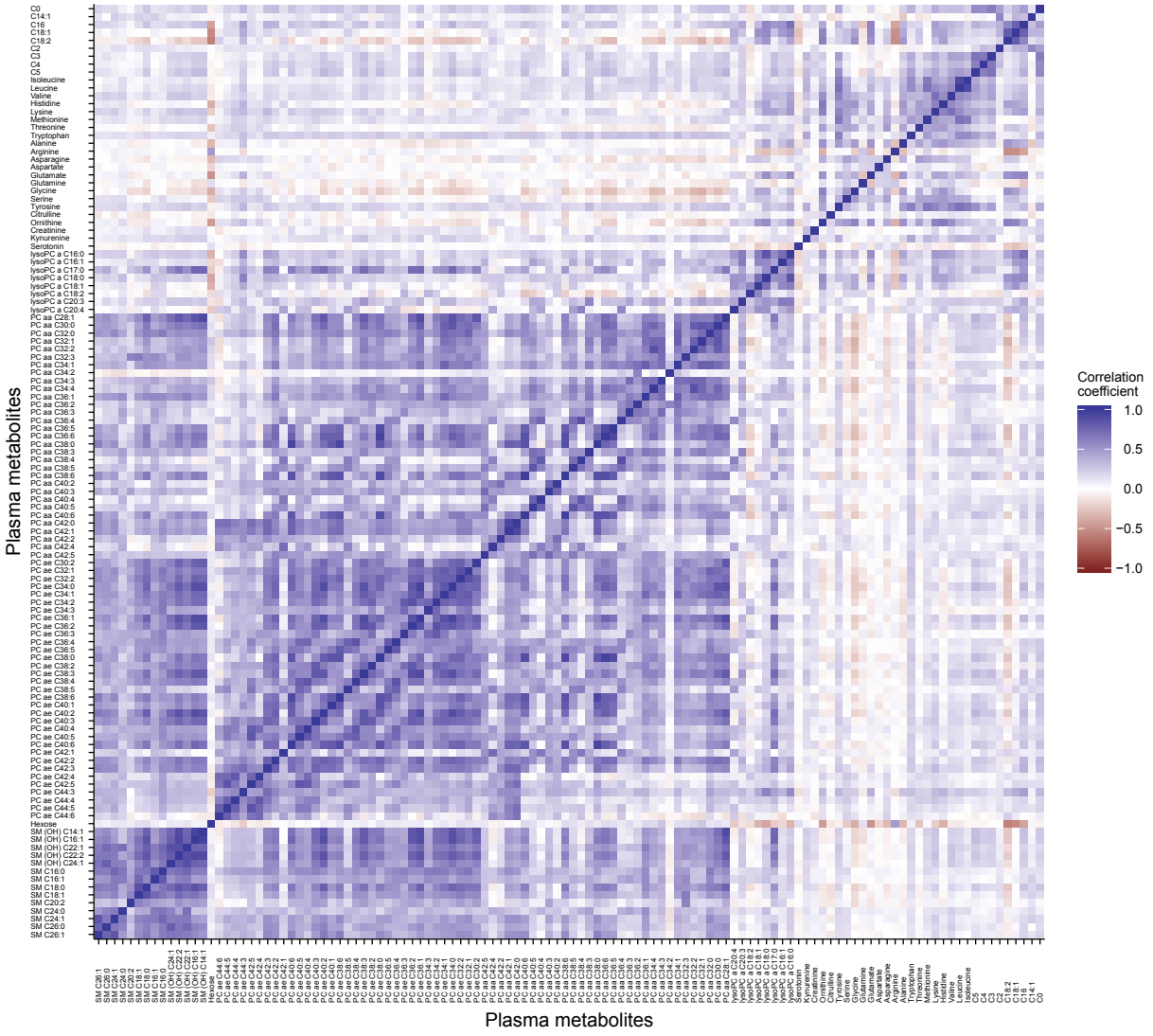


Figure 5.4: Correlation coefficients between plasma metabolite concentrations

The colours indicate the strength and direction of the Spearman's rank correlation coefficients between pairs of plasma metabolite concentrations.

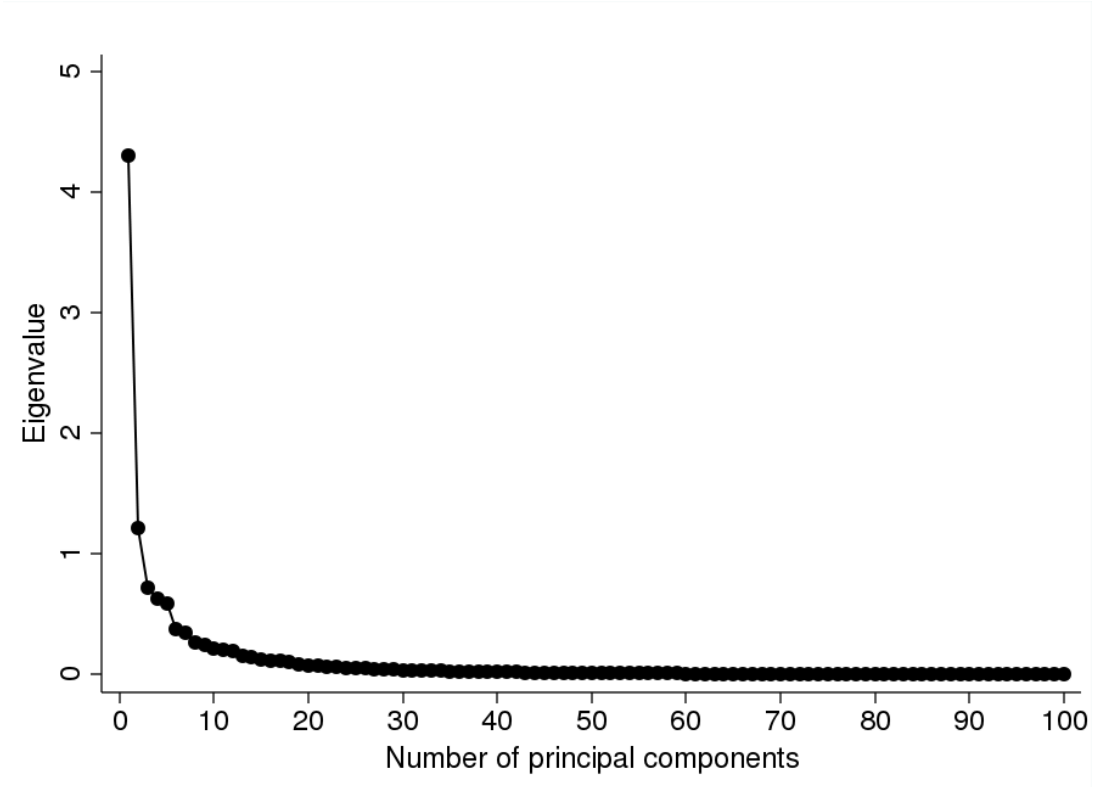


Figure 5.5: Scree plot for PCA of metabolite concentrations – main model

Metabolite concentrations were log-transformed and covariance matrix was used.

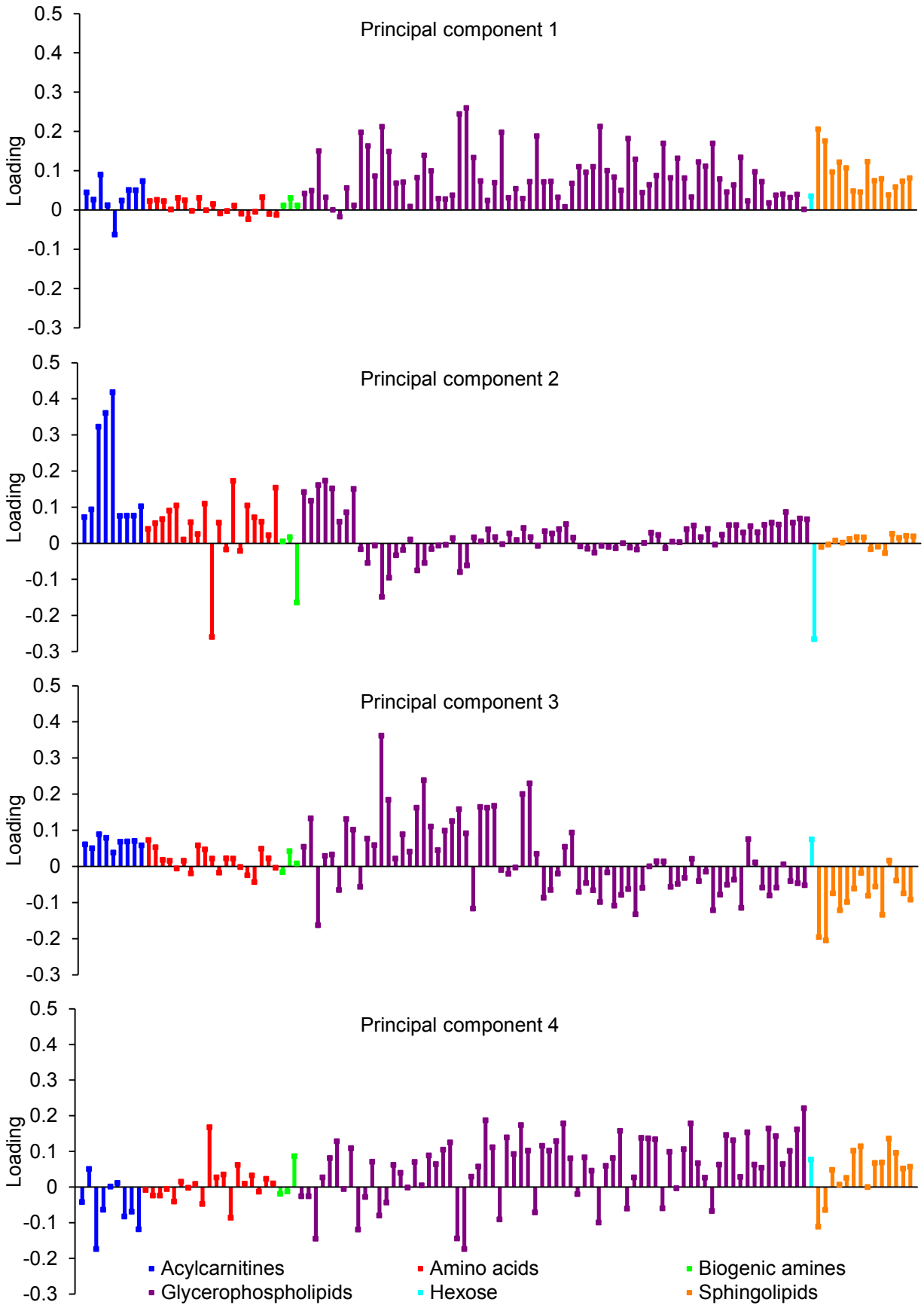


Figure 5.6: Loadings for the first four principal components by metabolite class – main model
Loadings were derived using principal component analysis based on the covariance matrix of log-transformed metabolite concentrations.

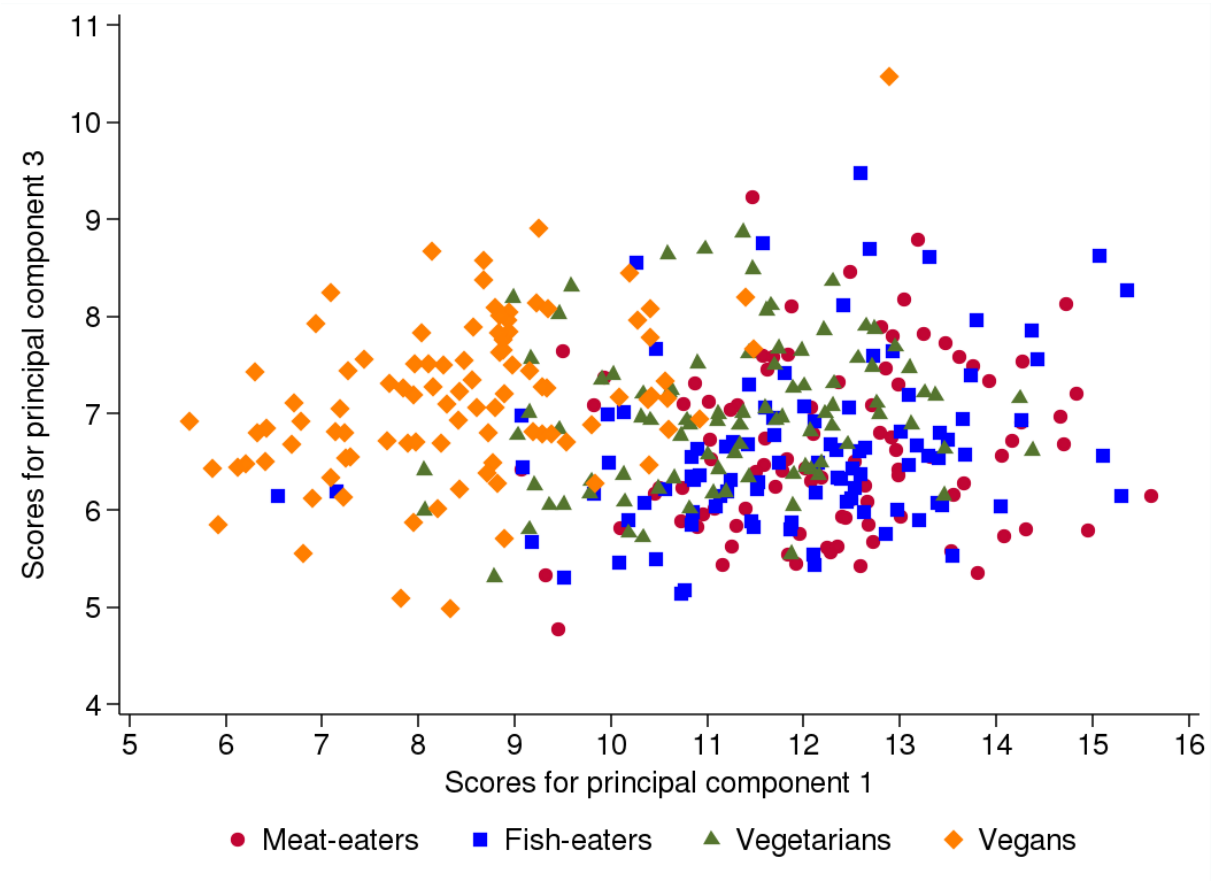


Figure 5.7: Score plot for principal component 1 and 3 by diet group – main model
Each point represents a participant. Scores were derived using principal component analysis based on the covariance matrix of log-transformed metabolite concentrations [156].

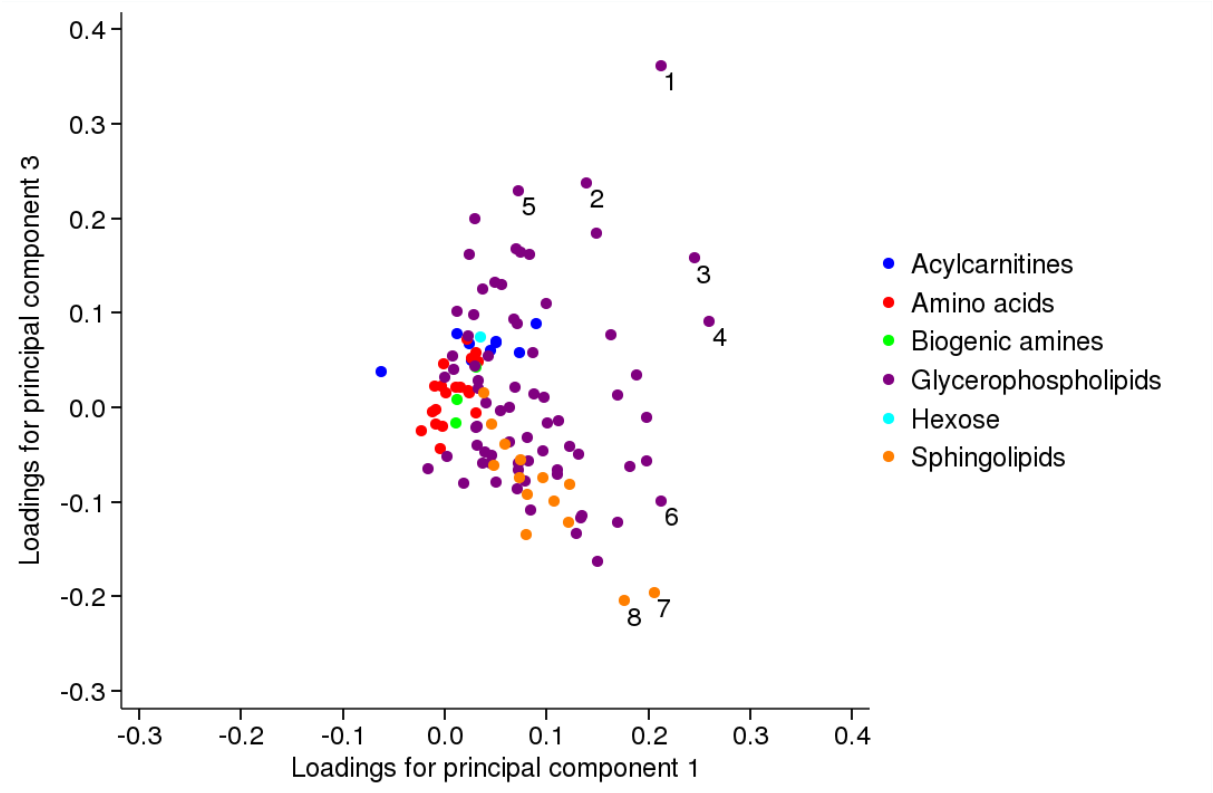


Figure 5.8: Loading plot for principal component 1 and 3 – main model

Each point represents a metabolite and metabolites with a loading more extreme than ±0.2 are marked: 1 PC aa C32:1; 2 PC aa C34:4; 3 PC aa C36:5; 4 PC aa C36:6; 5 PC aa C40:5; 6 PC ae C34:0; 7 SM (OH) C14:1; and 8 SM (OH) C16:1. Loadings were derived using principal component analysis based on the covariance matrix of log-transformed metabolite concentrations.

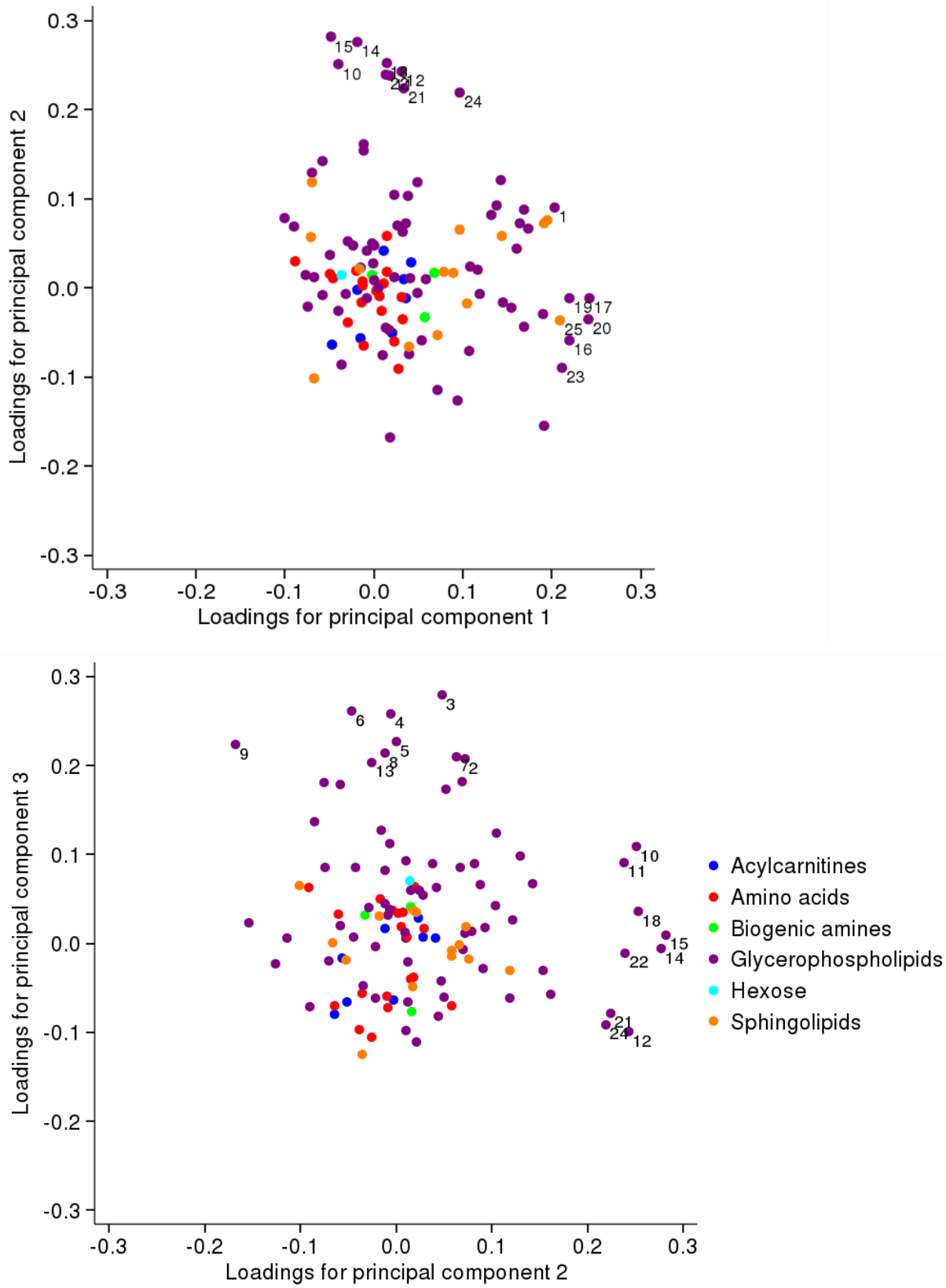


Figure 5.9: Loading plot for principal component 1 and 2, and 2 and 3 – alternative model
Each point represents a metabolite and metabolites with a loading more extreme than ±0.2 are marked: 1 lysoPC a C17:0; 2 PC aa C30:0; 3 PC aa C32:1; 4 PC aa C32:2; 5 PC aa C34:1; 6 PC aa C34:3; 7 PC aa C34:4; 8 PC aa C36:1; 9 PC aa C36:3; 10 PC aa C36:5; 11 PC aa C36:6; 12 PC aa C38:0; 13 PC aa C38:3; 14 PC aa C38:6; 15 PC aa C40:6; 16 PC ae C34:2; 17 PC ae C36:2; 18 PC ae C38:0; 19 PC ae C38:3; 20 PC ae C38:4; 21 PC ae C38:6; 22 PC ae C40:1; 23 PC ae C40:4; 24 PC ae C40:6; and 25 SM C20:2. Loadings were derived using principal component analysis based on the correlation matrix of log-transformed and z-score standardised metabolite concentrations.

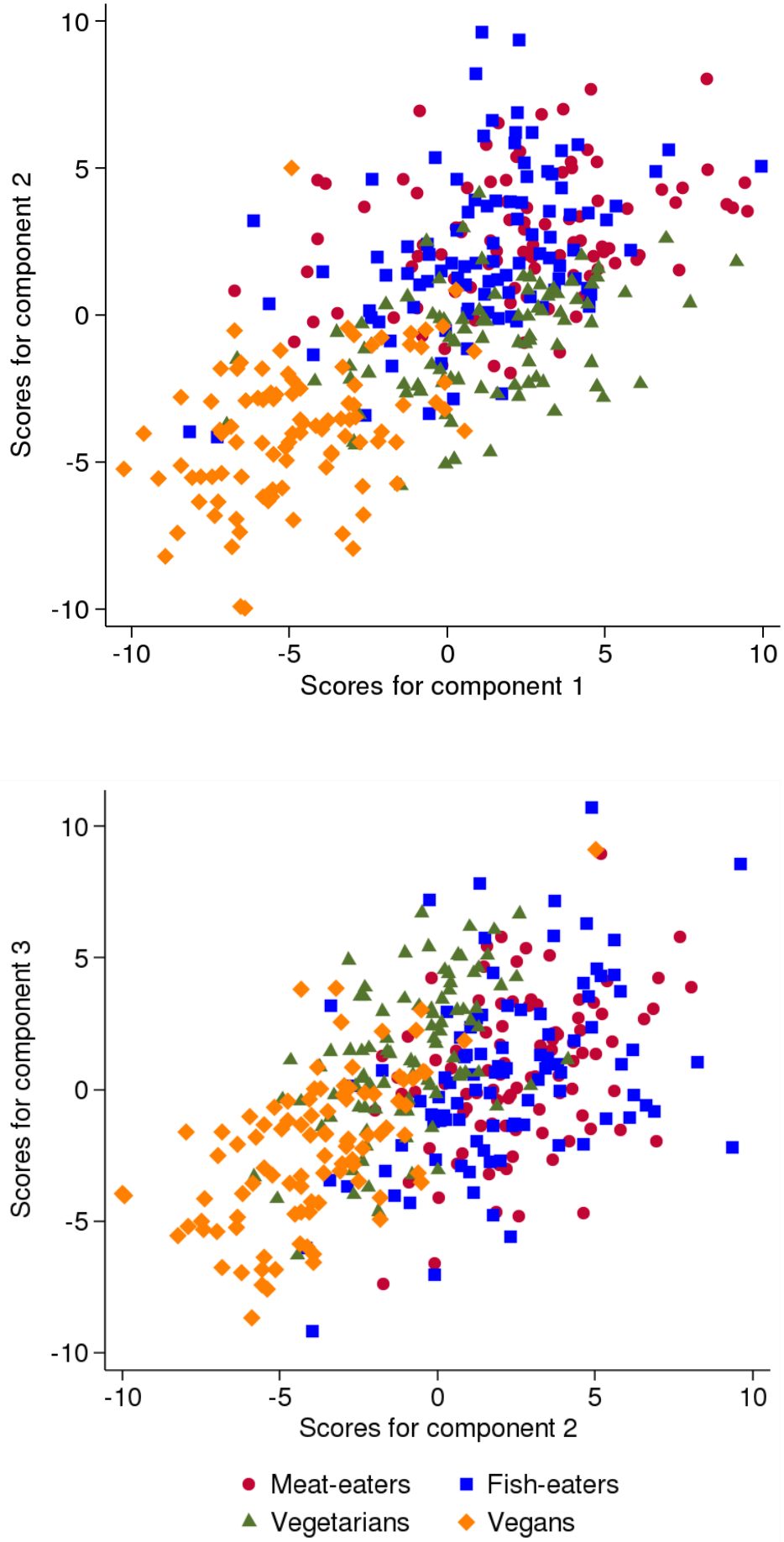


Figure 5.10: Score plots for principal component 1 and 2, and 2 and 3 by diet group – alternative model
Scores were derived using principal component analysis based on the correlation matrix of log-transformed and z-score standardised metabolite concentrations. Each point represents a participant.

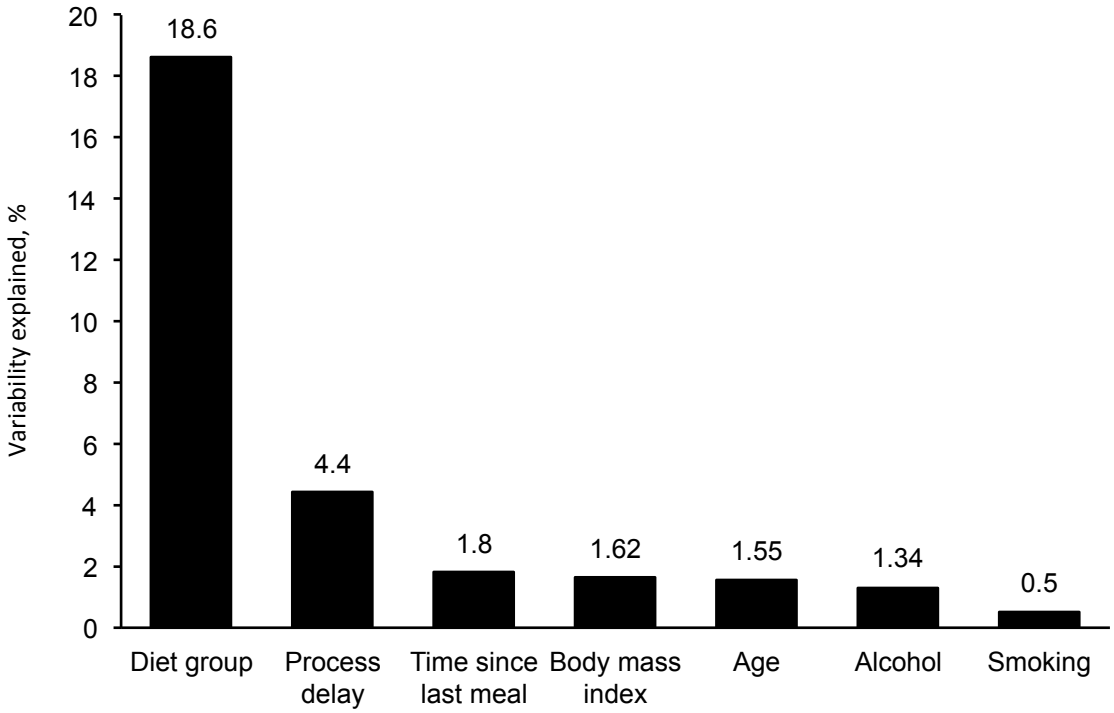


Figure 5.11: Variability in metabolite concentrations explained by diet group and other characteristics
The percentage of variability in metabolomics data explained by diet group and characteristics of the participants and blood samples were calculated using the PC-PR2 method [161].

6. Individual metabolite concentrations and prostate cancer risk

6.1 Introduction

6.1.1 Prostate cancer

Prostate cancer is the second most commonly diagnosed cancer in men worldwide [1], with the highest incidence rates in high-income regions, such as Western and Northern Europe, North America and Australia/New Zealand (85 to 112 per 100,000), but also in the Caribbean and Southern Africa (80 and 62 per 100,000, respectively) [1]. The only established risk factors for prostate cancer are older age, African ancestry [22], family history of prostate cancer [34], genetic factors [37] and circulating concentrations of insulin-like growth factor-I (IGF-I), which is the only potentially modifiable risk factor [2].

6.1.2 Metabolomics

Metabolomics is the identification and quantification of metabolites (i.e. low-molecular weight (up to 1000 to 1500 Dalton [13, 29]) reactants, intermediates or products of biochemical reactions) in a biological system and it is estimated that the human metabolome comprises many thousand metabolites [13, 93]. Since metabolite concentrations are affected by dietary, lifestyle, environmental and genetic factors, the measurements provide a snapshot of the biological state of a system, physiological as well as pathological [11, 14].

In recent years, metabolomics has been applied in epidemiological studies, with the aim of identifying novel risk factors for disease and biomarkers of exposure and disease, and to advance our understanding of disease mechanisms [10]. Examination of the metabolome may thus help us to understand the mechanisms underlying the association between dietary protein, IGF-I and prostate cancer risk (See Section 2.2.6 on page 15 for details) and to identify novel risk factors for prostate cancer.

6.1.3 Studies of metabolomics and prostate cancer

Most studies of prostate cancer using metabolomics have focused on identifying biomarkers for screening, for diagnostic and prognostic purposes. Reviews conclude that metabolite profiles, in different mediums including blood, urine, prostate tissue and bone, might be related to prostate cancer prognosis, recurrence and treatment effect [11, 85, 96]. Although results differ between studies, amino acids, citrate, choline containing compounds, lactate, phospholipids, polyamines and sarcosine have been associated with prostate cancer in several of these studies [11, 85, 96].

In contrast, little is known about how metabolic profiles prior to diagnosis might be related to risk of developing prostate cancer [28, 29, 82, 83, 84] and thus about the potential for preventive strategies. A lower risk of overall and aggressive prostate cancer in men with higher circulating concentrations of metabolites related to energy and lipid metabolism (including α-ketoglutarate, 1-stearoylglycerol and some glycerophospholipids) has been reported [28, 82, 84], with specific lipid species potentially being associated with different tumour stages [83]. Furthermore, a positive association with overall prostate cancer risk has been indicated for the glycerophospholipid phosphatidylcholine (PC) acyl-alkyl (ae) C30:0, two amino acids and a biogenic amine [84]. Contrasting results have been found in a population screened for prostate cancer, including positive associations with lipids and inverse associations with amino acids and peptides [29]. The four previously published prospective studies (five publications [28, 29, 82, 83, 84]), however, have a relatively limited number of cases (each ≤ 380). Only four of these papers have reported on prostate cancer risk by tumour characteristics [28, 29, 82, 83] and in all, except Huang et al. (2017) [83], advanced stage and high grade tumours were considered together as one category of aggressive disease rather than separately. Besides the paper published on the work described in this chapter [19], there are no published prospective data on metabolite concentrations and subsequent risk of death from prostate cancer, of which I am aware.

6.1.4 How metabolites might be involved in prostate cancer

Due to the novelty of metabolomics and the large number of metabolites measured, this is an exploratory analysis with few prior hypotheses about the associations between metabolites and risk of prostate cancer. However, evidence from basic and clinical research suggests that some of the measured metabolites might potentially be linked to prostate cancer.

Prostate cancer can form and develop under hyper-activation of the PI3K-Akt-mTOR pathway, a key regulator of cellular processes, and deregulation of this pathway has been observed in up to 50% of prostate cancers [89, 4]. The activity of this pathway might be influenced by dietary protein composition because branched-chain amino acids (i.e. leucine, isoleucine and valine) can upregulate mTOR activity and thereby increase proliferation and decrease apoptosis [4, 5, 89]. Furthermore, intake of essential amino acids may increase circulating IGF-I concentration [86, 87, 88] which in turn stimulates proliferation and cell survival through the PI3K-Akt-mTOR pathway [3].

The membrane lipids phosphatidylcholines and sphingomyelins play a role in several cellular regulatory processes, including proliferation, apoptosis and migration, and in cell signalling [177, 178].

Differences in concentrations of both amino acids and glycerophospholipids have been observed in cases of other cancers than prostate cancer, compared to controls; other cancers were

kidney, colorectal, lung, breast and gastric cancers for amino acids and brain, breast, lung and liver cancers for glycerophospholipids [11].

6.2 Aim

The aim of this study was to prospectively investigate the association between pre-diagnostic metabolite concentrations and risk of prostate cancer overall, as well as by time to diagnosis and tumour characteristics, and risk of death from prostate cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). The studied metabolites were acylcarnitines, amino acids, biogenic amines, glycerophospholipids, hexose and sphingolipids.

6.3 Methods

6.3.1 Study population

The EPIC cohort has been described in detail in Chapter 3 on page 21. In brief, the cohort includes 153,400 men recruited from 19 centres in eight European countries (Denmark, Germany, Greece, Italy, Netherlands, Spain, Sweden and United Kingdom) between 1992 and 2000, when they were aged mostly 35 to 70 years old. At recruitment, detailed information was collected on dietary intake, lifestyle, anthropometry and previous disease, and 139,600 men also gave a blood sample [99].

All participants gave written informed consent to participate in the EPIC cohort. Approval for the current study was obtained from the Internal Review Board of the International Agency for Research on Cancer (IARC), Lyon, France (Project No. 14-09) and participating centres as required.

For the current analysis, men were eligible if they had blood stored at the central biobank at IARC (samples coming from 15 centres in Germany, Greece, Italy, Netherlands, Spain and United Kingdom; Figure 3.1 on page 22), the date of blood collection was known and if no cancer (except non-melanoma skin cancer) had been diagnosed at the time of blood collection.

6.3.1.1 Follow-up and case and control selection

Information on cancer incidence, tumour sub-types and vital status was obtained via record linkage to population-based cancer and mortality registries in most countries. However in Germany and Greece, a combination of methods was used, including health insurance records, local (municipality-, hospital- or physician-based) cancer and pathology registries, and active follow-up via participants or their next-of-kin; self-reported cancer incidence was verified via medical records.

Cases: Prostate cancer was defined as code C61 in the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) and cases were men diagnosed with prostate cancer after blood collection and prior to the end of follow-up. For these analyses samples were available for cases from both the first and third phase of centralisation of follow-up data (phase I and III); with end of follow-up ranging between centres from 2001 to 2002 and from 2007 to 2008, respectively.

Histological grade was known for 83.8% of cases; low-intermediate grade (Gleason score < 8 or coded as well, moderately or poorly differentiated tumours) was diagnosed in 778 men and 124 men were diagnosed with high grade disease (Gleason score ≥ 8 or coded as undifferentiated tumours). Information on tumour stage was available for 61.7% of cases. 456 men had localised tumours (tumour-node-metastasis (TNM) system score of $\leq$ T2 and N0/X and M0, or stage coded as localised), 208 men had advanced stage tumours (TNM score of $\mathrm{T3-4}$ and/or $\mathrm{N1-3}$ and/or M1, or coded as advanced), 115 men had aggressive disease (a subset of advanced stage disease defined as TNM score of T4 and/or $\mathrm{N1-3}$ and/or M1), while 549 men had non-aggressive disease.

Matched controls: Each case was matched to one control participant, who was selected randomly among male cohort participants who were alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the case. Matching criteria were study centre, length of follow-up, age at blood collection ($\pm$6 months), and time of day ($\pm$1 hour) and fasting status (i.e. time since last meal; < 3, $3-6$, > 6 hours) at blood collection. An incidence density sampling procedure was used such that a control could become a case at a later date or be a control for more than one case.

6.3.2 Blood collection and laboratory analysis

6.3.2.1 Metabolomic assay

A standardised protocol for blood collection and processing was followed (see Section 3.2.1 on page 23 for details). Prior to this assay, samples had been thawed for another assay and refrozen. All plasma samples (with citrate as an anticoagulant) were assayed at IARC, using the Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbrcuk, Austria) and following the procedure recommended by the vendor. A triple quadropole mass spectrometer (Triple Quad 4500; AB Sciex, Framingham, Massachusetts, USA) was used to quantify a total of 142 metabolites; amino acids ($n = 22$) and biogenic amines ($n = 11$) were separated by ultra-high performance liquid chromatography before injection into the mass spectrometer, whereas flow injection analysis was used for acylcarnitines ($n = 18$), glycerophospholipids ($n = 78$), hexose

($n = 1$) and sphingolipids ($n = 12$). Please refer to Section 3.5.2 on page 28 for metabolite nomenclature.

Samples from matched case-control sets were assayed in the same analytical batch, each of which included quality control samples of pooled plasma (six to 11 per batch and a total of 255 from one plasma pool). Laboratory personnel were blinded to sample category, i.e. case, study control or quality control.

6.3.2.2 Total prostate-specific antigen

Measurements for total prostate-specific antigen (PSA) were conducted using the dual-label DELFIA Prostatus® total PSA-Assay (Perkin-Elmer, Turku, Finland) [179] calibrated against the World Health Organization (WHO) 96/670 (PSA-WHO) standard in Dr. Lilja's laboratory at the Wallenberg Research Laboratories, Department of Translational Medicine, Lund University, Skåne University Hospital, Malmö, Sweden, for a previous study [31]. Intra- and inter-assay coefficients of variation (CV) ranged from 7 to 14% (personal communication Karl Smith Byrne, Cancer Epidemiology Unit, University of Oxford). PSA concentration was available for 71.1% of men in the current study.

6.3.3 Exclusions of metabolites and participants

Concentrations of 142 metabolites were available for 2,169 men (Figure 6.1 on page 194). Table 6.1 on page 135 displays the completeness and CVs for all measured metabolites.

Metabolites ($n = 18$) were excluded if more than 15% of participants had measurements outside the measurable range (Figure 6.1 on page 194), and the remaining measurements outside the measurable range were imputed. Measurements below the limit of detection (applicable to 12 metabolites for one to 307 men) and below the limit of quantification (applicable to seven metabolites for one to 296 men) were set to half the lowest measured concentration and to half the limit of quantification, respectively. Measurements above the highest concentration calibration standards were set to the highest standards (applicable to one metabolite for one man).

Metabolites ($n = 2$) for which the overall CV was higher than 20% (Figure 6.1 on page 194 and Table 6.1 on page 135) were further excluded. This left 122 metabolites for the analysis (seven acylcarnitines, 21 amino acids, six biogenic amines, 75 glycerophospholipids, hexose and 12 sphingolipids. The median (range) of overall coefficients of variation was 12.3% ($7.0 - 17.2\%$) for acylcarnitines, 8.8% ($6.0 - 12.6\%$) for amino acids, 10.6% ($4.3 - 17.2\%$) for biogenic amines, 11.2% ($7.3 - 17.7\%$) for glycerophospholipids and 10.4% ($8.0 - 19.9\%$) for sphingolipids, and the coefficient of variation for hexose was 6.5%.

Men with missing information on any of the 122 metabolites ($n = 1$) and men in incomplete case-control sets ($n = 14$) were excluded, leaving 1,077 matched case-control sets for the statistical analysis.

6.3.4 Statistical analysis

Logarithmically-transformed concentrations of metabolites were used for all analyses.

6.3.4.1 Correlations between analytes

Partial correlations were calculated between log-transformed concentrations of total PSA and metabolites separately in all controls, controls with total PSA below 1 ng/mL, and cases, adjusting for age at blood collection ($< 55; 55-59; 60-64; 65-69, \geq 70$ years), body mass index (BMI; fourths of the distribution, corresponding to: $< 24.5; 24.5 - 26.5; 26.6 - 28.8; \geq 28.9$; unknown) and study centre. The analyses of metabolite-metabolite correlations included 1,077 controls, while correlations between metabolites and total PSA included 764 controls, 489 controls with total PSA concentration below 1 ng/mL and 768 cases, respectively.

6.3.4.2 Metabolite concentrations by covariates

In controls, we investigated associations between metabolite concentrations and potential covariates, including participant characteristics and factors related to blood sampling and handling.

Firstly, differences in geometric mean metabolite concentrations across categories of covariates were tested using analysis of variance, after adjustment for age at blood collection, BMI (both categorised as above), study centre and assay batch, where these were not the variable of interest. If each subsequent mean was lower or higher than the preceding mean, the direction of the association was defined as inverse or positive, respectively.

Secondly, the variability in metabolite concentrations attributed to each covariate, after controlling for all other variables in the model was calculated using the principal component partial R-square (PC-PR2) method [161] as described in Section 5.3.5.6 on page 62.

6.3.4.3 Risk of overall prostate cancer

Risk per one standard deviation – the continuous model: We used conditional logistic regression to estimate risk of prostate cancer for a one standard deviation (SD) increase in metabolite concentrations; using metabolite concentration divided by the standard deviation of metabolite concentration as a continuous variable.

Tests for linear trend were computed for metabolite concentrations as continuous variables. Departure from linearity was tested using the likelihood ratio χ^2 test comparing the linear model

with two non-linear models, one including the quadratic polynomial of metabolite concentration and one including the cubic polynomial. The cubic polynomial is more flexible than the quadratic and might thus better capture any non-linearity. Thus, the models with the quadratic and the cubic polynomial were compared using the likelihood ratio χ^2 test.

The analysis was conditioned on the matching variables and further adjusted for exact age (continuously) in one model, and additionally for BMI (fourths of the distribution; unknown), smoking (never; past; current; unknown), alcohol intake ($< 10; 10-19; 20-39; \geq 40$ g/d; unknown), education (primary; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown) in a second model. Results from the minimally- and multivariable-adjusted models did not materially differ and only results from the latter model are presented.

Risk by fifths of metabolite concentrations: Risk by fifths of metabolite concentrations (based on the distribution among controls) was additionally computed for overall risk of prostate cancer using conditional logistic regression.

Tests for linear trend were computed across the median concentrations in the fifths. Departure from linearity was tested using the likelihood ratio χ^2 test comparing models with the fifths scored with their medians and as indicator variables, respectively.

As for the continuous model, results were not materially changed when only adjusting for age as compared to adjusting for the full set of covariates described above; results from the multivariable-adjusted model are presented.

6.3.4.4 Sub-group analyses

Conditional logistic regression models similar to the continuous model for overall prostate cancer risk were fitted for sub-groups by time to diagnosis (≤ 5 and > 5 years) and by tumour characteristics (low-intermediate and high grade, localised and advanced stage, and non-aggressive and aggressive prostate cancer), after assigning the controls the value of their matched case for the outcome variables, e.g. grade. Tests for heterogeneity in the association between continuous metabolite concentrations and prostate cancer risk by sub-groups were done using the likelihood ratio χ^2 test, which compared models with and without an interaction term between the linear trend variables and the outcome variable of interest, e.g. grade.

We also assessed the risk of death from prostate cancer using the subset of cases who were known to have subsequently died of prostate cancer during follow-up (prostate cancer listed as the underlying cause of death on the death certificate). The original matching was preserved and matched sets in which the control had died before the case ($n = 13$) or vital status was unknown for the control ($n = 4$) were excluded.

6.3.4.5 Multiple testing

All tests of statistical significance were two-sided, and to account for multiple testing a false discovery rate (FDR)-controlling procedure (Benjamini–Hochberg [154]) was applied. First, the p-values were sorted and ranked from the lowest $p(1)$ to the highest $p(m)$. Let k be the largest rank $p(i)$ for which $p(i) < (i/m) \times \alpha$ is true; α is the significance level and set to 0.05 in all analyses. Then all the null hypotheses for p-values from $p(1)$ to $p(k)$ were rejected [154]. This method allows 5% of positive findings to be false, on average [140].

Additionally, an adjusted p-value (p_{adj}; sometimes referred to as a q-value by other authors) was computed. $p_{adj}(i)$ was defined as the minimum of $p(n) \times (m/n)$ for n being $i, i+1, \ldots, m$. Tests with a $p_{adj} < 0.05$ were declared significant after controlling the FDR at 5% [173].

Tests with p-values < 0.05 for which the null hypothesis was not rejected after controlling the FDR are denoted as conventionally statistically significant.

6.3.4.6 Sensitivity analyses

Overlap with analysis in EPIC-Heidelberg: In a sensitivity analysis, we excluded men who were also in a previous analysis conducted in EPIC-Heidelberg [84] (91 cases and 11 controls; personal communication Tilman Kühn, Division of Cancer Epidemiology, German Cancer Research Center (DKFZ)) and resulting incomplete matched sets. This left 985 matched sets for this analysis.

6.3.4.7 Statistical software

All analyses were done in Stata Statistical Software Package, version 14 (Stata Corporation, College Station, Texas, USA), except the PC-PR2 which was done using R: A language and environment for statistical computing, version 3.2.3 (Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/). The power calculation presented in the discussion (see Section 6.5.4.2 on page 134) was done in Quanto, version 1.2.4.

6.4 Results

6.4.1 Participant characteristics

At blood collection, participants were on average 60 years of age (range 40 to 77 years) and, at diagnosis, cases were on average 67 years old (range 47 to 88 years). No clear differences were seen in baseline characteristics between cases and controls (Table 6.2 on page 138).

6.4.2 Distribution of metabolite concentrations

The distribution of metabolite concentrations by case-control status is shown in Table 6.3 on page 139. The concentrations varied from hundredths of a μmol/L for some acylcarnitines to thousands of μmol/L for hexose. After logarithmical transformation, most metabolites were close to normally distributed indicated by a skewness and kurtosis of zero and three, respectively. However, some metabolites were slightly left-skewed and more peaked than the normal-distribution, e.g. PC ae C44:4.

6.4.3 Correlations

Strong positive correlations were observed within metabolite classes with correlation coefficients (r) up to 0.9 within amino acids, glycerophospholipids and sphingolipids but also between metabolites in the latter two classes ($r \leq 0.8$; Figure 6.2 on page 195). Total PSA and metabolite concentrations were not strongly correlated in controls, controls with low total PSA concentration or cases ($-0.16 \leq r \leq 0.13$; Figure 6.2 on page 195 and Table 6.4 on page 144).

6.4.4 Metabolite concentrations by covariates

Associations between metabolite concentrations and covariates, remaining after controlling the FDR, are shown in Table 6.5 on page 149. Study centre and assay batch were associated with concentrations of the vast majority of metabolites (81.1% and 80.3%, respectively). Furthermore, alcohol intake and BMI were related, in various directions, to concentrations of approximately half the metabolites overall, including metabolites from all the classes. Time of day and fasting status at blood collection were associated with 21.3% and 12.3% of metabolites, respectively. Fasting men had lower concentrations of several amino acids and a biogenic amine, and higher concentrations of two acylcarnitines, than non-fasting men. Similarly, time of day at blood collection was associated with several amino acids, two acylcarnitines and three biogenic amines, and also some glycerophospholipids. Age, smoking status and physical activity were associated with fewer metabolite concentrations (seven to 12 each); notably there were lower concentrations of several glycerophospholipids in older men and in men who were current smokers. There was little evidence of associations of metabolite concentrations with marital status or educational attainment.

Together the covariates explained 29.6% of the total variability in the metabolite data, with assay batch and centre explaining the most variability (7.4 and 5.4%, respectively), followed by alcohol intake and BMI, while other covariates all explained less than 1% each (Figure 6.3 on page 196).

6.4.5 Risk of overall prostate cancer

6.4.5.1 The continuous model

The statistical significance of associations between metabolite concentrations and overall prostate cancer risk are shown in Figure 6.4 on page 197, and in Figure 6.5 on page 198 odds ratios (OR) and 95% confidence intervals (CI) for a one SD increase in metabolite concentration are shown for metabolites with $p < 0.1$ (additional results in Table 6.6 on page 155). Conventionally statistically significant ($p < 0.05$) lower risks were seen in men with higher concentrations of the acylcarnitine C18:1 and the amino acid citrulline, while higher risks were seen for the amino acid trans-4-hydroxyproline (t4-OH-Pro), glycerophospholipids PC diacyl (aa) C28:1, PC ae C30:0 and PC ae C30:2, and sphingolipid SM (OH) C14:1, but none of the associations were statistically significant after controlling for multiple testing. The strongest association was with PC ae C30:0, for which the risk was 16% higher per SD increase in concentration ($OR_{1SD} = 1.16$; 95%CI: $1.04 - 1.30$).

There was no evidence of departure from linearity in the association with prostate cancer risk for the seven metabolites that were associated with overall prostate cancer risk, except for C18:1 for which both the quadratic and cubic polynomial fitted the data conventionally significantly better than a straight line, but not after controlling for multiple testing (Table 6.7 on page 158). Neither was there much evidence of departure from linearity for other metabolites.

6.4.5.2 The categorical model

The associations observed in the continuous model for t4-OH-Pro, PC ae C30:0 (conventionally significant in the top fifth) and SM (OH) C14:1 (conventionally significant from the 2nd to the 5th fifths) persisted in the categorical model (Table 6.8 on page 163). In addition, conventionally significant trends of lower risk were seen for acylcarnitine C14:1 and glycerophospholipids lysoPC a C16:1 and PC aa C36:4, while a higher risk was seen for amino acid isoleucine. For all four, the risk in the top fifth was conventionally significantly different from the bottom fifth.

There was no evidence of non-linearity for any of the metabolites showing associations with prostate cancer risk in the categorical model (Table 6.7 on page 158) and little evidence of departure from linearity for other metabolites.

6.4.5.3 Time to diagnosis

Of the cases, 428 (39.7%) were diagnosed within five years of blood collection. A one SD increase in citrulline concentration was associated with a 27% lower risk of prostate cancer diagnosed within five years of blood collection but not with prostate cancer diagnosed later ($OR_{1SD} = 0.73$; 95%CI: $0.62 - 0.86$; $p_{trend} = 0.0002$, which was significant after controlling for multiple testing,

and $OR_{1SD} = 1.02$; 95%CI: $0.90 - 1.16$, respectively; $p_{het} = 0.0009$; Figure 6.6 on page 199 and Table 6.9 on page 168).

Conventionally significant associations were seen in the first five years of follow-up, inversely for three acylcarnitines (including C18:1) and positively for two amino acids and a glycerophospholipid, and for follow-up beyond five years, for t4-OH-Pro, PC aa C28:1 and PC ae C30:0, although there was no evidence of heterogeneity by time to diagnosis.

6.4.6 Risk of prostate cancer sub-types

6.4.6.1 High grade

Conventionally significant inverse associations with high grade prostate cancer were seen for an acylcarnitine, 24 glycerophospholipids and three sphingolipids, with risk reductions up to 48% ($OR_{1SD} = 0.52$, 95%CI: $0.35 - 0.79$ for PC aa C32:3; Figures 6.7 on page 200 and 6.8 on page 201, and Table 6.10 on page 173). However the associations did not remain statistically significant after correcting for multiple testing.

6.4.6.2 Advanced stage

After controlling for multiple testing, 12 glycerophospholipids, i.e. lysoPC a C18:0, PC aa C36:2, PC aa C36:3, PC aa C38:3, PC aa C38:5, PC aa C40:2, PC aa C40:3, PC aa C40:4, PC aa C40:5, PC aa C42:4, PC aa C42:5 and PC ae C40:1, were inversely associated with risk of advanced prostate cancer, with risk reductions up to 46% ($OR_{1SD} = 0.54$; 95%CI: $0.40 - 0.72$ for PC aa C40:3; Figures 6.9 on page 202 and 6.10 on page 203, and Table 6.11 on page 178). For six of these and three additional glycerophospholipids, the test for heterogeneity by stage was significant after taking multiple testing into account, with inverse associations for advanced disease and no associations for localised disease.

Additionally, conventionally significant associations with advanced disease were seen for four acylcarnitines, two amino acids, a biogenic amine, 20 glycerophospholipids and two sphingolipids; all but the amino acid proline were inversely associated advanced disease.

6.4.6.3 Aggressive cancer

Conventionally significant associations were seen for 13 metabolites with aggressive prostate cancer. Positive association were observed with an acylcarnitine, an amino acid, a glycerophospholipid, hexose and a sphingomyelin, while inverse associations were seen for seven glycerophospholipids and a sphingolipid, but none remained after controlling the FDR (Figures 6.11 on page 204 and 6.12 on page 205, and Table 6.12 on page 183). The strongest association

was with glycerophospholipid lysoPC a C16:0 for which the estimated risk reduction was 48% ($OR_{1SD} = 0.52$, 95%CI: $0.31 - 0.86$).

6.4.7 Risk of death from prostate cancer

During follow-up, 144 men died of prostate cancer and after excluding matched sets in which the control had died before the case ($n = 13$) or vital status was unknown for the control ($n = 4$), 127 matched sets were available for analysis. Seven metabolites were conventionally significantly associated with death from prostate cancer. Men with higher concentrations of an acylcarnitine (C3), two amino acids (methionine and t4-OH-Pro), a biogenic amine (ADMA), hexose and a sphingolipid (SM (OH) C14:1) were at higher risk, while an inverse association was found for a glycerophospholipid (PC aa C42:4; Figures 6.13 on page 206 and 6.14 on page 207, and Table 6.13 on page 188). However, the associations did not remain statistically significant after controlling for multiple testing.

6.4.8 Sensitivity analyses

6.4.8.1 EPIC-Heidelberg analysis

Results were not materially changed after excluding the overlap in participants between the EPIC-Heidelberg [84] and our analysis (Table 6.14 on page 191). For the seven metabolites, which were associated with risk in our main analysis of overall prostate cancer risk, risk estimates were similar after exclusion of the overlap in study participants, and all but C18:1 and citrulline were still conventionally significant. In addition, a positive association with PC ae C34:0 was conventionally statistically significant.

6.5 Discussion

6.5.1 Main findings

In this prospective study of 122 plasma metabolite concentrations and prostate cancer risk, seven metabolites were associated with risk of overall prostate cancer at a conventional level of statistical significance but not after allowance for multiple testing. These metabolites were from several metabolite classes, suggesting that dysregulation of many metabolic pathways may be related to prostate cancer. The results stratified by time to diagnosis suggested that lower concentrations of citrulline might be a marker of sub-clinical prostate cancer as the association was observed for disease diagnosed in the first few years after blood draw only. In contrast, the association with other metabolites, including t4-OH-Pro, SM (OH) C14:1, PC ae C30:0 and other glycerophospholipids, did not vary by time to diagnosis and these might thus provide insights into

aetiology. Sub-group analyses indicated a possible link between higher concentrations of several glycerophospholipids and decreased risk of high-risk tumour sub-types, especially advanced stage prostate cancer. Similar to the results for overall prostate cancer, suggestive associations ($p < 0.05$) were observed between death from prostate cancer and seven metabolites from several metabolite classes.

6.5.2 Other studies

Besides the current study, four smaller prospective studies (five publications) of metabolomics and prostate cancer risk have been published; two were nested within the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) [28, 82, 83], one in EPIC-Heidelberg [84] and one in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) [29]. The publications included 200, 74, 137, 310 and 380 cases, respectively. Comparison of results from metabolomics studies is not straightforward owing to differences in population characteristics, biological medium [166], the technological assay and statistical tools, including procedures for dealing with the large number of metabolites. Nonetheless, replication of possible associations both with specific metabolites and also more globally with metabolite classes is essential for interpretation of results.

In the two ATBC studies, mass spectrometry was used to measure 420 and 626 metabolites, respectively, samples were fasting serum samples from smokers, and stage and grade were combined to define aggressive tumours (as opposed to our more restricted definition of aggressive tumours based on TNM-stage and not grade) [28, 82]. Amino acids, glycerophospholipids and sphingolipids were measured and with some overlap of specific metabolites with the current study (19 amino acids and two biogenic amines).

The first ATBC analysis [82] did not find a significant association between citrulline and prostate cancer, but for overall prostate cancer the OR per SD increase in citrulline (0.87, 95%CI: $0.62 - 1.23$) was similar to the observation in our study (0.90, 95%CI: $0.82 - 0.99$). We did not replicate the ATBC results suggesting a lower risk of prostate cancer ($p < 0.05$) in men with higher concentrations of some amino acids, i.e. alanine (for overall and aggressive prostate cancer), lysine (for overall and non-aggressive prostate cancer), methionine (for overall and aggressive prostate cancer) or phenylalanine (for overall and non-aggressive prostate cancer). In contrast, we observed a conventionally significant higher risk of death from prostate cancer with higher methionine concentration. The main finding of the first ATBC study was a strong inverse association between 1-stearoylglycerol, a product of lipid breakdown, and risk of overall and aggressive prostate cancer, but we did not have data on this metabolite.

In line with the first ATBC study, the second [83, 28], with no overlap in participants from the first [82], showed inverse associations of lipids (including glycerophospholipids) and energy

metabolites (involved in the Krebs cycle) with risk of aggressive prostate cancer. Similarly, we found inverse associations for several glycerophospholipids with risk for high grade, aggressive (defined based on stage) and especially advanced prostate cancer. In a separate publication, results by tumour stage [83] were reported to differ between men with T2 (tumour within the prostate), T3 (tumour broken through the capsule) and T4 (invasion of nearby structures) tumours compared to controls; alterations in histidine metabolism were observed in men diagnosed with T2 tumours, while different lipids were reported for T3 and T4 tumours. However, further data are needed to support these findings as the number of cases in each sub-type was very limited (71, 51 and 15 cases for T2, T3 and T4, respectively).

The PLCO analysis used the same metabolomic assay as the ATBC analyses, although on non-fasting rather than fasting serum samples, but, in contrast to any of the other studies, participants were all screened for prostate cancer using annual PSA test and digital rectal examination [29]. Lower risk of prostate cancer was suggested ($p < 0.05$) in relation to higher concentrations of some amino acids and their derivatives, including arginine (with aggressive prostate cancer; defined by combining stage and grade information) and tryptophan (with overall and aggressive prostate cancer). We did not replicate these findings. The reported positive associations with lipids for overall and aggressive prostate cancer were inconsistent with the results from ATBC and our findings. These differences may be due to the screening of the PLCO population and the resulting differences in case-mix [29] as screening was not common in ATBC [28] or EPIC.

In the EPIC-Heidelberg analysis [84], mostly the same metabolites were investigated as in the current analysis and a similar assay was used. The strongest finding for overall prostate cancer in both studies was a suggested ($p < 0.05$) positive association with the glycerophospholipid PC ae C30:0 (EPIC-Heidelberg: $OR_{\text{top vs bottom fourth}} = 1.89$, 95%CI: $1.06 - 3.36$). Our results were not materially changed after excluding the small overlap in participants between the two analyses. In both EPIC-Heidelberg and our analyses, the results for PC ae C30:0 did not differ by time to diagnosis. Additionally, the EPIC-Heidelberg analysis suggested inverse associations with three glycerophospholipids (all lysophosphatidylcholines) and positive associations with alanine, proline and methioninesulfoxide, the latter of which was not available for our analysis. Our results for these metabolites were mainly in the same direction but less strong for overall prostate cancer, while we observed stronger inverse associations for two of the lysophosphatidylcholines with aggressive disease and positive associations of proline with advanced and aggressive prostate cancer.

6.5.3 Possible mechanisms

The strongest findings in our study were an inverse association of citrulline with risk of prostate cancer diagnosis within the first five years of follow-up, positive associations of prostate cancer risk with t4-OH-Pro, SM (OH) C14:1 (for overall and death from prostate cancer for both metabolites) and PC ae C30:0 (for overall and aggressive prostate cancer), and a lower risk of advanced prostate cancer in relation to several glycerophospholipid concentrations. However, relatively little is known about the potential biological role of these specific metabolites in carcinogenesis of the prostate.

6.5.3.1 Citrulline

Citrulline has antioxidant functions and has been shown to protect DNA and enzymes from reactive oxygen species [180], which might otherwise promote progression of prostate cancer via continuous proliferation and impaired apoptosis [181].

In line with our results, lower urinary concentrations of citrulline have been observed in prostate cancer patients than in healthy controls [182]. This may potentially be due to altered citrulline metabolism in some tumour cells; the activity of an enzyme involved in the formation of citrulline (inducible nitric oxide synthase, which converts arginine to citrulline) might be lower in some prostate cancer cell lines than others. However, whether citrulline might be useful as a marker of sub-clinical prostate cancer would need to be further investigated.

6.5.3.2 Hydroxyproline

It is not clear why t4-OH-Pro (one variant of hydroxyproline), might be related to risk of prostate cancer. However, urinary excretion of hydroxyproline has been previously recognised as an early marker of bone metastases in patients with prostate cancer [183, 184], as hydroxyproline is released from collagen in tumour invasion [185]. While 4-hydroxyproline has also been suggested as a marker of red and processed meat intakes [186, 187], the evidence does not suggest an association between intake of these foods and prostate cancer risk [188].

Standard amino acid have often been reported to differ between controls and patients with various cancer types [11, 94] but it is not clear if the associations with the non-standard amino acids citrulline and t4-OH-Pro are specific to prostate cancer or not.

6.5.3.3 Sphingomyelins

A role of sphingomyelins (the type of sphingolipid investigated here) in carcinogenesis might be explained by their involvement in cell proliferation, migration and autophagy [178]. SM (OH) C14:1 has also been suggested as a marker of cream intake [187], and intake of dairy

products might be related to prostate cancer risk [8]. However, sphingomyelins might also be markers of existing prostate cancer, as higher sphingomyelin concentrations have been observed in prostate tumour tissue and patients' plasma than in benign prostatic hyperplasia tissue and control participants, respectively [189, 190].

6.5.3.4 Phosphatidylcholines

Phosphatidylcholines (the type of glycerophospholipids investigated here) for cell membranes are synthesised via the CDP-choline pathway, which is a key determinant of cell cycle progression and programmed cell death, due to its role in provision of membrane components and lipid second messengers [177]. Homeostasis of phosphatidylcholines (including PC ae C30:0) plays a critical role in proliferation of cancer cells, and choline phosphotransferase catalysing their synthesis has been reported to facilitate survival of metastatic prostate cancer cells [177]. On the other hand, disruption of their synthesis, in e.g. DNA damage, leads to apoptosis. Higher plasma phosphatidylcholine concentrations in prostate cancer patients than controls have been reported [190], which is in line with the positive association of PC ae C30:0 with risk.

While possible mechanisms for the inverse associations between several phosphatidylcholines and advanced prostate cancer are not clear, positive associations have been reported between concentrations of diacyl-phosphatidylcholines (denoted 'PC aa Cx:y' in the current paper) and type 2 diabetes [147], which is linked to a lower prostate cancer risk [54, 55].

The association between phosphatidylcholines and risk might apply to malignancies in general rather than being prostate cancer specific. Changes in phosphatidylcholine concentrations have been reported in patients with cancers of the bladder, brain, breast, kidney, liver, lung and ovaries [11, 94, 191], and in the prospective analysis in the EPIC-Heidelberg sub-cohort associations with breast and colorectal cancer were also suggested [84]. Altered circulating phosphatidylcholine concentrations in men who are subsequently diagnosed with cancer may be due to altered lipid uptake and metabolism by rapidly proliferating cancer cells in sub-clinical tumours; these metabolites are required for membrane synthesis and lipid-based cell signalling [192].

6.5.4 Strengths and limitations

6.5.4.1 Strengths

To date, this is the largest study of metabolite concentrations and risk of prostate cancer. The relatively large number of participants has enabled us to investigate not only total risk of prostate cancer but also to conduct exploratory analyses by time to diagnosis and by prostate tumour characteristics, and of death from prostate cancer. Furthermore, the detailed information on

covariates, several of which were associated with metabolite concentration in the current analysis, has reduced the risk of confounding biasing the results.

6.5.4.2 Limitations

The choice of metabolomics assay was determined by coverage of metabolites of *a priori* interest (amino acids) and cost, but use of a targeted assay limits our analysis to the metabolites covered by the assay. The study has a number of other limitations.

Reproducibility over time: Only one blood sample was available per participant, which will attenuate the risk estimate if a single measure does not represent long-term exposure. Reproducibility over time can be expressed by the intra-class correlation coefficient (ICC; calculated as the between-persons variance divided by the sum of between- and within-person variance in repeat measurements [139]), where an ICC of one represents perfect reproducibility. Reproducibility of metabolites over four months to two years and four months has been reported to be relatively high (median ICC were 0.54 to 0.70), suggesting that a single measurement may be adequate for most metabolites [126, 146, 193], although the reproducibility was lower for some acylcarnitines and a few amino acids [126, 193].

Pre-analytical conditions: Pre-analytical conditions, such as food consumption prior to blood collection, temperature of blood samples, freeze-thaw cycles and use of citrate as anticoagulant could potentially diminish our ability to detect associations [145, 171, 194].

Firstly, participants were not required to fast before blood collection. However, the current analysis and other studies have found that fasting status only explains a limited amount of the variability of metabolite concentrations [18, 147, 161, 171, 172]. Nevertheless, cases and controls were matched on time since last food or drink to minimise any risk of bias due to fasting status.

Secondly, it is not clear if the temperature at which blood samples are handled prior to long term storage has a very limited [144] or larger [145] effect on metabolite concentrations and thus potentially on the study results. In the current study, cases and controls were matched by centre as a proxy for sample handling to minimise bias.

Thirdly, that samples had been thawed and refrozen for a previous assay is unlikely to have had a notable effect on the results because metabolite concentrations have been shown to be stable after several freeze-thaw cycles [144, 145].

Finally, like other anticoagulants, use of citrate in plasma samples can affect measurements of metabolites, including amino acids, glycerophospholipid and sphingomyelins [194, 195]. Thus the metabolite concentrations reported here might not be directly comparable to concentrations

measured in serum, or in plasma treated with EDTA or heparin. However, relative risk estimates are unlikely to be affected as citrate was used in all samples.

Stage and grade classification: Variation in the histological grading and stage classification between pathologists and over time could result in some misclassification of tumours by stage and grade category [196, 197, 198, 199]. Furthermore, these tools are not strong predictors of clinical outcome, e.g. how aggressive the tumours are [200]. This might lead to attenuation of risk estimates by stage and grade.

Correction for multiple testing: To account for multiple testing, we used an FDR-controlling procedure, which is less over-conservative than the Bonferroni correction used in previous papers [28, 29, 82, 84] but might still be somewhat conservative as it does not take correlations between metabolites into account.

Statistical power: Although this is the largest study to date of metabolites and prostate cancer risk, the numbers and thus the statistical power are still limited, especially for analyses of tumour sub-types and death.

As explained in Section 5.3.5.4 on page 60, two power calculations were conducted to estimate the statistical power at the two extreme significance levels of FDR. For the analysis of overall prostate cancer risk including 1,077 matched case-control set, we had more than 80% power to detect OR_{1SD} between 1.13 and 1.21 or 0.83 and 0.88. For the analysis of advanced prostate cancer including 208 cases, the corresponding numbers were 1.32 to 1.55 and 0.64 to 0.76, while for the analyses of high grade, aggressive disease and death from prostate cancer, including 115 to 127 cases, we had more than 80% power to detect OR_{1SD} between 1.44 to 1.83 or 0.55 to 0.69. To facilitate more robust analyses, assays of blood from additional cases and controls are planned in EPIC.

6.6 Conclusion

Findings from this large study of pre-diagnostic plasma metabolites and prostate cancer risk indicate that several metabolites, including acylcarnitines, amino acids, glycerophospholipids and sphingolipids, might be related to prostate cancer risk. Analyses stratifying for time to diagnosis suggested that low concentrations of citrulline might be a marker of sub-clinical prostate cancer, while other metabolites might be related to aetiology. Higher concentrations of several glycerophospholipids might be associated with lower risk of advanced stage prostate cancer. These results need to be investigated further in other large prospective studies with data on prostate tumour characteristics and death.

6.7 Tables

Table 6.1: Completeness and coefficients of variation for the metabolomics assay and corresponding exclusions of metabolites for assay round I

Metabolites[1]	Non-quantifiable results[2]				CV[3], %
	Missing data	Outside the measurable range		Total, %	
		Samples, %	Type (conc, µmol/L)		
Acylcarnitines					
C0	0	0.05	<LOD	0.05	7.02
C2	0	0.97	<LOD	0.97	9.26
C3	0	14.39	<LOD	14.39	10.29
C4	0	23.00	<LOD	**23.00**	13.41
C5	0	30.50	<LOD	**30.50**	8.43
C8	0	98.24	<LOD	**98.24**	-
C10	0	88.80	<LOD	**88.80**	-
C10:1	0	44.56	<LOD	**44.56**	6.56
C12	0	54.93	<LOD	**54.93**	10.01
C12:1	0	48.31	<LOD	**48.31**	10.34
C14	0	61.08	<LOD	**61.08**	-
C14:1	0	11.15	<LOD	11.15	16.65
C14:2	0	48.54	<LOD	**48.54**	14.15
C16	0	0.60	<LOD	0.60	13.83
C16:1	0	92.27	<LOD	**92.27**	-
C18	0	15.13	<LOD	**15.13**	15.81
C18:1	0	0.23	<LOD	0.23	11.69
C18:2	0	3.89	<LOD	3.89	17.19
Amino acids					
Alanine	0	0		0	5.99
Arginine	0	0.65	<LOQ (5)	0.65	7.28
Asparagine	0	0		0	7.69
Aspartate	0	22.44	<LOQ (5)	**22.44**	15.45
Citrulline	0	0		0	9.44
Glutamate	0	0		0	7.93
Glutamine	0	0		0	8.98
Glycine	0	0		0	8.18
Histidine	0	0		0	7.11
Isoleucine	0	0		0	7.77
Leucine	0	0.05	<LOQ (50)	0.05	7.55
Lysine	0	0		0	11.93
Methionine	0	0.19	<LOQ (5)	0.19	12.63
Ornithine	0	0		0	11.71
Phenylalanine	0	0		0	8.31
Proline	0	0		0	8.87
Serine	0	0		0	8.49
t4-OH-Pro	0	0		0	9.49
Threonine	0	0		0	5.95
Tryptophane	0	0		0	9.66
Tyrosine	0	0		0	8.92
Valine	0	0		0	11.36
Biogenic amines					
ADMA	0.05	4.44	<LOQ (0.25)	4.49	17.29
alpha-AAA	0	66.08	<LOQ (1)	**66.08**	22.57

Continues...

Table 6.1 continued

Metabolites[1]	Non-quantifiable results[2]				CV[3], %
	Missing data	Outside the measurable range		Total, %	
		Samples, %	Type (conc, µmol/L)		
Creatinine	0	0		0	5.01
Kynurenine	0	0.37	<LOQ (1)	0.37	11.42
Putrescine	0	74.46	<LOQ (0.1)	**74.46**	-
Sarcosine	0	13.79	<LOQ (1)	13.79	9.06
SDMA	0	0		0	16.16
Serotonin	0	85.15	<LOQ (0.1)	**85.15**	-
Spermidine	0	93.24	<LOQ (0.25)	**93.24**	-
Spermine	0	96.81	<LOQ (0.25)	**96.81**	-
Taurine	0	0.05	> highest std. (200)	0.05	4.31
Glycerophospholipids					
lysoPC a C16:0	0	0		0	8.45
lysoPC a C16:1	0	0		0	8.53
lysoPC a C17:0	0	0		0	8.70
lysoPC a C18:0	0	0		0	10.27
lysoPC a C18:1	0	0		0	9.08
lysoPC a C18:2	0	0		0	8.45
lysoPC a C20:3	0	0		0	13.60
lysoPC a C20:4	0	0		0	12.03
lysoPC a C28:1	0	0		0	**26.20**
PC aa C28:1	0	0		0	7.59
PC aa C30:0	0	0		0	7.29
PC aa C32:0	0	0		0	9.46
PC aa C32:1	0	0		0	12.05
PC aa C32:2	0	0		0	16.66
PC aa C32:3	0	0		0	11.30
PC aa C34:1	0	0		0	11.52
PC aa C34:2	0	0		0	10.03
PC aa C34:3	0	0		0	9.18
PC aa C34:4	0	0		0	14.16
PC aa C36:0	0	0		0	10.73
PC aa C36:1	0	0		0	8.63
PC aa C36:2	0	0		0	9.07
PC aa C36:3	0	0		0	9.24
PC aa C36:4	0	0		0	10.84
PC aa C36:5	0	0		0	11.04
PC aa C36:6	0	0		0	12.83
PC aa C38:0	0	0		0	12.78
PC aa C38:3	0	0		0	13.07
PC aa C38:4	0	0		0	8.66
PC aa C38:5	0	0		0	8.49
PC aa C38:6	0	0		0	8.16
PC aa C40:1	0	68.16	<LOD	**68.16**	-
PC aa C40:2	0	0		0	14.70
PC aa C40:3	0	0		0	16.21
PC aa C40:4	0	0		0	8.15
PC aa C40:5	0	0		0	10.55
PC aa C40:6	0	0		0	9.59
PC aa C42:0	0	0		0	13.93
PC aa C42:1	0	0		0	17.61
PC aa C42:2	0	0.32	<LOD	0.32	14.44
PC aa C42:4	0	0		0	13.36
PC aa C42:5	0	0		0	14.23
PC aa C42:6	0	3.15	<LOD	3.15	13.60
PC ae C30:0	0	0.88	<LOD	0.88	10.28
PC ae C30:2	0	0		0	15.98

Continues...

Table 6.1 continued

Metabolites[1]	Non-quantifiable results[2]				CV[3], %
	Missing data	Outside the measurable range		Total, %	
		Samples, %	Type (conc, µmol/L)		
PC ae C32:1	0	0		0	10.55
PC ae C32:2	0	0		0	11.51
PC ae C34:0	0	0		0	11.39
PC ae C34:1	0	0		0	9.99
PC ae C34:2	0	0		0	10.24
PC ae C34:3	0	0		0	10.48
PC ae C36:0	0	0		0	15.99
PC ae C36:1	0	0		0	8.55
PC ae C36:2	0	0		0	11.39
PC ae C36:3	0	0		0	13.16
PC ae C36:4	0	0		0	9.14
PC ae C36:5	0	0		0	9.71
PC ae C38:0	0	0		0	10.53
PC ae C38:2	0	0		0	10.64
PC ae C38:3	0	0		0	8.37
PC ae C38:4	0	0		0	8.71
PC ae C38:5	0	0		0	9.67
PC ae C38:6	0	0		0	11.46
PC ae C40:1	0	0		0	14.76
PC ae C40:2	0	0		0	12.46
PC ae C40:3	0	0		0	9.05
PC ae C40:4	0	0		0	8.59
PC ae C40:5	0	0		0	11.16
PC ae C40:6	0	0		0	9.10
PC ae C42:1	0	0.69	<LOD	0.69	11.38
PC ae C42:2	0	0		0	11.21
PC ae C42:3	0	0		0	12.43
PC ae C42:4	0	0		0	11.69
PC ae C42:5	0	0		0	10.32
PC ae C44:3	0	2.55	<LOD	2.55	**22.19**
PC ae C44:4	0	0.14	<LOD	0.14	16.86
PC ae C44:5	0	0		0	11.09
PC ae C44:6	0	0		0	9.72
Hexose					
Hexose	0	3.52	<LOQ (1500)	3.52	6.53
Sphingolipids					
SM (OH) C14:1	0	0		0	8.56
SM (OH) C16:1	0	0		0	9.66
SM (OH) C22:1	0	0		0	9.74
SM (OH) C22:2	0	0		0	8.90
SM (OH) C24:1	0	0		0	13.59
SM C16:0	0	0		0	9.11
SM C16:1	0	0		0	8.64
SM C18:0	0	0		0	8.88
SM C18:1	0	0		0	8.00
SM C20:2	0	0		0	19.84
SM C24:0	0	0		0	9.14
SM C24:1	0	0		0	10.33

[1] Metabolites marked in grey were excluded from the statistical analyses because $>$ 15% of samples had non-quantifiable assay results or the coefficient of variation (CV) was $>$ 20%. The reason for exclusion is marked in bold

[2] Limit of detection (LOD) is applicable to acylcarnitines, glycerophospholipids, sphingolipids; limit of quantification (LOQ) is applicable to amino acids, biogenic amines and hexose

[3] CV could not be calculated for all metabolites because too many quality control samples having assay results outside the measurable range

Table 6.2: Characteristics of 1,077 prostate cancer cases and 1,077 controls in EPIC

Characteristic	Cases, n = 1,077	Controls, n = 1,077
Age at blood collection, years (SD)	60.0 (7.1)	60.0 (7.1)
Height, cm (SD)[1]	171.7 (6.9)	172.1 (7.1)
BMI, kg/m² (SD)[1]	26.8 (3.5)	26.9 (3.5)
Smoking, n (%)[1]		
Never	349 (33.0)	307 (29.0)
Former	467 (44.1)	501 (47.3)
Current	242 (22.9)	252 (23.8)
Alcohol consumption, n (%)[1]		
<10 g/day	479 (44.8)	480 (44.6)
10-19 g/day	199 (18.6)	201 (18.7)
20-40 g/day	218 (20.4)	221 (20.5)
≥40 g/day	174 (16.3)	174 (16.2)
Physical activity, n (%)[1]		
Inactive	292 (28.0)	273 (26.0)
Moderately inactive	328 (31.4)	326 (31.1)
Moderately active	242 (23.2)	246 (23.5)
Active	182 (17.4)	204 (19.4)
Marital status, n (%)[1]		
Married or cohabiting	718 (89.4)	730 (90.2)
Not married or cohabiting	85 (10.6)	79 (9.8)
Educational attainment, n (%)[1]		
Primary or equivalent	413 (41.1)	425 (42.2)
Secondary	369 (36.7)	396 (39.4)
Degree	223 (22.2)	185 (18.4)
Cases only		
Age at diagnosis, mean years (SD)	66.9 (7.0)	–
Time to diagnosis, n (%)[2]		
<2 years	137 (12.7)	–
2-<4 years	177 (16.4)	–
4-<6 years	191 (17.7)	–
6-<8 years	120 (11.1)	–
8-<10 years	136 (12.6)	–
≥10 years	316 (29.3)	–
Year of diagnosis, median (range)	2001 (1994-2008)	–
Grade, n (%)[1, 3]		
Low-intermediate grade	778 (86.3)	–
High grade	124 (13.7)	–
Stage, n (%)[1, 4]		
Localised	456 (68.7)	–
Advanced	208 (31.3)	–
Non-aggressive	549 (82.7)	–
Aggressive	115 (17.3)	–
Death from prostate cancer, n (%)[1, 5]	127 (12.3)	–

[1] Unknown values for some participants; the calculations of percentages exclude missing values

[2] Time between blood collection and diagnosis

[3] Gleason score $<$8 or coded as well, moderately or poorly differentiated for low-intermediate grade and Gleason score $\geq$8 or coded as undifferentiated for high grade

[4] The TNM system was used to categorise stages of prostate cancer; localised: $\leq$ T2 and N0/X and M0, or coded as localised; advanced: T3 $-$ 4 and/or N1 $-$ 3 and/or M1, or coded as advanced; and aggressive: T4 and/or N1 $-$ 3 and/or M1. All categories are not mutually exclusive so numbers do not add up; percentages were calculated separately for localised and advanced, and for non-aggressive and aggressive

[5] Death during follow-up with prostate cancer listed as the underlying cause of death on the death certificate. 144 died of prostate cancer; 17 were excluded from further analysis as their matched control had died before them (n $=$ 13) or vital status was unknown for the control (n $=$ 4)

Table 6.3: Distribution of metabolite concentrations by case-control status in EPIC

Metabolites	Controls							Cases						
	Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data				Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data			
		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]
Acylcarnitines														
C0	32.61	32.23	32.99	3.48	0.19	-0.54	5.76	32.25	31.85	32.65	3.47	0.21	-1.27	11.93
C2	5.10	5.00	5.20	1.63	0.34	-0.48	5.98	5.00	4.90	5.11	1.61	0.34	-0.51	5.57
C3	0.28	0.27	0.29	-1.27	0.69	-1.37	3.70	0.29	0.28	0.30	-1.25	0.68	-1.42	3.99
C14:1	0.04	0.04	0.05	-3.12	0.71	-1.18	4.04	0.04	0.04	0.04	-3.16	0.73	-1.06	3.87
C16	0.09	0.09	0.10	-2.36	0.33	-0.41	9.62	0.09	0.09	0.10	-2.36	0.37	-0.19	11.91
C18:1	0.11	0.10	0.11	-2.24	0.38	0.61	5.10	0.10	0.10	0.11	-2.27	0.41	0.53	5.87
C18:2	0.04	0.04	0.04	-3.21	0.60	-0.63	6.04	0.04	0.04	0.04	-3.23	0.61	-0.39	6.21
Alanine	332.43	327.69	337.24	5.81	0.24	-0.02	2.91	332.63	327.74	337.60	5.81	0.25	-0.08	2.84
Arginine	46.42	44.89	48.00	3.84	0.56	-1.96	8.63	47.03	45.48	48.63	3.85	0.56	-2.18	9.97
Asparagine	39.23	38.77	39.70	3.67	0.20	0.31	3.17	39.45	38.96	39.96	3.68	0.21	0.30	3.42
Citrulline	28.82	28.36	29.28	3.36	0.27	-0.44	4.52	28.28	27.83	28.74	3.34	0.27	-0.22	3.60
Glutamate	51.10	49.57	52.67	3.93	0.51	0.59	3.53	51.39	49.81	53.02	3.94	0.52	0.67	3.81
Glutamine	607.53	602.07	613.04	6.41	0.15	-0.16	3.35	608.69	603.04	614.39	6.41	0.16	-0.53	4.44
Glycine	172.07	169.77	174.41	5.15	0.23	0.38	3.42	173.32	170.92	175.75	5.16	0.23	0.52	4.06
Histidine	68.41	67.77	69.05	4.23	0.16	0.40	4.62	68.56	67.90	69.22	4.23	0.16	0.21	3.71
Isoleucine	66.72	65.73	67.71	4.20	0.25	0.43	3.88	68.11	67.05	69.19	4.22	0.26	0.13	3.18
Leucine	124.85	123.20	126.52	4.83	0.22	0.37	3.85	126.49	124.68	128.32	4.84	0.24	-0.16	4.87
Lysine	182.01	179.62	184.44	5.20	0.22	-0.06	3.63	182.03	179.55	184.54	5.20	0.23	-0.26	3.65
Methionine	18.68	18.41	18.96	2.93	0.25	0.24	4.33	18.81	18.50	19.13	2.93	0.28	-1.37	12.19
Ornithine	79.32	77.73	80.94	4.37	0.34	0.27	2.96	79.38	77.70	81.11	4.37	0.36	0.28	3.12
Phenylalanine	55.98	55.40	56.57	4.02	0.17	0.25	3.41	56.60	55.94	57.27	4.04	0.20	0.43	5.92
Proline	181.50	178.41	184.64	5.20	0.29	0.09	3.08	185.82	182.57	189.14	5.22	0.30	0.19	3.32
Serine	81.56	80.51	82.62	4.40	0.22	0.07	3.30	82.32	81.22	83.43	4.41	0.22	0.09	3.53
t4-OH-Pro	7.98	7.78	8.19	2.08	0.43	0.58	3.77	8.29	8.07	8.52	2.12	0.45	0.57	3.27
Threonine	103.15	101.79	104.53	4.64	0.22	-0.16	3.52	104.12	102.75	105.52	4.65	0.22	-0.34	3.51
Tryptophan	50.76	50.20	51.32	3.93	0.18	-0.29	3.74	50.90	50.30	51.50	3.93	0.20	-0.54	4.84
Tyrosine	61.38	60.54	62.22	4.12	0.23	0.03	3.32	61.37	60.49	62.26	4.12	0.24	0.11	3.03
Valine	208.30	205.80	210.83	5.34	0.20	0.01	3.41	210.26	207.65	212.91	5.35	0.21	-0.11	3.18

Continues...

Table 6.3 continued

Metabolites	Controls							Cases						
	Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data				Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data			
		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]
ADMA	0.37	0.36	0.37	-1.00	0.32	-1.68	6.96	0.37	0.36	0.38	-0.99	0.30	-1.54	7.34
Creatinine	69.08	68.34	69.83	4.24	0.18	0.31	4.47	69.62	68.83	70.41	4.24	0.19	1.26	12.95
Kynurenine	2.02	1.99	2.06	0.71	0.26	-0.50	5.80	2.05	2.02	2.08	0.72	0.26	-0.32	6.54
Sarcosine	1.33	1.29	1.37	0.28	0.48	-0.71	3.32	1.34	1.30	1.37	0.29	0.47	-0.74	3.51
SDMA	0.38	0.38	0.39	-0.96	0.22	0.19	3.82	0.38	0.38	0.39	-0.97	0.24	0.82	8.85
Taurine	47.92	46.95	48.91	3.87	0.34	0.87	3.97	47.63	46.66	48.63	3.86	0.34	0.85	4.15
Glycerophoapholipids														
lysoPC a C16:0	96.95	95.60	98.32	4.57	0.23	0.29	3.58	96.64	95.21	98.09	4.57	0.25	0.48	3.79
lysoPC a C16:1	2.79	2.74	2.84	1.03	0.32	0.15	3.12	2.74	2.69	2.80	1.01	0.33	0.17	3.24
lysoPC a C17:0	1.76	1.72	1.79	0.56	0.32	-0.06	3.04	1.78	1.74	1.81	0.58	0.34	-0.17	3.61
lysoPC a C18:0	31.30	30.83	31.77	3.44	0.25	0.13	3.66	31.00	30.49	31.51	3.43	0.27	0.32	3.63
lysoPC a C18:1	22.12	21.72	22.54	3.10	0.31	0.06	3.08	21.75	21.36	22.14	3.08	0.30	0.00	3.06
lysoPC a C18:2	31.40	30.79	32.01	3.45	0.33	0.06	3.11	31.42	30.80	32.05	3.45	0.33	0.02	2.94
lysoPC a C20:3	2.20	2.15	2.24	0.79	0.35	-0.09	2.93	2.17	2.12	2.21	0.77	0.36	-0.25	3.59
lysoPC a C20:4	6.12	6.00	6.24	1.81	0.34	-0.06	3.02	6.08	5.96	6.21	1.81	0.35	-0.10	3.06
PC aa C28:1	2.08	2.04	2.11	0.73	0.27	-0.33	3.15	2.13	2.10	2.17	0.76	0.27	-0.21	3.01
PC aa C30:0	2.91	2.84	2.97	1.07	0.36	-0.05	2.95	2.95	2.89	3.02	1.08	0.35	-0.16	2.95
PC aa C32:0	20.22	19.95	20.49	3.01	0.22	0.01	3.05	20.35	20.09	20.61	3.01	0.21	-0.03	3.04
PC aa C32:1	21.92	21.22	22.63	3.09	0.54	0.00	2.99	21.83	21.16	22.51	3.08	0.52	0.05	3.23
PC aa C32:2	5.47	5.35	5.60	1.70	0.39	-0.53	3.84	5.50	5.38	5.63	1.71	0.39	-0.34	3.50
PC aa C32:3	0.73	0.72	0.75	-0.31	0.29	0.12	2.90	0.73	0.72	0.75	-0.31	0.28	0.21	2.82
PC aa C34:1	268.00	263.07	273.03	5.59	0.31	-0.12	3.13	266.29	261.67	271.00	5.58	0.29	-0.25	3.49
PC aa C34:2	574.47	567.85	581.16	6.35	0.19	-0.13	3.11	574.90	568.41	581.47	6.35	0.19	0.00	3.07
PC aa C34:3	20.55	20.14	20.97	3.02	0.34	-0.08	2.74	20.39	19.99	20.81	3.02	0.34	0.06	3.00
PC aa C34:4	2.19	2.14	2.24	0.78	0.37	-0.33	4.34	2.19	2.14	2.23	0.78	0.36	-0.13	2.88
PC aa C36:0	2.97	2.92	3.03	1.09	0.29	-0.06	3.16	3.00	2.94	3.05	1.10	0.30	0.03	3.08
PC aa C36:1	46.46	45.66	47.28	3.84	0.29	0.03	3.18	45.98	45.22	46.76	3.83	0.28	0.00	3.39
PC aa C36:2	308.10	304.52	311.73	5.73	0.20	-0.28	3.66	305.02	301.46	308.61	5.72	0.20	-0.14	3.25
PC aa C36:3	165.85	163.68	168.04	5.11	0.22	-0.18	3.15	163.36	161.27	165.48	5.10	0.22	0.10	2.92

Continues...

Table 6.3 continued

Metabolites	Controls							Cases						
	Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data				Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data			
		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]
PC aa C36:4	215.98	212.53	219.48	5.38	0.27	-0.05	3.34	215.03	211.62	218.49	5.37	0.27	-0.08	3.14
PC aa C36:5	28.74	27.85	29.66	3.36	0.53	0.31	3.38	29.30	28.35	30.29	3.38	0.55	0.32	3.14
PC aa C36:6	1.16	1.14	1.19	0.15	0.42	0.01	3.43	1.18	1.16	1.22	0.17	0.42	0.14	2.96
PC aa C38:0	2.87	2.82	2.92	1.05	0.30	-0.06	2.90	2.88	2.83	2.94	1.06	0.32	0.00	3.33
PC aa C38:3	42.64	41.93	43.35	3.75	0.28	0.00	2.98	41.92	41.22	42.64	3.74	0.28	0.00	3.03
PC aa C38:4	93.31	91.77	94.88	4.54	0.28	-0.15	3.58	91.79	90.26	93.34	4.52	0.28	-0.13	2.97
PC aa C38:5	49.34	48.53	50.16	3.90	0.28	0.03	3.24	49.09	48.28	49.92	3.89	0.28	0.15	3.28
PC aa C38:6	87.22	85.56	88.91	4.47	0.32	-0.18	3.25	87.96	86.31	89.65	4.48	0.32	-0.17	3.31
PC aa C40:2	0.26	0.25	0.26	-1.35	0.33	0.51	3.51	0.26	0.25	0.26	-1.36	0.35	0.85	4.48
PC aa C40:3	0.44	0.43	0.45	-0.82	0.34	0.50	3.88	0.44	0.43	0.45	-0.82	0.35	0.82	4.42
PC aa C40:4	2.36	2.32	2.40	0.86	0.28	-0.02	3.62	2.33	2.29	2.37	0.84	0.27	0.11	2.97
PC aa C40:5	5.76	5.65	5.88	1.75	0.33	-0.09	2.97	5.69	5.58	5.80	1.74	0.32	-0.04	2.97
PC aa C40:6	24.03	23.54	24.52	3.18	0.34	-0.13	3.28	23.95	23.47	24.45	3.18	0.34	-0.11	3.27
PC aa C42:0	0.29	0.28	0.29	-1.25	0.30	-0.11	2.81	0.29	0.29	0.30	-1.23	0.30	0.03	3.23
PC aa C42:1	0.15	0.14	0.15	-1.92	0.29	-0.19	2.78	0.15	0.15	0.15	-1.90	0.31	0.04	3.03
PC aa C42:2	0.14	0.14	0.15	-1.94	0.32	-0.41	6.00	0.14	0.14	0.15	-1.93	0.32	0.08	6.02
PC aa C42:4	0.09	0.09	0.09	-2.41	0.24	-0.20	3.73	0.09	0.09	0.09	-2.42	0.25	0.04	3.39
PC aa C42:5	0.20	0.20	0.21	-1.60	0.36	0.11	2.95	0.20	0.20	0.20	-1.61	0.36	0.37	3.28
PC aa C42:6	0.29	0.28	0.29	-1.24	0.35	-1.60	8.58	0.29	0.28	0.29	-1.25	0.38	-1.40	7.66
PC ae C30:0	0.26	0.26	0.27	-1.33	0.34	-0.62	5.16	0.27	0.27	0.28	-1.30	0.33	-0.54	4.95
PC ae C30:2	0.06	0.06	0.06	-2.76	0.29	-0.10	2.84	0.07	0.06	0.07	-2.73	0.28	-0.02	2.79
PC ae C32:1	3.70	3.65	3.75	1.31	0.23	-0.19	3.53	3.74	3.69	3.79	1.32	0.22	-0.18	3.45
PC ae C32:2	1.07	1.05	1.08	0.06	0.24	-0.16	3.34	1.07	1.06	1.09	0.07	0.24	0.04	3.37
PC ae C34:0	1.94	1.90	1.97	0.66	0.30	-0.12	3.30	1.98	1.95	2.01	0.68	0.29	-0.24	3.68
PC ae C34:1	12.62	12.44	12.80	2.54	0.24	-0.13	3.26	12.73	12.55	12.90	2.54	0.23	-0.15	3.09
PC ae C34:2	15.53	15.31	15.75	2.74	0.24	0.03	3.34	15.73	15.51	15.94	2.76	0.23	-0.04	3.09
PC ae C34:3	10.28	10.11	10.45	2.33	0.27	-0.03	3.38	10.29	10.12	10.46	2.33	0.27	0.06	3.05
PC ae C36:0	0.98	0.97	1.00	-0.02	0.28	0.26	3.90	1.00	0.98	1.01	0.00	0.28	0.52	4.56
PC ae C36:1	9.26	9.12	9.40	2.23	0.26	-0.06	3.07	9.37	9.23	9.51	2.24	0.25	-0.06	3.28
PC ae C36:2	17.80	17.53	18.08	2.88	0.26	0.00	2.91	18.01	17.73	18.28	2.89	0.26	-0.21	3.58

Continues...

Table 6.3 continued

Metabolites	Controls							Cases						
	Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data				Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data			
		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]
PC ae C36:3	9.44	9.31	9.57	2.25	0.24	-0.01	3.39	9.47	9.34	9.60	2.25	0.23	-0.03	2.98
PC ae C36:4	19.23	18.93	19.54	2.96	0.27	0.16	3.11	19.30	19.00	19.61	2.96	0.26	-0.01	3.22
PC ae C36:5	13.66	13.43	13.90	2.61	0.28	-0.04	3.28	13.69	13.46	13.93	2.62	0.28	0.04	3.26
PC ae C38:0	2.32	2.27	2.36	0.84	0.33	0.12	2.98	2.33	2.28	2.37	0.84	0.34	0.15	2.80
PC ae C38:2	2.17	2.14	2.20	0.78	0.26	-0.04	2.93	2.18	2.15	2.21	0.78	0.25	0.04	2.91
PC ae C38:3	3.90	3.84	3.96	1.36	0.24	-0.14	2.89	3.93	3.88	3.99	1.37	0.24	-0.10	3.16
PC ae C38:4	12.66	12.48	12.85	2.54	0.24	-0.15	3.24	12.72	12.54	12.91	2.54	0.24	-0.35	3.37
PC ae C38:5	17.53	17.30	17.77	2.86	0.23	-0.01	3.10	17.51	17.28	17.75	2.86	0.22	-0.05	3.03
PC ae C38:6	7.95	7.83	8.09	2.07	0.27	-0.12	3.41	8.02	7.88	8.15	2.08	0.28	0.08	3.16
PC ae C40:1	1.16	1.14	1.18	0.15	0.27	-0.22	3.16	1.16	1.14	1.18	0.15	0.27	0.00	2.99
PC ae C40:2	1.37	1.35	1.39	0.32	0.27	-0.05	2.82	1.39	1.37	1.41	0.33	0.27	-0.04	3.19
PC ae C40:3	0.72	0.71	0.73	-0.33	0.22	-0.02	3.01	0.73	0.72	0.74	-0.32	0.23	0.20	3.07
PC ae C40:4	1.92	1.90	1.94	0.65	0.21	-0.05	3.10	1.94	1.91	1.96	0.66	0.21	-0.15	2.96
PC ae C40:5	2.84	2.80	2.87	1.04	0.22	-0.04	3.10	2.86	2.82	2.90	1.05	0.22	0.01	3.28
PC ae C40:6	4.01	3.95	4.08	1.39	0.25	-0.07	2.83	4.07	4.01	4.14	1.40	0.27	-0.05	3.17
PC ae C42:1	0.29	0.29	0.30	-1.23	0.24	-1.36	10.73	0.29	0.29	0.29	-1.24	0.22	-0.83	8.94
PC ae C42:2	0.37	0.37	0.38	-0.99	0.25	0.04	3.00	0.37	0.36	0.38	-0.99	0.26	-0.07	3.46
PC ae C42:3	0.48	0.47	0.49	-0.74	0.26	-0.14	2.89	0.48	0.47	0.49	-0.73	0.26	-0.18	3.12
PC ae C42:4	0.50	0.49	0.51	-0.70	0.26	-0.15	3.18	0.50	0.50	0.51	-0.69	0.26	-0.06	3.28
PC ae C42:5	1.34	1.32	1.36	0.29	0.22	0.00	3.12	1.35	1.34	1.37	0.30	0.22	-0.10	3.17
PC ae C44:4	0.17	0.16	0.17	-1.79	0.31	-3.45	43.46	0.17	0.16	0.17	-1.78	0.30	-2.06	28.82
PC ae C44:5	0.69	0.68	0.70	-0.37	0.28	-0.03	3.32	0.70	0.68	0.71	-0.36	0.28	-0.16	3.44
PC ae C44:6	0.59	0.58	0.60	-0.53	0.27	-0.06	2.95	0.60	0.59	0.61	-0.51	0.28	-0.03	3.05
Hexose														
Hexose	3589	3499	3680	8.19	0.42	-1.39	8.29	3614	3527	3704	8.19	0.41	-1.75	8.83
Sphingolipids														
SM (OH) C14:1	3.38	3.33	3.44	1.22	0.28	-0.33	3.14	3.49	3.43	3.55	1.25	0.28	-0.32	3.42
SM (OH) C16:1	1.58	1.56	1.61	0.46	0.25	-0.36	3.52	1.62	1.59	1.64	0.48	0.26	-0.46	3.99
SM (OH) C22:1	4.50	4.44	4.56	1.50	0.23	-0.39	3.68	4.51	4.45	4.58	1.51	0.24	-0.62	4.46
SM (OH) C22:2	3.56	3.51	3.61	1.27	0.24	-0.41	3.58	3.61	3.55	3.66	1.28	0.24	-0.46	3.84
SM (OH) C24:1	0.33	0.32	0.33	-1.12	0.28	-0.34	3.83	0.33	0.32	0.33	-1.11	0.30	-0.44	3.78

Continues...

Table 6.3 continued

Metabolites	Controls							Cases						
	Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data				Geometric mean, µmol/L	95% CI, µmol/L		Logarithmically-transformed data			
		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]		LL	UL	Mean, ln(µmol/L)	SD, ln(µmol/L)	Skew[1]	Kurt[2]
SM C16:0	61.86	61.19	62.54	4.12	0.18	-0.51	4.21	62.37	61.70	63.05	4.13	0.18	-0.35	3.55
SM C16:1	8.46	8.37	8.56	2.14	0.18	-0.38	4.07	8.50	8.41	8.59	2.14	0.18	-0.19	3.34
SM C18:0	10.22	10.08	10.36	2.32	0.23	-0.35	3.87	10.31	10.17	10.45	2.33	0.23	-0.23	2.97
SM C18:1	4.96	4.90	5.03	1.60	0.22	-0.35	3.66	4.97	4.91	5.04	1.60	0.22	-0.14	3.15
SM C20:2	0.27	0.26	0.27	-1.33	0.54	0.09	3.35	0.27	0.26	0.27	-1.32	0.52	-0.07	4.22
SM C24:0	6.11	6.03	6.19	1.81	0.23	-0.04	3.97	6.12	6.03	6.20	1.81	0.23	-0.17	3.42
SM C24:1	13.38	13.19	13.57	2.59	0.24	-0.39	3.55	13.49	13.30	13.68	2.60	0.24	-0.29	3.44

[1] Skewness (Skew) is a measure of the non-symmetry of a distribution. The normal distribution has a skewness of 0, while negative and positive skewness indicte left and right skewed distributions, respectively

[2] Kurtosis (Kurt) is a measure of how peaked the distribution is. The normal distribution has a kurtosis of 3, while a kurtosis above and below 3 indicate that the distribution is narrower and wider, respectively, than the normal distribution

Table 6.4: Partial correlation coefficients between concentrations of total PSA and metabolites in EPIC[1]

Metabolites	Controls (n=764) mean total PSA 1.21ng/ml			Controls with PSA < 1ng/ml (n=489) mean total PSA 0.55ng/ml			Cases (n=768) mean total PSA 5.59 ng/ml		
	r	p^2	p_{adj}	r	p^2	p_{adj}	r	p^2	p_{adj}
Acylcarnitines									
C0	-0.090	0.015	0.977	-0.055	0.237	0.987	-0.008	0.831	0.874
C2	-0.035	0.336	0.977	-0.077	0.098	0.987	0.081	0.027	0.232
C3	0.005	0.900	0.977	0.008	0.870	0.987	-0.083	0.023	0.213
C14:1	-0.072	0.049	0.977	-0.057	0.219	0.987	-0.046	0.206	0.648
C16	0.018	0.622	0.977	-0.020	0.673	0.987	0.120	**0.001**	0.021
C18:1	-0.008	0.822	0.977	-0.015	0.753	0.987	0.046	0.214	0.648
C18:2	-0.005	0.893	0.977	0.005	0.923	0.987	0.011	0.764	0.862
Amino acids									
Alanine	0.038	0.297	0.977	0.037	0.431	0.987	0.007	0.856	0.892
Arginine	0.019	0.600	0.977	0.004	0.936	0.987	-0.033	0.370	0.769
Asparagine	-0.026	0.475	0.977	0.022	0.629	0.987	0.023	0.528	0.796
Citrulline	-0.018	0.625	0.977	0.089	0.055	0.987	0.017	0.635	0.862
Glutamate	-0.023	0.539	0.977	0.017	0.711	0.987	0.108	**0.003**	0.043
Glutamine	0.017	0.634	0.977	-0.030	0.524	0.987	-0.024	0.508	0.795
Glycine	0.072	0.049	0.977	0.005	0.910	0.987	0.030	0.410	0.769
Histidine	-0.029	0.429	0.977	-0.044	0.344	0.987	-0.002	0.962	0.962
Isoleucine	-0.008	0.830	0.977	0.001	0.990	0.990	0.028	0.437	0.769
Leucine	-0.012	0.750	0.977	-0.019	0.686	0.987	0.026	0.481	0.782
Lysine	0.025	0.495	0.977	-0.016	0.731	0.987	-0.029	0.422	0.769
Methionine	-0.028	0.441	0.977	-0.003	0.953	0.987	-0.015	0.674	0.862
Ornithine	0.024	0.508	0.977	0.018	0.694	0.987	0.030	0.409	0.769
Phenylalanine	-0.049	0.185	0.977	-0.043	0.355	0.987	0.040	0.280	0.727
Proline	0.011	0.769	0.977	0.008	0.856	0.987	-0.037	0.311	0.727
Serine	0.005	0.897	0.977	-0.006	0.898	0.987	-0.027	0.460	0.769
t4-OH-Pro	-0.016	0.662	0.977	-0.072	0.121	0.987	0.068	0.064	0.393
Threonine	-0.023	0.530	0.977	-0.016	0.725	0.987	0.048	0.187	0.648
Tryptophan	-0.023	0.532	0.977	-0.041	0.378	0.987	-0.031	0.396	0.769
Tyrosine	-0.057	0.124	0.977	-0.069	0.135	0.987	-0.009	0.815	0.868
Valine	-0.008	0.832	0.977	0.006	0.8963	0.9875	0.029	0.4311	0.7692

Continues...

Table 6.4 continued

Metabolites	Controls (n=764) mean total PSA 1.21ng/ml			Controls with PSA < 1ng/ml (n=489) mean total PSA 0.55ng/ml			Cases (n=768) mean total PSA 5.59 ng/ml		
	r	p^2	p_{adj}	r	p^2	p_{adj}	r	p^2	p_{adj}
Biogenic amines									
ADMA	0.021	0.563	0.977	0.065	0.1634	0.9875	0.035	0.3358	0.7546
Creatinine	0.066	0.074	0.977	0.031	0.5061	0.9875	0.134	**0.0003**	0.0082
Kynurenine	0.039	0.287	0.977	0.012	0.7970	0.9875	0.089	0.0146	0.1619
Sarcosine	-0.021	0.566	0.977	0.010	0.8277	0.9875	0.016	0.6537	0.8617
SDMA	0.070	0.055	0.977	-0.002	0.9713	0.9875	0.116	**0.0015**	0.0257
Taurine	-0.002	0.946	0.977	-0.008	0.8659	0.9875	0.085	0.0202	0.2050
Glycerophospholipids									
lysoPC a C16:0	0.041	0.269	0.977	-0.027	0.5659	0.9875	0.064	0.0799	0.4602
lysoPC a C16:1	-0.010	0.786	0.977	-0.040	0.3941	0.9875	0.051	0.1609	0.6479
lysoPC a C17:0	0.038	0.299	0.977	-0.054	0.2432	0.9875	-0.045	0.2171	0.6479
lysoPC a C18:0	0.008	0.821	0.977	-0.047	0.3095	0.9875	0.018	0.6263	0.8617
lysoPC a C18:1	-0.024	0.521	0.977	-0.016	0.7256	0.9875	0.012	0.7478	0.8617
lysoPC a C18:2	-0.028	0.449	0.977	-0.004	0.9309	0.9875	-0.012	0.7378	0.8617
lysoPC a C20:3	-0.046	0.214	0.977	-0.053	0.2501	0.9875	0.011	0.7699	0.8617
lysoPC a C20:4	-0.040	0.277	0.977	-0.079	0.0883	0.9875	0.042	0.2511	0.7123
PC aa C28:1	0.049	0.179	0.977	-0.023	0.6264	0.9875	-0.013	0.7214	0.8617
PC aa C30:0	0.020	0.594	0.977	-0.002	0.9579	0.9875	-0.013	0.7165	0.8617
PC aa C32:0	0.006	0.866	0.977	0.007	0.8723	0.9875	0.069	0.0609	0.3912
PC aa C32:1	0.006	0.872	0.977	-0.004	0.9372	0.9875	0.028	0.4457	0.7692
PC aa C32:2	0.020	0.582	0.977	-0.007	0.8862	0.9875	0.018	0.6270	0.8617
PC aa C32:3	0.032	0.389	0.977	0.017	0.7086	0.9875	0.024	0.5049	0.7952
PC aa C34:1	0.014	0.694	0.977	0.010	0.8340	0.9875	0.062	0.0931	0.4940
PC aa C34:2	0.014	0.707	0.977	0.004	0.9320	0.9875	0.105	0.0043	0.0520
PC aa C34:3	0.009	0.796	0.977	-0.001	0.9901	0.9902	0.027	0.4666	0.7692
PC aa C34:4	0.004	0.919	0.977	-0.019	0.6756	0.9875	0.052	0.1550	0.6479
PC aa C36:0	0.010	0.788	0.977	0.057	0.2182	0.9875	0.058	0.1155	0.5849
PC aa C36:1	-0.018	0.621	0.977	0.002	0.9683	0.9875	-0.024	0.5205	0.7958
PC aa C36:2	-0.009	0.798	0.977	-0.005	0.9078	0.9875	0.010	0.7939	0.8648
PC aa C36:3	-0.026	0.481	0.977	-0.013	0.7749	0.9875	-0.014	0.7077	0.8617

Continues...

Table 6.4 continued

Metabolites	Controls (n=764) mean total PSA 1.21ng/ml			Controls with PSA < 1ng/ml (n=489) mean total PSA 0.55ng/ml			Cases (n=768) mean total PSA 5.59 ng/ml		
	r	p^2	p_{adj}	r	p^2	p_{adj}	r	p^2	p_{adj}
PC aa C36:4	-0.008	0.821	0.977	-0.047	0.3067	0.9875	0.027	0.4569	0.7692
PC aa C36:5	0.021	0.564	0.977	0.047	0.3114	0.9875	0.046	0.2110	0.6479
PC aa C36:6	0.028	0.444	0.977	0.067	0.1492	0.9875	0.046	0.2065	0.6479
PC aa C38:0	0.027	0.460	0.977	0.060	0.1964	0.9875	0.046	0.2138	0.6479
PC aa C38:3	-0.032	0.381	0.977	-0.033	0.4758	0.9875	-0.012	0.7354	0.8617
PC aa C38:4	-0.042	0.251	0.977	-0.072	0.1219	0.9875	0.020	0.5930	0.8511
PC aa C38:5	-0.002	0.959	0.977	-0.003	0.9506	0.9875	0.034	0.3563	0.7692
PC aa C38:6	0.026	0.476	0.977	0.066	0.1556	0.9875	0.035	0.3402	0.7546
PC aa C40:2	-0.036	0.331	0.977	0.019	0.6782	0.9875	0.023	0.5364	0.7981
PC aa C40:3	-0.043	0.247	0.977	0.014	0.7679	0.9875	0.011	0.7670	0.8617
PC aa C40:4	-0.041	0.260	0.977	-0.062	0.1780	0.9875	0.017	0.6516	0.8617
PC aa C40:5	0.016	0.666	0.977	-0.006	0.9046	0.9875	0.038	0.2960	0.7274
PC aa C40:6	0.011	0.768	0.977	0.057	0.2172	0.9875	0.003	0.9419	0.9576
PC aa C42:0	0.013	0.723	0.977	0.018	0.6986	0.9875	-0.037	0.3104	0.7274
PC aa C42:1	-0.006	0.874	0.977	-0.007	0.8872	0.9875	-0.011	0.7653	0.8617
PC aa C42:2	0.004	0.920	0.977	0.004	0.9316	0.9875	0.002	0.9523	0.9601
PC aa C42:4	-0.074	0.043	0.977	-0.067	0.1482	0.9875	0.027	0.4643	0.7692
PC aa C42:5	-0.006	0.879	0.977	0.054	0.2450	0.9875	0.009	0.8052	0.8676
PC aa C42:6	0.015	0.689	0.977	-0.020	0.6692	0.9875	0.022	0.5482	0.8058
PC ae C30:0	0.039	0.295	0.977	-0.023	0.6221	0.9875	-0.012	0.7516	0.8617
PC ae C30:2	0.056	0.125	0.977	-0.057	0.2201	0.9875	0.045	0.2177	0.6479
PC ae C32:1	-0.014	0.703	0.977	-0.036	0.4335	0.9875	0.050	0.1702	0.6479
PC ae C32:2	0.002	0.958	0.977	-0.005	0.9189	0.9875	0.078	0.0321	0.2447
PC ae C34:0	0.007	0.843	0.977	-0.044	0.3376	0.9875	0.006	0.8761	0.8982
PC ae C34:1	0.021	0.566	0.977	-0.019	0.6857	0.9875	0.012	0.7382	0.8617
PC ae C34:2	0.034	0.350	0.977	-0.002	0.9632	0.9875	0.051	0.1641	0.6479
PC ae C34:3	0.033	0.365	0.977	0.010	0.8239	0.9875	-0.027	0.4636	0.7692
PC ae C36:0	0.023	0.531	0.977	0.041	0.3740	0.9875	0.042	0.2463	0.7123
PC ae C36:1	0.041	0.267	0.977	-0.021	0.6518	0.9875	-0.029	0.4306	0.7692
PC ae C36:2	0.034	0.354	0.977	-0.030	0.5170	0.9875	-0.031	0.4017	0.7692

Continues...

Table 6.4 continued

Metabolites	Controls (n=764) mean total PSA 1.21ng/ml			Controls with PSA < 1ng/ml (n=489) mean total PSA 0.55ng/ml			Cases (n=768) mean total PSA 5.59 ng/ml		
	r	p^2	p_{adj}	r	p^2	p_{adj}	r	p^2	p_{adj}
PC ae C36:3	0.012	0.738	0.977	-0.003	0.9507	0.9875	0.056	0.1279	0.6002
PC ae C36:4	0.014	0.707	0.977	-0.028	0.5392	0.9875	0.064	0.0830	0.4602
PC ae C36:5	-0.002	0.949	0.977	-0.005	0.9167	0.9875	0.040	0.2730	0.7242
PC ae C38:0	0.011	0.774	0.977	0.044	0.3402	0.9875	0.023	0.5235	0.7958
PC ae C38:2	-0.016	0.655	0.977	-0.007	0.8770	0.9875	-0.144	**0.0001**	0.0046
PC ae C38:3	0.007	0.858	0.977	-0.042	0.3648	0.9875	-0.040	0.2718	0.7242
PC ae C38:4	-0.005	0.892	0.977	-0.090	0.0528	0.9875	-0.006	0.8702	0.8982
PC ae C38:5	0.004	0.905	0.977	-0.014	0.7567	0.9875	0.017	0.6368	0.8617
PC ae C38:6	0.029	0.437	0.977	0.039	0.4018	0.9875	0.046	0.2079	0.6479
PC ae C40:1	-0.008	0.824	0.977	0.023	0.6158	0.9875	0.010	0.7878	0.8648
PC ae C40:2	0.050	0.177	0.977	0.028	0.5444	0.9875	0.030	0.4055	0.7692
PC ae C40:3	-0.001	0.981	0.981	-0.035	0.4451	0.9875	-0.041	0.2679	0.7242
PC ae C40:4	-0.015	0.690	0.977	-0.089	0.0557	0.9875	0.015	0.6771	0.8617
PC ae C40:5	0.022	0.554	0.977	-0.047	0.3154	0.9875	0.037	0.3160	0.7274
PC ae C40:6	0.041	0.265	0.977	0.027	0.5539	0.9875	0.017	0.6436	0.8617
PC ae C42:1	-0.004	0.906	0.977	-0.025	0.5838	0.9875	0.034	0.3606	0.7692
PC ae C42:2	-0.002	0.961	0.977	0.003	0.9553	0.9875	-0.014	0.6951	0.8617
PC ae C42:3	0.016	0.671	0.977	-0.008	0.8640	0.9875	-0.014	0.7061	0.8617
PC ae C42:4	-0.021	0.561	0.977	-0.116	0.0118	0.9875	-0.016	0.6690	0.8617
PC ae C42:5	-0.009	0.809	0.977	-0.056	0.2251	0.9875	-0.020	0.5797	0.8420
PC ae C44:4	0.003	0.929	0.977	0.014	0.7657	0.9875	-0.050	0.1703	0.6479
PC ae C44:5	0.005	0.888	0.977	-0.063	0.1718	0.9875	0.008	0.8178	0.8676
PC ae C44:6	-0.021	0.559	0.977	-0.058	0.2112	0.9875	-0.010	0.7899	0.8648
Hexose									
Hexose	0.001	0.969	0.977	0.030	0.5208	0.9875	-0.027	0.4607	0.7692
Sphingolipids									
SM (OH) C14:1	0.042	0.256	0.977	-0.037	0.4215	0.9875	-0.057	0.1199	0.5849
SM (OH) C16:1	0.042	0.258	0.977	-0.020	0.6713	0.9875	-0.108	**0.0032**	0.0434
SM (OH) C22:1	0.002	0.967	0.977	-0.035	0.4530	0.9875	-0.156	**0.0000**	0.0022
SM (OH) C22:2	0.042	0.258	0.977	-0.028	0.5435	0.9875	-0.131	**0.0003**	0.0082
SM (OH) C24:1	-0.009	0.802	0.977	-0.047	0.3137	0.9875	-0.132	**0.0003**	0.0082

Continues...

Table 6.4 continued

Metabolites	Controls (n=764) mean total PSA 1.21ng/ml			Controls with PSA < 1ng/ml (n=489) mean total PSA 0.55ng/ml			Cases (n=768) mean total PSA 5.59 ng/ml		
	r	p^2	p_{adj}	r	p^2	p_{adj}	r	p^2	p_{adj}
SM C16:0	-0.013	0.726	0.977	-0.010	0.8364	0.9875	-0.073	0.0451	0.3060
SM C16:1	-0.004	0.921	0.977	-0.033	0.4789	0.9875	-0.024	0.5041	0.7952
SM C18:0	0.058	0.116	0.977	0.009	0.8490	0.9875	-0.080	0.0287	0.2336
SM C18:1	0.065	0.076	0.977	0.022	0.6389	0.9875	-0.074	0.0426	0.3058
SM C20:2	-0.011	0.774	0.977	-0.061	0.1881	0.9875	-0.039	0.2862	0.7274
SM C24:0	-0.045	0.226	0.977	-0.028	0.5396	0.9875	-0.038	0.3008	0.7274
SM C24:1	-0.005	0.897	0.977	0.047	0.3097	0.9875	-0.051	0.1659	0.6479

[1] Adjusted for age at blood collection (< 55, $55 - 59$, $60 - 64$, $65 - 69$, ≥ 70 years), BMI (fourths, unknown) and study centre

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

Table 6.5: Metabolites significantly associated with covariates in 1,077 control participants from EPIC[1,2,3]

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Study centre	C2	Alanine	ADMA	lysoPC a C16:0	Hexose	SM (OH) C14:1	99	81.1
	C14:1	Arginine	Creatinine	lysoPC a C16:1		SM C18:0		
	C16	Asparagine	Kynurenine	lysoPC a C17:0		SM C20:2		
	C18:1	Citrulline	Taurine	lysoPC a C18:0		SM C24:1		
	C18:2	Glutamate		lysoPC a C18:1				
		Glycine		lysoPC a C18:2				
		Histidine		lysoPC a C20:3				
		Isoleucine		lysoPC a C20:4				
		Leucine		PC aa C28:1				
		Lysine		PC aa C30:0				
		Methionine		PC aa C32:0				
		Ornithine		PC aa C32:1				
		Phenylalanine		PC aa C32:2				
		Proline		PC aa C32:3				
		Serine		PC aa C34:1				
		t4-OH-Pro		PC aa C34:2				
		Threonine		PC aa C34:3				
		Tyrosine		PC aa C34:4				
		Valine		PC aa C36:0				
				PC aa C36:1				
				PC aa C36:2				
				PC aa C36:3				
				PC aa C36:4				
				PC aa C36:5				
				PC aa C36:6				
				PC aa C38:0				
				PC aa C38:3				
				PC aa C38:4				
				PC aa C38:5				
				PC aa C38:6				
				PC aa C40:2				
				PC aa C40:3				
				PC aa C40:4				
				PC aa C40:5				

Continues...

Table 6.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Assay batch	C2	Alanine	ADMA	lysoPC a C16:0		SM (OH) C14:1	98	80.3
	C3	Asparagine	Creatinine	lysoPC a C16:1		SM (OH) C16:1		
	C14:1	Citrulline	Kynurenine	lysoPC a C18:0		SM (OH) C22:1		
	C18:1	Glutamate	Sarcosine	lysoPC a C18:1		SM (OH) C22:2		
	C18:2	Glutamine	SDMA	lysoPC a C20:3		SM (OH) C24:1		
		Glycine	Taurine	lysoPC a C20:4		SM C16:0		
		Histidine		PC aa C32:0		SM C16:1		
		Isoleucine		PC aa C32:3		SM C18:0		
		Lysine		PC aa C34:1		SM C18:1		
		Methionine		PC aa C34:2		SM C20:2		
		Ornithine		PC aa C34:3		SM C24:0		
		Phenylalanine		PC aa C34:4		SM C24:1		
		Serine		PC aa C36:0				
		t4-OH-Pro		PC aa C36:2				
		Threonine		PC aa C36:3				
		Tryptophan		PC aa C36:4				
		Tyrosine		PC aa C38:0				
		Valine		PC aa C38:3				
				PC aa C38:4				
				PC aa C40:2				
				PC aa C40:3				
				PC aa C40:5				
				PC aa C40:6				
				PC aa C42:0				
				PC aa C42:1				
				PC aa C42:2				
				PC aa C42:4				
				PC aa C42:5				
				PC aa C42:6				
				PC ae C30:2				
				PC ae C32:1				
				PC ae C32:2				
				PC ae C34:0				
				PC ae C34:1				

Continues...

Table 6.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Alcohol intake (<10; 10-19; 20-39; ≥40 g/d)	C0 C2 C16 C18:1	Arginine Citrulline Glutamate Glutamine Phenylalanine Serine t4-OH-Pro ↑	Sarcosine ↑	lysoPC a C16:0 ↑ lysoPC a C16:1 ↑ lysoPC a C17:0 ↓ lysoPC a C18:1 ↑ lysoPC a C20:3 lysoPC a C20:4 ↑ PC aa C28:1 PC aa C30:0 ↑ PC aa C32:0 ↑ PC aa C32:1 ↑ PC aa C32:2 ↑ PC aa C32:3 PC aa C34:1 ↑ PC aa C34:2 ↑ PC aa C34:3 ↑ PC aa C34:4 ↑ PC aa C36:0 PC aa C36:1 ↑ PC aa C36:3 PC aa C36:4 ↑ PC aa C36:5 ↑ PC aa C36:6 ↑ PC aa C38:4 ↑ PC aa C38:5 ↑ PC aa C38:6 PC aa C40:4 PC aa C40:5 PC aa C40:6 PC aa C42:0 ↓ PC ae C30:2 PC ae C32:1 ↑ PC ae C34:3 PC ae C36:0 ↑ PC ae C36:2 ↓		SM (OH) C14:1 SM (OH) C16:1 SM (OH) C22:1 SM (OH) C22:2 SM (OH) C24:1 SM C18:0 SM C20:2 ↓ SM C24:0 ↑ SM C24:1 ↑	64	52.5

Continues...

Table 6.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Body mass index (Fourths of the distribution: <24.5; 24.5-26.5; 26.6-28.8; ≥28.9)	C0 C2 ↑ C3 ↑ C16 ↑	Asparagine Glutamate ↑ Glutamine ↓ Glycine ↓ Histidine Isoleucine Leucine Lysine Phenylalanine Serine ↓ Threonine Tryptophan Tyrosine ↑ Valine	Creatinine Kynurenine ↑	lysoPC a C16:0 lysoPC a C17:0 ↓ lysoPC a C18:0 lysoPC a C18:1 ↓ lysoPC a C18:2 ↓ lysoPC a C20:4 ↓ PC aa C32:1 ↑ PC aa C32:2 ↑ PC aa C34:4 ↑ PC aa C36:0 PC aa C38:3 ↑ PC aa C38:4 ↑ PC aa C40:4 ↑ PC aa C40:6 ↑ PC aa C42:0 ↓ PC aa C42:1 ↓ PC aa C42:2 PC ae C34:2 PC ae C34:3 ↓ PC ae C36:1 PC ae C36:2 ↓ PC ae C36:3 ↓ PC ae C38:2 ↓ PC ae C40:1 ↓ PC ae C40:3 ↓ PC ae C40:5 ↓ PC ae C40:6 PC ae C42:2 ↓ PC ae C42:3 ↓ PC ae C42:4 ↓ PC ae C42:5 ↓ PC ae C44:4 ↓ PC ae C44:5 ↓ PC ae C44:6 ↓	Hexose	SM (OH) C14:1 SM (OH) C16:1 SM (OH) C22:2 SM C16:0 ↓ SM C24:1	60	49.2

Continues...

Table 6.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Fasting status (No; yes, i.e. at leats six hours)	C2 ↑ C14:1 ↑	Alanine ↓ Asparagine ↓ Histidine ↓ Isoleucine ↓ Leucine ↓ Lysine ↓ Methionine ↓ Ornithine ↓ Phenylalanine ↓ Proline ↓ Tyrosine ↓ Valine ↓	Sarcosine ↓				15	12.3
Time of day at blood collection (6:00-8:59; 9:00-11:59; 12:00-14:59; 15:00-20:00)	C0 C2	Alanine ↓ Asparagine Histidine Isoleucine Leucine Ornithine Phenylalanine t4-OH-Pro Threonine Tryptophan Valine	Kynurenine Sarcosine SDMA	lysoPC a C16:0 lysoPC a C18:0 lysoPC a C18:1 lysoPC a C18:2 lysoPC a C20:3 lysoPC a C20:4 PC aa C34:2 PC aa C36:1 PC aa C36:2 PC ae C34:3			26	21.3

Continues...

Table 6.5 continued

	Acylcarnitines	Amino acids	Biogenic amines	Glycero-phospholipids	Hexose	Sphingolipids	Total, n	Total, %
Age (<55; 55-59; 60-64; 65-69, ≥70 years)		Citrulline ↑	Kynurenine ↑ SDMA	lysoPC a C18:2 lysoPC a C20:3 ↓ lysoPC a C20:4 PC aa C36:4 ↓ PC aa C38:4 ↓ PC aa C40:4			9	7.4
Smoking status (Never; past; current)		t4-OH-Pro ↑	Kynurenine	lysoPC a C17:0 ↓ PC aa C38:0 ↓ PC ae C30:2 ↓ PC ae C40:6 ↓		SM (OH) C14:1 ↓	7	5.7
Physical activity (Inactive; moderately inactive; moderately active; active)		Glutamate ↓ Proline		lysoPC a C18:0 lysoPC a C18:2 lysoPC a C20:4 ↑ PC aa C36:0 PC aa C38:5 PC aa C40:2 PC ae C34:3 ↑ PC ae C36:5 PC ae C38:6 PC ae C40:1			12	9.8
Marital status							0	0
Educational attainment (Primary; secondary; degree level)				PC ae C30:0 ↑		SM (OH) C14:1 ↑	2	1.6

[1] Differences in geometric mean metabolite concentrations by covariates were tested using ANOVA, after adjustment for age group at blood collection, BMI, study centre and assay batch

[2] Metabolites were included if the association remained significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

[3] Upward arrows indicate that the mean metabolite concentration in each subsequent category of the covariate was higher than in the preceding category. Downward arrows indicate that the mean metabolite concentration in each subsequent category of the covariate was lower in than the preceding category. No directions were provided for study centre and assay batch.

Table 6.6: Risk of prostate cancer per 1 standard deviation increase in metabolite concentrations in 1,077 matched case-control sets in EPIC[1]

Metabolites	1 SD increase			p-trend[2]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
Acylcarnitines					
C0	0.93	0.85	1.02	0.1293	0.7988
C2	0.92	0.84	1.02	0.1030	0.7988
C3	1.05	0.93	1.19	0.4417	0.9321
C14:1	0.91	0.82	1.00	0.0610	0.7491
C16	0.98	0.89	1.08	0.7280	0.9700
C18:1	0.89	0.79	0.99	0.0294	0.7076
C18:2	0.94	0.84	1.06	0.3249	0.9321
Amino acids					
Alanine	0.99	0.90	1.09	0.8311	0.9752
Arginine	1.04	0.92	1.19	0.5182	0.9346
Asparagine	1.02	0.93	1.12	0.6777	0.9700
Citrulline	0.90	0.82	0.99	0.0356	0.7076
Glutamate	1.05	0.92	1.19	0.4809	0.9321
Glutamine	1.00	0.91	1.09	0.9274	0.9816
Glycine	1.02	0.93	1.13	0.6581	0.9700
Histidine	1.00	0.91	1.11	0.9540	0.9816
Isoleucine	1.10	1.00	1.21	0.0526	0.7491
Leucine	1.06	0.96	1.16	0.2373	0.9321
Lysine	0.99	0.90	1.09	0.8641	0.9816
Methionine	1.02	0.92	1.12	0.7271	0.9700
Ornithine	1.02	0.90	1.15	0.7774	0.9752
Phenylalanine	1.07	0.97	1.18	0.1777	0.9321
Proline	1.08	0.99	1.19	0.0924	0.7988
Serine	1.05	0.95	1.16	0.3365	0.9321
t4-OH-Pro	1.11	1.02	1.22	0.0190	0.7076
Threonine	1.04	0.95	1.14	0.3980	0.9321
Tryptophan	1.01	0.92	1.10	0.8944	0.9816
Tyrosine	0.99	0.90	1.09	0.8707	0.9816
Valine	1.05	0.95	1.15	0.3663	0.9321
Biogenic amines					
ADMA	1.05	0.95	1.15	0.3705	0.9321
Creatinine	1.03	0.94	1.13	0.5669	0.9346
Kynurenine	1.05	0.95	1.16	0.3408	0.9321
Sarcosine	1.00	0.92	1.10	0.9181	0.9816
SDMA	0.96	0.87	1.07	0.4782	0.9321
Taurine	0.96	0.83	1.10	0.5517	0.9346
Glycerophospholipids					
lysoPC a C16:0	0.96	0.87	1.07	0.4798	0.9321
lysoPC a C16:1	0.92	0.83	1.01	0.0939	0.7988
lysoPC a C17:0	1.03	0.93	1.13	0.6217	0.9700
lysoPC a C18:0	0.93	0.84	1.04	0.2015	0.9321
lysoPC a C18:1	0.91	0.82	1.00	0.0614	0.7491
lysoPC a C18:2	0.98	0.89	1.09	0.7659	0.9752
lysoPC a C20:3	0.93	0.83	1.03	0.1375	0.7988
lysoPC a C20:4	0.96	0.86	1.06	0.4095	0.9321
PC aa C28:1	1.12	1.01	1.24	0.0270	0.7076
PC aa C30:0	1.04	0.94	1.14	0.4589	0.9321
PC aa C32:0	1.02	0.93	1.12	0.6931	0.9700
PC aa C32:1	0.98	0.89	1.09	0.7315	0.9700
PC aa C32:2	1.01	0.92	1.11	0.8803	0.9816
PC aa C32:3	0.97	0.88	1.08	0.6290	0.9700

Continues...

Table 6.6 continued

Metabolites	1 SD increase			p-trend[2]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
PC aa C34:1	0.97	0.88	1.07	0.4961	0.9321
PC aa C34:2	1.00	0.91	1.10	0.9900	0.9900
PC aa C34:3	0.95	0.85	1.06	0.3690	0.9321
PC aa C34:4	0.99	0.90	1.08	0.7865	0.9752
PC aa C36:0	1.02	0.93	1.12	0.7021	0.9700
PC aa C36:1	0.95	0.87	1.05	0.3128	0.9321
PC aa C36:2	0.94	0.86	1.03	0.2054	0.9321
PC aa C36:3	0.92	0.84	1.01	0.0753	0.7656
PC aa C36:4	0.97	0.88	1.07	0.5780	0.9402
PC aa C36:5	1.04	0.95	1.14	0.4186	0.9321
PC aa C36:6	1.04	0.95	1.14	0.4125	0.9321
PC aa C38:0	1.01	0.91	1.12	0.8393	0.9752
PC aa C38:3	0.92	0.84	1.02	0.1094	0.7988
PC aa C38:4	0.93	0.85	1.02	0.1372	0.7988
PC aa C38:5	0.97	0.88	1.07	0.5146	0.9346
PC aa C38:6	1.02	0.93	1.13	0.6108	0.9678
PC aa C40:2	0.97	0.87	1.08	0.5509	0.9346
PC aa C40:3	0.96	0.87	1.07	0.4823	0.9321
PC aa C40:4	0.95	0.87	1.04	0.2526	0.9321
PC aa C40:5	0.94	0.85	1.04	0.2032	0.9321
PC aa C40:6	0.98	0.89	1.08	0.7159	0.9700
PC aa C42:0	1.06	0.96	1.16	0.2720	0.9321
PC aa C42:1	1.05	0.95	1.15	0.3529	0.9321
PC aa C42:2	1.02	0.93	1.12	0.6950	0.9700
PC aa C42:4	0.93	0.85	1.02	0.1172	0.7988
PC aa C42:5	0.95	0.85	1.06	0.3474	0.9321
PC aa C42:6	0.97	0.88	1.08	0.6024	0.9670
PC ae C30:0	1.16	1.04	1.30	0.0071	0.7076
PC ae C30:2	1.12	1.00	1.25	0.0406	0.7076
PC ae C32:1	1.03	0.94	1.13	0.5632	0.9346
PC ae C32:2	1.02	0.93	1.12	0.6624	0.9700
PC ae C34:0	1.07	0.98	1.18	0.1343	0.7988
PC ae C34:1	1.03	0.94	1.13	0.5587	0.9346
PC ae C34:2	1.05	0.96	1.15	0.3192	0.9321
PC ae C34:3	0.99	0.90	1.08	0.8097	0.9752
PC ae C36:0	1.06	0.97	1.16	0.2324	0.9321
PC ae C36:1	1.04	0.95	1.14	0.4049	0.9321
PC ae C36:2	1.05	0.95	1.16	0.3590	0.9321
PC ae C36:3	1.00	0.92	1.10	0.9309	0.9816
PC ae C36:4	1.01	0.92	1.11	0.8082	0.9752
PC ae C36:5	1.00	0.91	1.10	0.9670	0.9831
PC ae C38:0	1.00	0.90	1.10	0.9575	0.9816
PC ae C38:2	1.01	0.91	1.12	0.8289	0.9752
PC ae C38:3	1.03	0.94	1.13	0.4857	0.9321
PC ae C38:4	1.01	0.92	1.11	0.7756	0.9752
PC ae C38:5	0.99	0.90	1.08	0.7807	0.9752
PC ae C38:6	1.02	0.93	1.13	0.6523	0.9700
PC ae C40:1	0.98	0.89	1.08	0.7102	0.9700
PC ae C40:2	1.05	0.95	1.16	0.3346	0.9321
PC ae C40:3	1.04	0.94	1.14	0.4966	0.9321
PC ae C40:4	1.04	0.95	1.13	0.4225	0.9321
PC ae C40:5	1.03	0.94	1.13	0.5606	0.9346
PC ae C40:6	1.05	0.95	1.15	0.3293	0.9321

Continues...

Table 6.6 continued

Metabolites	1 SD increase			p-trend[2]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
PC ae C42:1	0.97	0.88	1.07	0.5500	0.9346
PC ae C42:2	0.96	0.88	1.06	0.4467	0.9321
PC ae C42:3	0.99	0.90	1.10	0.8564	0.9816
PC ae C42:4	1.04	0.94	1.14	0.4702	0.9321
PC ae C42:5	1.03	0.94	1.13	0.4760	0.9321
PC ae C44:4	1.00	0.91	1.10	0.9479	0.9816
PC ae C44:5	1.01	0.92	1.10	0.8204	0.9752
PC ae C44:6	1.03	0.95	1.13	0.4658	0.9321
Hexose					
Hexose	1.03	0.91	1.16	0.6445	0.9700
Sphingolipids					
SM (OH) C14:1	1.13	1.02	1.25	0.0196	0.7076
SM (OH) C16:1	1.09	0.99	1.20	0.0715	0.7656
SM (OH) C22:1	0.99	0.90	1.09	0.8386	0.9752
SM (OH) C22:2	1.05	0.95	1.15	0.3429	0.9321
SM (OH) C24:1	1.00	0.90	1.10	0.9220	0.9816
SM C16:0	1.04	0.95	1.15	0.3933	0.9321
SM C16:1	1.02	0.93	1.11	0.7396	0.9702
SM C18:0	1.04	0.95	1.15	0.4000	0.9321
SM C18:1	1.00	0.91	1.10	0.9870	0.9900
SM C20:2	1.00	0.88	1.13	0.9519	0.9816
SM C24:0	1.00	0.91	1.10	0.9459	0.9816
SM C24:1	1.04	0.94	1.15	0.4158	0.9321

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

Table 6.7: Departure from linearity of the association between metabolites and prostate cancer risk

Metabolites	Continuous model[1]						Categorical model[2]	
	Linear vs. quadratic		Linear vs. cubic		Quadratic vs. cubic			
	p	p_{adj}	p	p_{adj}	p	p_{adj}	p	p_{adj}
Acylcarnitines								
C0	0.4636	0.8827	0.7605	0.8844	0.9181	0.9573	0.9236	0.9883
C2	0.8414	0.9577	0.7734	0.8844	0.4912	0.8209	0.1840	0.7931
C3	0.2808	0.8357	0.3084	0.8237	0.2753	0.7840	0.2138	0.8152
C14:1	0.7678	0.9343	0.4500	0.8237	0.2192	0.7403	0.1537	0.7931
C16	0.0753	0.5478	0.1575	0.6933	0.4655	0.8209	0.5539	0.9810
C18:1	0.0078	0.4746	0.0274	0.3709	0.7358	0.9034	0.1829	0.7931
C18:2	0.9304	0.9620	0.2266	0.8082	0.0853	0.5779	0.9120	0.9883
Amino acids								
Alanine	0.2254	0.7939	0.3662	0.8237	0.4627	0.8209	0.5121	0.9810
Arginine	0.7969	0.9343	0.5548	0.8237	0.2916	0.7840	0.7775	0.9881
Asparagine	0.0252	0.5115	0.0778	0.5271	0.7583	0.9034	0.7177	0.9810
Citrulline	0.9797	0.9797	0.4427	0.8237	0.2018	0.7403	0.4351	0.9810
Glutamate	0.2002	0.7939	0.4041	0.8237	0.6790	0.8897	0.7441	0.9810
Glutamine	0.2447	0.7939	0.1456	0.6832	0.1138	0.6613	0.7299	0.9810
Glycine	0.3123	0.8494	0.4957	0.8237	0.5361	0.8320	0.3718	0.9810
Histidine	0.3551	0.8804	0.2673	0.8154	0.1817	0.7403	0.0380	0.7247
Isoleucine	0.1257	0.6969	0.0078	0.1911	0.0067	0.2955	0.1885	0.7931
Leucine	0.0419	0.5478	0.0069	0.1911	0.0158	0.2955	0.0564	0.7247
Lysine	0.3650	0.8804	0.3269	0.8237	0.2341	0.7508	0.9222	0.9883
Methionine	0.0509	0.5478	0.0062	0.1911	0.0118	0.2955	0.4038	0.9810
Ornithine	0.0182	0.5115	0.0602	0.4899	0.8307	0.9213	0.5623	0.9810
Phenylalanine	0.0026	0.3160	0.0049	0.1911	0.2092	0.7403	0.0435	0.7247
Proline	0.4687	0.8827	0.6958	0.8760	0.6544	0.8897	0.9058	0.9883
Serine	0.2056	0.7939	0.4489	0.8237	0.9901	0.9901	0.6722	0.9810
t4-OH-Pro	0.6612	0.9096	0.8569	0.9198	0.7326	0.9034	0.0505	0.7247
Threonine	0.8976	0.9595	0.4645	0.8237	0.2180	0.7403	0.4174	0.9810
Tryptophan	0.1014	0.6511	0.2319	0.8082	0.6244	0.8897	0.9721	0.9883
Tyrosine	0.1464	0.7441	0.2599	0.8129	0.4441	0.8203	0.0525	0.7247
Valine	0.3133	0.8494	0.3505	0.8237	0.2986	0.7840	0.2053	0.8078

Continues...

Table 6.7 continued

Metabolites	Continuous model[1]						Categorical model[2]	
	Linear vs. quadratic		Linear vs. cubic		Quadratic vs. cubic			
	p	p_{adj}	p	p_{adj}	p	p_{adj}	p	p_{adj}
Biogenic amines								
ADMA	0.4728	0.8827	0.3244	0.8237	0.1876	0.7403	0.0381	0.7247
Creatinine	0.4747	0.8827	0.5357	0.8237	0.3905	0.8021	0.1820	0.7931
Kynurenine	0.7138	0.9096	0.8595	0.9198	0.6816	0.8897	0.7980	0.9883
Sarcosine	0.5192	0.8948	0.7934	0.8844	0.8275	0.9213	0.8953	0.9883
SDMA	0.3044	0.8494	0.3699	0.8237	0.3337	0.7944	0.3089	0.9810
Taurine	0.5436	0.9085	0.7974	0.8844	0.7721	0.9034	0.9321	0.9883
Glycerophospholipids								
lysoPC a C16:0	0.0424	0.5478	0.0717	0.5145	0.2830	0.7840	0.6729	0.9810
lysoPC a C16:1	0.2327	0.7939	0.4833	0.8237	0.8620	0.9299	0.9060	0.9883
lysoPC a C17:0	0.0551	0.5478	0.1301	0.6832	0.5277	0.8320	0.5546	0.9810
lysoPC a C18:0	0.0123	0.4996	0.0139	0.2420	0.1307	0.6915	0.3028	0.9810
lysoPC a C18:1	0.2923	0.8490	0.4998	0.8237	0.5979	0.8896	0.8540	0.9883
lysoPC a C18:2	0.4032	0.8804	0.6885	0.8760	0.8273	0.9213	0.4496	0.9810
lysoPC a C20:3	0.4041	0.8804	0.5829	0.8465	0.5357	0.8320	0.9552	0.9883
lysoPC a C20:4	0.1227	0.6969	0.2988	0.8237	0.8547	0.9299	0.6408	0.9810
PC aa C28:1	0.6002	0.9096	0.5243	0.8237	0.3133	0.7840	0.4794	0.9810
PC aa C30:0	0.5037	0.8913	0.4670	0.8237	0.2996	0.7840	0.6910	0.9810
PC aa C32:0	0.1137	0.6933	0.2518	0.8129	0.6128	0.8897	0.6558	0.9810
PC aa C32:1	0.2207	0.7939	0.3942	0.8237	0.5474	0.8348	0.3298	0.9810
PC aa C32:2	0.8713	0.9577	0.3249	0.8237	0.1360	0.6915	0.7541	0.9810
PC aa C32:3	0.5702	0.9096	0.4717	0.8237	0.2773	0.7840	0.4805	0.9810
PC aa C34:1	0.0399	0.5478	0.0393	0.4366	0.1335	0.6915	0.3541	0.9810
PC aa C34:2	0.5281	0.8948	0.3461	0.8237	0.1891	0.7403	0.9316	0.9883
PC aa C34:3	0.6733	0.9096	0.1731	0.7039	0.0680	0.5661	0.7375	0.9810
PC aa C34:4	0.4872	0.8871	0.5576	0.8237	0.4076	0.8021	0.1034	0.7931
PC aa C36:0	0.0763	0.5478	0.0966	0.5627	0.2155	0.7403	0.6114	0.9810
PC aa C36:1	0.2730	0.8357	0.5338	0.8237	0.8167	0.9213	0.0725	0.7931
PC aa C36:2	0.9201	0.9595	0.4539	0.8237	0.2103	0.7403	0.1951	0.7935
PC aa C36:3	0.4419	0.8827	0.0119	0.2416	0.0040	0.2955	0.8816	0.9883

Continues...

Table 6.7 continued

Metabolites	Continuous model[1]						Categorical model[2]	
	Linear vs. quadratic		Linear vs. cubic		Quadratic vs. cubic			
	p	p_{adj}	p	p_{adj}	p	p_{adj}	p	p_{adj}
PC aa C36:4	0.8537	0.9577	0.9739	0.9820	0.8907	0.9449	0.0594	0.7247
PC aa C36:5	0.2278	0.7939	0.4816	0.8237	0.9348	0.9665	0.1870	0.7931
PC aa C36:6	0.6258	0.9096	0.4459	0.8237	0.2405	0.7508	0.5608	0.9810
PC aa C38:0	0.0723	0.5478	0.1200	0.6654	0.3149	0.7840	0.5067	0.9810
PC aa C38:3	0.7145	0.9096	0.8952	0.9362	0.7673	0.9034	0.8025	0.9883
PC aa C38:4	0.8590	0.9577	0.8978	0.9362	0.6679	0.8897	0.7409	0.9810
PC aa C38:5	0.6247	0.9096	0.5311	0.8237	0.3110	0.7840	0.6486	0.9810
PC aa C38:6	0.7150	0.9096	0.9346	0.9502	0.9655	0.9816	0.8598	0.9883
PC aa C40:2	0.0558	0.5478	0.0223	0.3405	0.0470	0.4511	0.8927	0.9883
PC aa C40:3	0.1896	0.7939	0.0474	0.4446	0.0364	0.4441	0.9239	0.9883
PC aa C40:4	0.5273	0.8948	0.5410	0.8237	0.3626	0.7944	0.7644	0.9816
PC aa C40:5	0.7023	0.9096	0.7324	0.8760	0.4898	0.8209	0.1060	0.7931
PC aa C40:6	0.7283	0.9146	0.8975	0.9362	0.7571	0.9034	0.6148	0.9810
PC aa C42:0	0.3805	0.8804	0.2787	0.8237	0.1814	0.7403	0.6762	0.9810
PC aa C42:1	0.0247	0.5115	0.0028	0.1911	0.0097	0.2955	0.5501	0.9810
PC aa C42:2	0.6894	0.9096	0.1877	0.7385	0.0742	0.5661	0.0579	0.7247
PC aa C42:4	0.2388	0.7939	0.0708	0.5145	0.0481	0.4511	0.7457	0.9810
PC aa C42:5	0.5894	0.9096	0.0394	0.4366	0.0129	0.2955	0.9693	0.9883
PC aa C42:6	0.1658	0.7778	0.0470	0.4446	0.0406	0.4502	0.7070	0.9810
PC ae C30:0	0.9139	0.9595	0.7185	0.8760	0.4203	0.8021	0.5687	0.9810
PC ae C30:2	0.4540	0.8827	0.5568	0.8237	0.4346	0.8157	0.8573	0.9883
PC ae C32:1	0.6984	0.9096	0.9139	0.9448	0.8623	0.9299	0.8553	0.9883
PC ae C32:2	0.6653	0.9096	0.2560	0.8129	0.1111	0.6613	0.3863	0.9810
PC ae C34:0	0.4645	0.8827	0.3906	0.8237	0.2462	0.7508	0.5604	0.9810
PC ae C34:1	0.1855	0.7939	0.3764	0.8237	0.6536	0.8897	0.8045	0.9883
PC ae C34:2	0.6167	0.9096	0.6865	0.8760	0.4787	0.8209	0.1641	0.7931
PC ae C34:3	0.7994	0.9343	0.7002	0.8760	0.4208	0.8021	0.9593	0.9883
PC ae C36:0	0.9515	0.9697	0.6601	0.8760	0.3632	0.7944	0.3141	0.9810
PC ae C36:1	0.3302	0.8758	0.6178	0.8760	0.9022	0.9489	0.6318	0.9810
PC ae C36:2	0.9013	0.9595	0.2572	0.8129	0.1003	0.6441	0.3755	0.9810

Continues...

Table 6.7 continued

Metabolites	Continuous model[1]						Categorical model[2]	
	Linear vs. quadratic		Linear vs. cubic		Quadratic vs. cubic			
	p	p_{adj}	p	p_{adj}	p	p_{adj}	p	p_{adj}
PC ae C36:3	0.3746	0.8804	0.6479	0.8760	0.7775	0.9034	0.7559	0.9810
PC ae C36:4	0.7347	0.9146	0.4496	0.8237	0.2232	0.7403	0.4456	0.9810
PC ae C36:5	0.9635	0.9715	0.7848	0.8844	0.4873	0.8209	0.9930	0.9981
PC ae C38:0	0.5971	0.9096	0.7890	0.8844	0.6592	0.8897	0.0394	0.7247
PC ae C38:2	0.1347	0.7146	0.2181	0.8062	0.3686	0.7944	0.9085	0.9883
PC ae C38:3	0.7158	0.9096	0.7740	0.8844	0.5377	0.8320	0.9981	0.9981
PC ae C38:4	0.5041	0.8913	0.2864	0.8237	0.1518	0.7403	0.1877	0.7931
PC ae C38:5	0.7114	0.9096	0.9321	0.9502	0.9512	0.9751	0.1646	0.7931
PC ae C38:6	0.4007	0.8804	0.1357	0.6832	0.0697	0.5661	0.6453	0.9810
PC ae C40:1	0.7941	0.9343	0.0969	0.5627	0.0320	0.4331	0.3451	0.9810
PC ae C40:2	0.3983	0.8804	0.6797	0.8760	0.8085	0.9213	0.2607	0.9086
PC ae C40:3	0.6906	0.9096	0.0825	0.5295	0.0279	0.4259	0.1754	0.7931
PC ae C40:4	0.6387	0.9096	0.6348	0.8760	0.4067	0.8021	0.1764	0.7931
PC ae C40:5	0.4775	0.8827	0.7039	0.8760	0.6565	0.8897	0.3534	0.9810
PC ae C40:6	0.0938	0.6358	0.2137	0.8062	0.5975	0.8896	0.6822	0.9810
PC ae C42:1	0.0665	0.5478	0.1710	0.7039	0.6855	0.8897	0.5554	0.9810
PC ae C42:2	0.6675	0.9096	0.8127	0.8932	0.6314	0.8897	0.7163	0.9810
PC ae C42:3	0.4713	0.8827	0.7292	0.8760	0.7372	0.9034	0.1248	0.7931
PC ae C42:4	0.8006	0.9343	0.6179	0.8760	0.3431	0.7944	0.5436	0.9810
PC ae C42:5	0.8693	0.9577	0.7134	0.8760	0.4207	0.8021	0.5423	0.9810
PC ae C44:4	0.4557	0.8827	0.4853	0.8237	0.3456	0.7944	0.0908	0.7931
PC ae C44:5	0.9074	0.9595	0.6659	0.8760	0.3712	0.7944	0.9545	0.9883
PC ae C44:6	0.2473	0.7939	0.4239	0.8237	0.5388	0.8320	0.2606	0.9086
Hexose								
Hexose	0.8041	0.9343	0.0561	0.4886	0.0170	0.2955	0.4541	0.9810
Sphingolipids								
SM (OH) C14:1	0.9095	0.9595	0.9935	0.9935	0.9890	0.9901	0.1420	0.7931
SM (OH) C16:1	0.1584	0.7731	0.3508	0.8237	0.7446	0.9034	0.4931	0.9810
SM (OH) C22:1	0.0602	0.5478	0.1591	0.6933	0.7028	0.9026	0.4723	0.9810
SM (OH) C22:2	0.2784	0.8357	0.5483	0.8237	0.8689	0.9299	0.2876	0.9745
SM (OH) C24:1	0.2157	0.7939	0.4430	0.8237	0.7567	0.9034	0.7432	0.9810

Continues...

Table 6.7 continued

Metabolites	Continuous model[1]						Categorical model[2]	
	Linear vs. quadratic		Linear vs. cubic		Quadratic vs. cubic			
	p	p_{adj}	p	p_{adj}	p	p_{adj}	p	p_{adj}
SM C16:0	0.8529	0.9577	0.7895	0.8844	0.5079	0.8320	0.2304	0.8518
SM C16:1	0.5632	0.9096	0.5604	0.8237	0.3640	0.7944	0.1210	0.7931
SM C18:0	0.7680	0.9343	0.7202	0.8760	0.4505	0.8203	0.4643	0.9810
SM C18:1	0.3622	0.8804	0.1428	0.6832	0.0801	0.5750	0.0402	0.7247
SM C20:2	0.3427	0.8804	0.3049	0.8237	0.2245	0.7403	0.0845	0.7931
SM C24:0	0.9538	0.9697	0.7015	0.8760	0.4008	0.8021	0.4139	0.9810
SM C24:1	0.6995	0.9096	0.8223	0.8957	0.6224	0.8897	0.7026	0.9810

[1] Departure from linearity of the continuous model was tested using the likelihood ratio χ^2 test comparing the linear model (metabolite concentration, continuously) with two non-linear models, one including the quadratic polynomial of metabolite concentration and one including the cubic polynomial

[2] Departure from linearity of the categorical model was tested using the likelihood ratio χ^2 test comparing models with the fifths scored with their medians and as indicator variables, respectively

Table 6.8: Risk of prostate cancer in relation to fifths of metabolite concentrations in 1,077 matched case-control sets in EPIC[1]

Metabolites	Fifths of metabolite concentrations												p-trend[2]	p_{adj}-trend
	1 (Ref)	2			3			4			5			
		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		
			LL	UL		LL	UL		LL	UL		LL	UL	
Acylcarnitines														
C0	Ref	0.95	0.72	1.25	0.89	0.68	1.17	0.95	0.72	1.24	0.92	0.69	1.21	0.5689	0.9025
C2	Ref	1.05	0.80	1.38	0.76	0.57	1.02	0.87	0.65	1.15	0.89	0.67	1.19	0.2617	0.8447
C3	Ref	1.35	0.95	1.93	1.03	0.72	1.49	1.20	0.84	1.71	1.14	0.82	1.59	0.4816	0.9025
C14:1	Ref	1.04	0.77	1.38	0.73	0.54	1.00	0.79	0.59	1.07	0.74	0.55	0.99	0.0348	0.6501
C16	Ref	0.87	0.66	1.14	1.03	0.78	1.36	0.88	0.66	1.19	0.93	0.68	1.27	0.7055	0.9025
C18:1	Ref	0.82	0.63	1.08	0.76	0.58	1.00	0.66	0.50	0.88	0.80	0.58	1.11	0.0554	0.7510
C18:2	Ref	0.93	0.70	1.24	0.95	0.73	1.26	0.93	0.69	1.24	0.79	0.56	1.10	0.2130	0.8447
Amino acids														
Alanine	Ref	0.88	0.67	1.16	0.89	0.67	1.17	0.94	0.71	1.25	1.06	0.79	1.43	0.6852	0.9025
Arginine	Ref	1.00	0.74	1.36	1.00	0.73	1.37	1.14	0.83	1.57	1.02	0.74	1.41	0.6944	0.9025
Asparagine	Ref	1.06	0.81	1.40	0.97	0.73	1.29	0.99	0.74	1.32	1.16	0.86	1.56	0.4153	0.8955
Citrulline	Ref	0.77	0.59	1.01	0.80	0.61	1.05	0.85	0.65	1.13	0.72	0.54	0.97	0.0653	0.7967
Glutamate	Ref	0.92	0.70	1.22	1.03	0.78	1.37	1.10	0.82	1.47	1.02	0.72	1.43	0.6149	0.9025
Glutamine	Ref	1.07	0.82	1.40	1.00	0.76	1.31	0.95	0.72	1.27	1.11	0.84	1.47	0.7081	0.9025
Glycine	Ref	1.08	0.83	1.42	0.87	0.66	1.16	1.05	0.79	1.40	1.12	0.83	1.51	0.5711	0.9025
Histidine	Ref	0.71	0.54	0.94	0.91	0.69	1.20	0.85	0.64	1.13	1.05	0.78	1.41	0.5666	0.9025
Isoleucine	Ref	0.90	0.68	1.19	1.02	0.77	1.36	0.98	0.73	1.31	1.40	1.05	1.87	0.0116	0.5338
Leucine	Ref	0.73	0.55	0.98	0.89	0.67	1.19	0.97	0.73	1.29	1.15	0.86	1.53	0.1328	0.8099
Lysine	Ref	1.07	0.81	1.41	0.98	0.74	1.29	1.02	0.77	1.36	1.05	0.79	1.41	0.8382	0.9647
Methionine	Ref	0.88	0.66	1.17	0.98	0.74	1.31	1.18	0.88	1.57	1.12	0.83	1.51	0.1686	0.8099
Ornithine	Ref	0.87	0.66	1.16	0.84	0.62	1.14	0.95	0.71	1.29	0.98	0.69	1.39	0.8915	0.9749
Phenylalanine	Ref	0.77	0.59	1.02	0.87	0.66	1.16	0.91	0.69	1.20	1.22	0.90	1.64	0.1260	0.8099
Proline	Ref	1.06	0.80	1.40	1.18	0.89	1.57	1.12	0.84	1.49	1.27	0.95	1.70	0.0924	0.8099
Serine	Ref	1.02	0.78	1.34	1.13	0.86	1.49	0.99	0.75	1.31	1.17	0.87	1.57	0.3820	0.8955
t4-OH-Pro	Ref	0.74	0.56	0.97	1.04	0.80	1.37	1.01	0.76	1.33	1.21	0.92	1.59	0.0373	0.6501
Threonine	Ref	0.90	0.68	1.20	1.01	0.76	1.33	1.19	0.89	1.57	1.06	0.79	1.42	0.2621	0.8447
Tryptophan	Ref	0.95	0.72	1.26	1.01	0.76	1.35	1.01	0.76	1.34	1.02	0.76	1.37	0.7798	0.9236
Tyrosine	Ref	0.89	0.68	1.17	0.71	0.53	0.95	1.00	0.75	1.33	0.92	0.69	1.23	0.9384	0.9766
Valine	Ref	0.82	0.62	1.08	0.90	0.68	1.20	0.92	0.69	1.23	1.12	0.84	1.50	0.2756	0.8447

Continues...

Table 6.8 continued

Metabolites	Fifths of metabolite concentrations													
	1 (Ref)	2			3			4			5			p-trend[2]	p_{adj}-trend
		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI			
			LL	UL		LL	UL		LL	UL		LL	UL		
Biogenic amines															
ADMA	Ref	1.27	0.96	1.67	0.86	0.64	1.14	1.11	0.83	1.48	0.99	0.72	1.35	0.7118	0.9025
Creatinine	Ref	0.90	0.69	1.18	0.81	0.61	1.07	1.09	0.83	1.43	0.92	0.69	1.23	0.9446	0.9766
Kynurenine	Ref	1.05	0.79	1.39	1.05	0.79	1.41	1.21	0.91	1.62	1.13	0.82	1.54	0.3003	0.8447
Sarcosine	Ref	1.14	0.87	1.51	1.11	0.85	1.46	1.07	0.80	1.42	1.10	0.83	1.46	0.5027	0.9025
SDMA	Ref	1.04	0.79	1.36	1.04	0.78	1.37	0.81	0.60	1.08	0.96	0.70	1.31	0.4151	0.8955
Taurine	Ref	0.90	0.68	1.19	0.95	0.70	1.30	0.89	0.63	1.25	0.87	0.60	1.27	0.5021	0.9025
Glycerophospholipids															
lysoPC a C16:0	Ref	0.87	0.66	1.15	0.82	0.62	1.08	0.78	0.58	1.06	0.85	0.62	1.17	0.2242	0.8447
lysoPC a C16:1	Ref	0.84	0.64	1.11	0.81	0.61	1.07	0.75	0.56	1.00	0.72	0.54	0.98	0.0270	0.6501
lysoPC a C17:0	Ref	0.89	0.67	1.17	1.03	0.78	1.36	0.93	0.70	1.24	1.08	0.80	1.46	0.6012	0.9025
lysoPC a C18:0	Ref	0.77	0.59	1.00	0.88	0.67	1.16	0.77	0.58	1.03	0.83	0.61	1.13	0.2511	0.8447
lysoPC a C18:1	Ref	0.98	0.75	1.28	0.97	0.73	1.29	0.91	0.68	1.21	0.78	0.57	1.06	0.1135	0.8099
lysoPC a C18:2	Ref	0.85	0.65	1.12	0.99	0.76	1.30	0.83	0.62	1.11	0.93	0.69	1.27	0.6521	0.9025
lysoPC a C20:3	Ref	0.97	0.74	1.28	0.89	0.67	1.18	0.87	0.65	1.17	0.87	0.64	1.20	0.3083	0.8447
lysoPC a C20:4	Ref	0.89	0.68	1.18	0.92	0.69	1.21	0.82	0.61	1.11	0.94	0.68	1.30	0.5782	0.9025
PC aa C28:1	Ref	1.26	0.95	1.68	1.08	0.81	1.45	1.31	0.95	1.79	1.33	0.97	1.82	0.1127	0.8099
PC aa C30:0	Ref	1.22	0.91	1.63	1.16	0.87	1.55	1.31	0.97	1.77	1.27	0.94	1.73	0.1471	0.8099
PC aa C32:0	Ref	1.20	0.91	1.58	1.16	0.88	1.54	1.20	0.90	1.60	1.14	0.85	1.54	0.4457	0.9025
PC aa C32:1	Ref	1.07	0.81	1.42	1.19	0.90	1.58	0.95	0.70	1.28	0.96	0.71	1.30	0.6269	0.9025
PC aa C32:2	Ref	0.89	0.67	1.18	0.98	0.74	1.29	1.02	0.77	1.36	0.95	0.71	1.26	0.9718	0.9848
PC aa C32:3	Ref	0.89	0.67	1.16	1.03	0.78	1.36	0.97	0.73	1.31	0.82	0.59	1.14	0.3927	0.8955
PC aa C34:1	Ref	1.09	0.82	1.44	1.20	0.91	1.59	0.97	0.73	1.29	0.97	0.71	1.31	0.6520	0.9025
PC aa C34:2	Ref	1.00	0.76	1.31	1.05	0.80	1.38	1.01	0.76	1.33	0.95	0.70	1.28	0.7777	0.9236
PC aa C34:3	Ref	1.10	0.82	1.47	1.04	0.77	1.41	0.95	0.69	1.31	0.88	0.62	1.24	0.2939	0.8447
PC aa C34:4	Ref	0.85	0.65	1.13	0.87	0.66	1.16	0.74	0.56	0.98	0.99	0.75	1.31	0.6122	0.9025
PC aa C36:0	Ref	1.02	0.77	1.33	0.89	0.67	1.18	0.91	0.69	1.20	1.03	0.77	1.37	0.9401	0.9766
PC aa C36:1	Ref	0.86	0.65	1.14	0.79	0.59	1.04	1.06	0.80	1.39	0.77	0.58	1.03	0.3185	0.8447
PC aa C36:2	Ref	0.76	0.58	1.00	0.91	0.70	1.20	0.91	0.70	1.20	0.75	0.56	1.00	0.1726	0.8099
PC aa C36:3	Ref	0.92	0.71	1.20	0.93	0.71	1.21	0.81	0.61	1.07	0.72	0.54	0.96	0.0175	0.5338

Continues...

Table 6.8 continued

Metabolites	Fifths of metabolite concentrations												p-trend[2]	p_{adj}-trend
	1 (Ref)	2			3			4			5			
		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		
			LL	UL		LL	UL		LL	UL		LL	UL	
PC aa C36:4	Ref	0.80	0.60	1.06	1.10	0.84	1.44	1.02	0.77	1.35	0.84	0.61	1.15	0.6823	0.9025
PC aa C36:5	Ref	0.89	0.68	1.17	0.88	0.67	1.16	0.90	0.67	1.20	1.18	0.89	1.58	0.1981	0.8447
PC aa C36:6	Ref	0.93	0.70	1.22	0.97	0.74	1.27	0.93	0.70	1.22	1.12	0.85	1.49	0.4640	0.9025
PC aa C38:0	Ref	1.11	0.84	1.46	1.06	0.80	1.41	0.92	0.68	1.25	1.11	0.80	1.53	0.9131	0.9766
PC aa C38:3	Ref	0.87	0.66	1.15	0.95	0.73	1.24	0.82	0.61	1.10	0.79	0.58	1.07	0.1284	0.8099
PC aa C38:4	Ref	0.86	0.65	1.12	0.93	0.71	1.22	0.83	0.62	1.10	0.86	0.64	1.15	0.3128	0.8447
PC aa C38:5	Ref	1.05	0.80	1.38	0.94	0.72	1.24	0.88	0.66	1.17	0.98	0.74	1.31	0.6202	0.9025
PC aa C38:6	Ref	1.12	0.85	1.48	1.06	0.80	1.41	1.03	0.77	1.38	1.10	0.82	1.47	0.6900	0.9025
PC aa C40:2	Ref	0.87	0.66	1.15	0.85	0.64	1.13	0.80	0.59	1.08	0.80	0.58	1.10	0.1634	0.8099
PC aa C40:3	Ref	0.93	0.71	1.23	0.91	0.68	1.21	0.82	0.60	1.12	0.84	0.61	1.17	0.2463	0.8447
PC aa C40:4	Ref	1.10	0.84	1.43	1.02	0.77	1.34	0.93	0.70	1.23	0.91	0.69	1.21	0.3375	0.8761
PC aa C40:5	Ref	1.14	0.86	1.51	0.79	0.59	1.06	0.97	0.72	1.31	0.83	0.61	1.13	0.1379	0.8099
PC aa C40:6	Ref	1.12	0.84	1.48	1.13	0.85	1.50	0.97	0.72	1.30	1.03	0.77	1.38	0.9064	0.9766
PC aa C42:0	Ref	0.98	0.74	1.29	0.89	0.67	1.19	0.93	0.70	1.24	1.05	0.78	1.41	0.8722	0.9749
PC aa C42:1	Ref	0.97	0.73	1.28	1.04	0.80	1.36	0.91	0.69	1.20	1.11	0.83	1.48	0.6204	0.9025
PC aa C42:2	Ref	1.17	0.89	1.55	1.19	0.90	1.57	0.85	0.63	1.15	1.06	0.78	1.44	0.7589	0.9232
PC aa C42:4	Ref	0.84	0.64	1.10	0.81	0.61	1.07	0.75	0.56	0.99	0.78	0.59	1.04	0.0538	0.7510
PC aa C42:5	Ref	1.00	0.76	1.32	0.92	0.68	1.23	0.91	0.66	1.25	0.83	0.59	1.16	0.2355	0.8447
PC aa C42:6	Ref	0.89	0.67	1.17	0.97	0.73	1.28	0.86	0.64	1.16	0.95	0.69	1.30	0.7154	0.9025
PC ae C30:0	Ref	0.95	0.71	1.27	1.20	0.89	1.63	1.20	0.88	1.64	1.47	1.07	2.04	0.0071	0.5338
PC ae C30:2	Ref	1.10	0.83	1.47	1.22	0.90	1.64	1.14	0.84	1.56	1.33	0.95	1.85	0.1112	0.8099
PC ae C32:1	Ref	1.03	0.79	1.35	0.98	0.74	1.30	1.03	0.77	1.36	1.16	0.87	1.55	0.3545	0.8826
PC ae C32:2	Ref	1.21	0.92	1.58	1.01	0.77	1.34	0.99	0.74	1.32	1.09	0.81	1.47	0.9767	0.9848
PC ae C34:0	Ref	1.25	0.94	1.65	1.25	0.94	1.66	1.15	0.86	1.53	1.29	0.96	1.73	0.1716	0.8099
PC ae C34:1	Ref	1.18	0.90	1.55	1.12	0.85	1.47	1.12	0.84	1.49	1.16	0.87	1.57	0.4278	0.8999
PC ae C34:2	Ref	1.06	0.81	1.38	0.89	0.67	1.17	1.25	0.96	1.64	1.15	0.87	1.51	0.2200	0.8447
PC ae C34:3	Ref	1.01	0.76	1.33	1.01	0.77	1.33	0.93	0.70	1.23	0.94	0.70	1.25	0.5267	0.9025
PC ae C36:0	Ref	1.24	0.94	1.62	1.06	0.80	1.40	1.06	0.80	1.40	1.28	0.97	1.69	0.2294	0.8447
PC ae C36:1	Ref	1.19	0.90	1.57	1.24	0.94	1.65	1.14	0.86	1.53	1.18	0.89	1.56	0.3488	0.8826
PC ae C36:2	Ref	1.09	0.82	1.43	1.25	0.95	1.65	1.02	0.76	1.36	1.26	0.92	1.71	0.2380	0.8447

Continues...

Table 6.8 continued

Metabolites	Fifths of metabolite concentrations												p-trend[2]	p_{adj}-trend
	1 (Ref)	2			3			4			5			
		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		
			LL	UL		LL	UL		LL	UL		LL	UL	
PC ae C36:3	Ref	0.95	0.72	1.26	0.94	0.71	1.25	1.10	0.83	1.44	1.04	0.78	1.39	0.5405	0.9025
PC ae C36:4	Ref	1.07	0.82	1.41	1.05	0.80	1.39	1.24	0.93	1.65	1.03	0.76	1.38	0.6315	0.9025
PC ae C36:5	Ref	1.02	0.78	1.34	1.02	0.77	1.36	1.01	0.76	1.35	0.98	0.73	1.33	0.8922	0.9749
PC ae C38:0	Ref	0.81	0.61	1.07	1.04	0.78	1.38	0.75	0.55	1.01	1.00	0.75	1.35	0.9247	0.9766
PC ae C38:2	Ref	0.92	0.69	1.21	0.96	0.72	1.27	0.99	0.74	1.32	0.92	0.67	1.25	0.7292	0.9078
PC ae C38:3	Ref	1.01	0.76	1.33	1.05	0.79	1.40	1.05	0.79	1.40	1.07	0.80	1.43	0.5859	0.9025
PC ae C38:4	Ref	1.21	0.92	1.59	0.97	0.73	1.28	1.26	0.96	1.66	1.12	0.84	1.50	0.4173	0.8955
PC ae C38:5	Ref	1.14	0.88	1.49	0.86	0.65	1.14	1.10	0.83	1.45	1.00	0.74	1.34	0.9592	0.9834
PC ae C38:6	Ref	1.00	0.76	1.30	0.90	0.68	1.19	0.88	0.66	1.17	1.01	0.75	1.36	0.8131	0.9513
PC ae C40:1	Ref	1.08	0.82	1.43	0.97	0.73	1.29	0.82	0.60	1.10	0.98	0.73	1.32	0.4184	0.8955
PC ae C40:2	Ref	1.15	0.87	1.51	1.29	0.98	1.71	1.03	0.77	1.38	1.20	0.89	1.63	0.4032	0.8955
PC ae C40:3	Ref	1.08	0.83	1.42	0.95	0.71	1.28	0.83	0.62	1.13	1.10	0.81	1.49	0.9874	0.9874
PC ae C40:4	Ref	1.09	0.83	1.43	1.06	0.81	1.39	0.92	0.70	1.20	1.27	0.97	1.68	0.2840	0.8447
PC ae C40:5	Ref	1.05	0.80	1.37	0.92	0.70	1.22	0.90	0.68	1.18	1.13	0.84	1.51	0.7374	0.9087
PC ae C40:6	Ref	0.88	0.67	1.16	1.05	0.80	1.37	0.97	0.73	1.28	0.99	0.74	1.33	0.8609	0.9749
PC ae C42:1	Ref	0.81	0.62	1.06	0.93	0.70	1.22	0.80	0.60	1.06	0.80	0.60	1.08	0.1688	0.8099
PC ae C42:2	Ref	0.84	0.64	1.12	0.91	0.69	1.20	0.91	0.69	1.22	0.82	0.61	1.10	0.3033	0.8447
PC ae C42:3	Ref	1.09	0.82	1.45	0.92	0.69	1.23	1.22	0.90	1.65	0.92	0.67	1.26	0.7643	0.9232
PC ae C42:4	Ref	1.16	0.88	1.52	1.06	0.80	1.40	0.99	0.74	1.33	1.16	0.87	1.55	0.5960	0.9025
PC ae C42:5	Ref	1.09	0.83	1.44	0.93	0.70	1.23	1.12	0.84	1.48	1.13	0.86	1.50	0.3992	0.8955
PC ae C44:4	Ref	1.17	0.89	1.53	0.83	0.62	1.10	0.94	0.70	1.25	1.05	0.78	1.41	0.8187	0.9513
PC ae C44:5	Ref	0.98	0.75	1.28	0.97	0.74	1.27	1.04	0.80	1.36	1.07	0.81	1.41	0.5777	0.9025
PC ae C44:6	Ref	1.03	0.79	1.35	0.83	0.63	1.09	1.04	0.79	1.36	1.06	0.81	1.40	0.6770	0.9025
Hexose															
Hexose	Ref	0.97	0.71	1.33	1.28	0.92	1.79	1.29	0.92	1.82	1.30	0.92	1.84	0.0752	0.8099
Sphingolipids															
SM (OH) C14:1	Ref	1.44	1.07	1.94	1.62	1.20	2.20	1.53	1.12	2.11	1.56	1.13	2.15	0.0144	0.5338
SM (OH) C16:1	Ref	0.88	0.67	1.17	1.04	0.78	1.39	1.14	0.85	1.52	1.18	0.88	1.58	0.1129	0.8099
SM (OH) C22:1	Ref	1.03	0.78	1.37	0.85	0.64	1.14	1.02	0.76	1.37	1.05	0.77	1.42	0.8722	0.9749
SM (OH) C22:2	Ref	0.80	0.61	1.06	1.01	0.76	1.32	0.96	0.72	1.28	1.04	0.78	1.39	0.5572	0.9025
SM (OH) C24:1	Ref	0.91	0.69	1.20	1.00	0.75	1.33	1.08	0.81	1.45	1.02	0.75	1.39	0.5749	0.9025

Continues...

Table 6.8 continued

Metabolites	Fifths of metabolite concentrations												p-trend[2]	p_{adj}-trend
	1 (Ref)	2			3			4			5			
		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		
			LL	UL		LL	UL		LL	UL		LL	UL	
SM C16:0	Ref	1.04	0.79	1.38	1.30	0.98	1.71	1.01	0.75	1.36	1.19	0.89	1.60	0.2853	0.8447
SM C16:1	Ref	0.99	0.75	1.30	0.84	0.64	1.11	1.17	0.89	1.54	0.92	0.69	1.23	0.8950	0.9749
SM C18:0	Ref	0.94	0.71	1.24	0.87	0.65	1.15	1.07	0.80	1.43	1.05	0.79	1.42	0.5354	0.9025
SM C18:1	Ref	0.74	0.57	0.97	0.72	0.55	0.95	0.92	0.69	1.21	0.88	0.66	1.16	0.6283	0.9025
SM C20:2	Ref	0.98	0.73	1.32	1.38	1.00	1.92	1.20	0.84	1.71	1.10	0.75	1.61	0.7176	0.9025
SM C24:0	Ref	0.87	0.66	1.15	1.09	0.83	1.43	0.95	0.72	1.26	1.02	0.76	1.37	0.7148	0.9025
SM C24:1	Ref	0.97	0.73	1.29	1.02	0.77	1.37	1.17	0.86	1.59	1.05	0.77	1.43	0.4759	0.9025

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

Table 6.9: Metabolite concentrations and risk of prostate cancer by time to diagnosis in 1,077 matched case-control sets in EPIC[1]

Metabolites	≤5 years (cases, n=428)					>5 years (cases, n=649)					p-het[2]	p_{adj}-het[2]
	1 SD increase			p-trend[2]	p_{adj}-trend[2]	1 SD increase			p-trend[2]	p_{adj}-trend[2]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Acylcarnitines												
C0	0.94	0.81	1.09	0.4186	0.8090	0.92	0.82	1.04	0.1843	0.9964	0.9881	0.9919
C2	0.86	0.75	0.99	0.0376	0.7456	0.99	0.86	1.13	0.8467	0.9964	0.2859	0.9897
C3	1.13	0.98	1.31	0.0947	0.7456	0.88	0.68	1.13	0.3036	0.9964	0.1652	0.9897
C14:1	0.87	0.77	0.99	0.0338	0.7456	0.99	0.82	1.19	0.9014	0.9964	0.3431	0.9897
C16	0.90	0.78	1.03	0.1133	0.7456	1.12	0.96	1.31	0.1651	0.9964	0.0909	0.9897
C18:1	0.83	0.70	0.98	0.0255	0.7456	0.93	0.80	1.08	0.3539	0.9964	0.5164	0.9897
C18:2	0.91	0.78	1.05	0.1935	0.7456	1.00	0.83	1.20	0.9923	0.9988	0.5213	0.9897
Amino acids												
Alanine	1.08	0.92	1.27	0.3592	0.8090	0.93	0.82	1.06	0.2920	0.9964	0.2168	0.9897
Arginine	1.25	1.00	1.56	0.0454	0.7456	0.95	0.81	1.12	0.5531	0.9964	0.1511	0.9897
Asparagine	1.02	0.88	1.19	0.7606	0.9279	1.01	0.89	1.16	0.8364	0.9964	0.9171	0.9897
Citrulline	0.73	0.62	0.86	**0.0002**	0.0244	1.02	0.90	1.16	0.7193	0.9964	0.0009	0.1098
Glutamate	0.95	0.77	1.16	0.5991	0.8701	1.11	0.94	1.30	0.2281	0.9964	0.1758	0.9897
Glutamine	0.92	0.80	1.06	0.2596	0.7725	1.04	0.92	1.17	0.5413	0.9964	0.2472	0.9897
Glycine	1.04	0.90	1.21	0.5855	0.8606	1.00	0.88	1.14	0.9908	0.9988	0.7659	0.9897
Histidine	0.93	0.80	1.09	0.3768	0.8090	1.05	0.92	1.20	0.4414	0.9964	0.3092	0.9897
Isoleucine	1.17	1.01	1.36	0.0334	0.7456	1.04	0.91	1.19	0.5307	0.9964	0.3063	0.9897
Leucine	1.08	0.94	1.24	0.2998	0.7886	1.04	0.91	1.19	0.5452	0.9964	0.8383	0.9897
Lysine	1.03	0.89	1.20	0.6621	0.8842	0.96	0.85	1.09	0.5401	0.9964	0.4484	0.9897
Methionine	1.06	0.92	1.22	0.4154	0.8090	0.97	0.85	1.11	0.6684	0.9964	0.3615	0.9897
Ornithine	0.96	0.79	1.17	0.6929	0.8898	1.05	0.90	1.23	0.5378	0.9964	0.5672	0.9897
Phenylalanine	1.14	0.98	1.33	0.0907	0.7456	1.01	0.88	1.15	0.9075	0.9964	0.3340	0.9897
Proline	1.10	0.95	1.26	0.1976	0.7456	1.07	0.95	1.21	0.2796	0.9964	0.9425	0.9897
Serine	1.03	0.89	1.19	0.6668	0.8842	1.06	0.92	1.21	0.4238	0.9964	0.7965	0.9897
t4-OH-Pro	1.09	0.95	1.26	0.2043	0.7456	1.14	1.01	1.29	0.0346	0.9964	0.5275	0.9897
Threonine	1.03	0.90	1.19	0.6547	0.8842	1.04	0.91	1.18	0.5648	0.9964	0.9611	0.9897
Tryptophan	1.06	0.92	1.23	0.3931	0.8090	0.97	0.85	1.10	0.6070	0.9964	0.3903	0.9897
Tyrosine	0.96	0.83	1.12	0.6278	0.8842	1.02	0.89	1.16	0.8066	0.9964	0.6158	0.9897
Valine	1.05	0.91	1.22	0.4845	0.8131	1.04	0.91	1.18	0.5961	0.9964	0.9654	0.9897

Continues...

Table 6.9 continued

Metabolites	≤5 years (cases, n=428)					>5 years (cases, n=649)					p-het[2]	p_{adj}-het[2]
	1 SD increase			p-trend[2]	p_{adj}-trend[2]	1 SD increase			p-trend[2]	p_{adj}-trend[2]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Biogenic amines												
ADMA	1.03	0.90	1.19	0.6594	0.8842	1.04	0.90	1.19	0.5963	0.9964	0.8234	0.9897
Creatinine	0.99	0.86	1.13	0.8378	0.9495	1.05	0.93	1.20	0.4183	0.9964	0.4695	0.9897
Kynurenine	1.08	0.93	1.25	0.3244	0.8090	1.02	0.89	1.16	0.8224	0.9964	0.5179	0.9897
Sarcosine	1.03	0.89	1.19	0.7366	0.9170	0.99	0.88	1.12	0.8939	0.9964	0.7152	0.9897
SDMA	0.93	0.79	1.09	0.3496	0.8090	0.98	0.85	1.13	0.7643	0.9964	0.7203	0.9897
Taurine	0.86	0.69	1.08	0.1926	0.7456	1.03	0.85	1.24	0.7711	0.9964	0.1968	0.9897
Glycerophospholipids												
lysoPC a C16:0	0.89	0.76	1.06	0.1878	0.7456	1.01	0.88	1.17	0.8391	0.9964	0.4555	0.9897
lysoPC a C16:1	0.88	0.75	1.03	0.1047	0.7456	0.95	0.83	1.08	0.4338	0.9964	0.6271	0.9897
lysoPC a C17:0	0.98	0.84	1.16	0.8430	0.9495	1.04	0.91	1.19	0.5459	0.9964	0.6202	0.9897
lysoPC a C18:0	0.87	0.74	1.03	0.1170	0.7456	0.97	0.85	1.11	0.6389	0.9964	0.4513	0.9897
lysoPC a C18:1	0.87	0.73	1.03	0.1084	0.7456	0.92	0.81	1.05	0.2236	0.9964	0.8011	0.9897
lysoPC a C18:2	0.95	0.80	1.11	0.4924	0.8131	1.00	0.88	1.14	0.9669	0.9988	0.7461	0.9897
lysoPC a C20:3	0.90	0.77	1.05	0.1926	0.7456	0.94	0.82	1.08	0.3905	0.9964	0.8187	0.9897
lysoPC a C20:4	0.88	0.74	1.03	0.1156	0.7456	1.02	0.88	1.19	0.7464	0.9964	0.4510	0.9897
PC aa C28:1	1.08	0.92	1.26	0.3549	0.8090	1.15	1.01	1.32	0.0382	0.9964	0.5060	0.9897
PC aa C30:0	1.05	0.90	1.23	0.5332	0.8131	1.04	0.91	1.18	0.5517	0.9964	0.8668	0.9897
PC aa C32:0	1.06	0.90	1.24	0.4965	0.8131	1.01	0.89	1.14	0.8783	0.9964	0.5382	0.9897
PC aa C32:1	0.99	0.84	1.17	0.9522	0.9681	0.99	0.87	1.13	0.9362	0.9964	0.9111	0.9897
PC aa C32:2	1.02	0.88	1.18	0.8101	0.9412	1.01	0.89	1.14	0.9155	0.9964	0.8990	0.9897
PC aa C32:3	1.07	0.91	1.26	0.4173	0.8090	0.90	0.78	1.05	0.1751	0.9964	0.2101	0.9897
PC aa C34:1	0.99	0.84	1.17	0.9056	0.9524	0.96	0.85	1.09	0.5706	0.9964	0.6961	0.9897
PC aa C34:2	0.98	0.83	1.15	0.7801	0.9390	1.02	0.90	1.15	0.8143	0.9964	0.7607	0.9897
PC aa C34:3	0.99	0.82	1.18	0.8768	0.9495	0.94	0.81	1.08	0.3670	0.9964	0.6703	0.9897
PC aa C34:4	1.00	0.86	1.15	0.9707	0.9707	0.99	0.87	1.12	0.8861	0.9964	0.7144	0.9897
PC aa C36:0	1.05	0.90	1.23	0.5267	0.8131	0.99	0.88	1.13	0.9368	0.9964	0.3989	0.9897
PC aa C36:1	0.98	0.84	1.14	0.7569	0.9279	0.95	0.84	1.06	0.3577	0.9964	0.8151	0.9897
PC aa C36:2	0.94	0.81	1.09	0.4244	0.8090	0.94	0.83	1.05	0.2796	0.9964	0.9630	0.9897
PC aa C36:3	0.91	0.78	1.06	0.2312	0.7623	0.92	0.82	1.04	0.1728	0.9964	0.8252	0.9897

Continues...

Table 6.9 continued

Metabolites	≤5 years (cases, n=428)					>5 years (cases, n=649)					p-het[2]	p_{adj}-het[2]
	1 SD increase			p-trend[2]	p_{adj}-trend[2]	1 SD increase			p-trend[2]	p_{adj}-trend[2]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
PC aa C36:4	0.93	0.79	1.09	0.3466	0.8090	1.02	0.89	1.17	0.7643	0.9964	0.7094	0.9897
PC aa C36:5	1.06	0.91	1.24	0.4768	0.8131	1.03	0.91	1.17	0.5856	0.9964	0.5218	0.9897
PC aa C36:6	1.09	0.94	1.26	0.2714	0.7839	1.01	0.90	1.14	0.8592	0.9964	0.3065	0.9897
PC aa C38:0	1.09	0.92	1.30	0.3038	0.7886	0.96	0.84	1.09	0.5216	0.9964	0.1664	0.9897
PC aa C38:3	0.93	0.79	1.09	0.3661	0.8090	0.92	0.81	1.05	0.2151	0.9964	0.8760	0.9897
PC aa C38:4	0.89	0.77	1.04	0.1368	0.7456	0.96	0.85	1.09	0.5607	0.9964	0.6415	0.9897
PC aa C38:5	0.99	0.86	1.15	0.9134	0.9524	0.95	0.84	1.09	0.4848	0.9964	0.3957	0.9897
PC aa C38:6	1.08	0.93	1.26	0.3036	0.7886	0.99	0.87	1.12	0.8428	0.9964	0.1973	0.9897
PC aa C40:2	1.06	0.90	1.25	0.4738	0.8131	0.91	0.79	1.04	0.1680	0.9964	0.1587	0.9897
PC aa C40:3	1.04	0.87	1.23	0.6817	0.8898	0.91	0.79	1.04	0.1750	0.9964	0.2081	0.9897
PC aa C40:4	0.94	0.81	1.09	0.3978	0.8090	0.96	0.85	1.09	0.5362	0.9964	0.8453	0.9897
PC aa C40:5	0.96	0.81	1.13	0.6206	0.8842	0.93	0.81	1.06	0.2793	0.9964	0.6606	0.9897
PC aa C40:6	1.04	0.89	1.21	0.6531	0.8842	0.95	0.84	1.07	0.3839	0.9964	0.3415	0.9897
PC aa C42:0	1.15	0.99	1.34	0.0636	0.7456	0.98	0.86	1.11	0.7138	0.9964	0.1362	0.9897
PC aa C42:1	1.24	1.07	1.45	0.0053	0.3233	0.92	0.81	1.04	0.1787	0.9964	0.0029	0.1769
PC aa C42:2	1.12	0.97	1.29	0.1203	0.7456	0.93	0.81	1.06	0.2677	0.9964	0.0492	0.9897
PC aa C42:4	0.94	0.82	1.09	0.4358	0.8131	0.92	0.82	1.04	0.1795	0.9964	0.7737	0.9897
PC aa C42:5	0.99	0.83	1.18	0.9044	0.9524	0.92	0.79	1.06	0.2613	0.9964	0.4281	0.9897
PC aa C42:6	1.00	0.86	1.16	0.9503	0.9681	0.96	0.83	1.10	0.5224	0.9964	0.6359	0.9897
PC ae C30:0	1.17	0.98	1.40	0.0766	0.7456	1.16	1.01	1.34	0.0377	0.9964	0.9430	0.9897
PC ae C30:2	1.11	0.94	1.32	0.2139	0.7456	1.13	0.97	1.30	0.1179	0.9964	0.8356	0.9897
PC ae C32:1	1.08	0.92	1.27	0.3266	0.8090	1.01	0.90	1.14	0.8446	0.9964	0.5638	0.9897
PC ae C32:2	1.11	0.94	1.30	0.2121	0.7456	0.98	0.87	1.11	0.7438	0.9964	0.2302	0.9897
PC ae C34:0	1.11	0.96	1.28	0.1757	0.7456	1.06	0.93	1.20	0.3734	0.9964	0.7239	0.9897
PC ae C34:1	1.07	0.92	1.24	0.3881	0.8090	1.01	0.89	1.14	0.8657	0.9964	0.6861	0.9897
PC ae C34:2	1.05	0.91	1.22	0.4812	0.8131	1.05	0.93	1.18	0.4540	0.9964	0.9919	0.9919
PC ae C34:3	1.02	0.88	1.18	0.7935	0.9390	0.97	0.87	1.10	0.6693	0.9964	0.5497	0.9897
PC ae C36:0	1.13	0.97	1.31	0.1124	0.7456	1.02	0.90	1.15	0.7545	0.9964	0.2052	0.9897
PC ae C36:1	1.05	0.91	1.22	0.4939	0.8131	1.03	0.92	1.17	0.6005	0.9964	0.9551	0.9897
PC ae C36:2	1.05	0.90	1.22	0.5133	0.8131	1.03	0.90	1.18	0.6421	0.9964	0.9737	0.9899

Continues...

Table 6.9 continued

Metabolites	≤5 years (cases, n=428)					>5 years (cases, n=649)					p-het[2]	p_{adj}-het[2]
	1 SD increase			p-trend[2]	p_{adj}-trend[2]	1 SD increase			p-trend[2]	p_{adj}-trend[2]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
PC ae C36:3	1.04	0.90	1.22	0.5774	0.8591	0.98	0.87	1.11	0.7907	0.9964	0.6334	0.9897
PC ae C36:4	0.99	0.85	1.15	0.8795	0.9495	1.04	0.91	1.18	0.5558	0.9964	0.8125	0.9897
PC ae C36:5	0.99	0.84	1.15	0.8506	0.9495	1.03	0.90	1.17	0.7036	0.9964	0.9466	0.9897
PC ae C38:0	1.06	0.91	1.24	0.4675	0.8131	0.95	0.84	1.08	0.4614	0.9964	0.1691	0.9897
PC ae C38:2	1.09	0.93	1.28	0.2832	0.7852	0.95	0.83	1.09	0.4737	0.9964	0.2564	0.9897
PC ae C38:3	1.05	0.91	1.22	0.4792	0.8131	1.02	0.90	1.15	0.7921	0.9964	0.9511	0.9897
PC ae C38:4	1.00	0.87	1.16	0.9618	0.9697	1.03	0.91	1.17	0.6323	0.9964	0.8017	0.9897
PC ae C38:5	0.99	0.85	1.15	0.8653	0.9495	0.99	0.88	1.13	0.9239	0.9964	0.7921	0.9897
PC ae C38:6	1.06	0.90	1.24	0.5006	0.8131	1.00	0.88	1.14	0.9895	0.9988	0.3730	0.9897
PC ae C40:1	1.05	0.89	1.24	0.5528	0.8326	0.94	0.83	1.07	0.3535	0.9964	0.1701	0.9897
PC ae C40:2	1.09	0.94	1.27	0.2531	0.7720	1.02	0.89	1.16	0.8009	0.9964	0.4788	0.9897
PC ae C40:3	1.11	0.95	1.30	0.1925	0.7456	0.98	0.86	1.12	0.7518	0.9964	0.3800	0.9897
PC ae C40:4	1.10	0.96	1.26	0.1821	0.7456	1.00	0.89	1.13	0.9988	0.9988	0.3915	0.9897
PC ae C40:5	1.10	0.95	1.27	0.2002	0.7456	0.98	0.86	1.11	0.7604	0.9964	0.2194	0.9897
PC ae C40:6	1.12	0.97	1.30	0.1243	0.7456	0.99	0.87	1.12	0.8539	0.9964	0.1852	0.9897
PC ae C42:1	0.98	0.84	1.15	0.8178	0.9412	0.97	0.86	1.10	0.6626	0.9964	0.8101	0.9897
PC ae C42:2	1.03	0.88	1.20	0.7333	0.9170	0.92	0.81	1.04	0.1843	0.9964	0.2468	0.9897
PC ae C42:3	1.09	0.93	1.28	0.2763	0.7839	0.93	0.81	1.06	0.2670	0.9964	0.1138	0.9897
PC ae C42:4	1.10	0.95	1.27	0.2111	0.7456	1.00	0.88	1.13	0.9482	0.9964	0.4412	0.9897
PC ae C42:5	1.12	0.97	1.29	0.1114	0.7456	0.97	0.86	1.10	0.6540	0.9964	0.1730	0.9897
PC ae C44:4	1.03	0.88	1.19	0.7270	0.9170	0.98	0.86	1.11	0.7323	0.9964	0.8030	0.9897
PC ae C44:5	1.06	0.92	1.22	0.4017	0.8090	0.98	0.87	1.10	0.7041	0.9964	0.4351	0.9897
PC ae C44:6	1.09	0.94	1.25	0.2521	0.7720	1.00	0.88	1.12	0.9522	0.9964	0.3965	0.9897
Hexose												
Hexose	1.04	0.87	1.24	0.6877	0.8898	1.01	0.85	1.20	0.9115	0.9964	0.7915	0.9897
Sphingolipids												
SM (OH) C14:1	1.11	0.95	1.28	0.1881	0.7456	1.14	0.99	1.31	0.0625	0.9964	0.6249	0.9897
SM (OH) C16:1	1.09	0.94	1.25	0.2515	0.7720	1.09	0.96	1.25	0.1832	0.9964	0.8050	0.9897
SM (OH) C22:1	0.96	0.83	1.10	0.5305	0.8131	1.02	0.89	1.17	0.7959	0.9964	0.4547	0.9897
SM (OH) C22:2	1.04	0.90	1.19	0.6322	0.8842	1.05	0.92	1.20	0.4299	0.9964	0.7648	0.9897
SM (OH) C24:1	0.99	0.86	1.14	0.9007	0.9524	1.00	0.87	1.16	0.9556	0.9964	0.9239	0.9897

Continues...

Table 6.9 continued

Metabolites	≤5 years (cases, n=428)					>5 years (cases, n=649)					p-het[2]	p_{adj}-het[2]
	1 SD increase			p-trend[2]	p_{adj}-trend[2]	1 SD increase			p-trend[2]	p_{adj}-trend[2]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
SM C16:0	1.07	0.93	1.24	0.3562	0.8090	1.02	0.89	1.16	0.7969	0.9964	0.6680	0.9897
SM C16:1	1.02	0.89	1.17	0.7926	0.9390	1.02	0.90	1.15	0.8131	0.9964	0.9184	0.9897
SM C18:0	1.05	0.90	1.22	0.5319	0.8131	1.05	0.92	1.19	0.4707	0.9964	0.9572	0.9897
SM C18:1	1.01	0.87	1.16	0.9438	0.9681	1.00	0.89	1.14	0.9397	0.9964	0.9471	0.9897
SM C20:2	1.02	0.85	1.23	0.8005	0.9390	0.96	0.80	1.16	0.6821	0.9964	0.8759	0.9897
SM C24:0	0.99	0.85	1.15	0.8634	0.9495	1.01	0.89	1.14	0.9147	0.9964	0.8453	0.9897
SM C24:1	1.10	0.94	1.29	0.2240	0.7591	1.00	0.87	1.14	0.9554	0.9964	0.2999	0.9897

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (<10; $10-19$; $20-39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

Table 6.10: Metabolite concentrations and risk of prostate cancer by histological grade in 902 matched case-control sets in EPIC[1]

Metabolites	Low-intermediate grade[2] (cases, n=778)					High grade[2] (cases, n=124)						
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend	p-het[3]	p_{adj}-het
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Acylcarnitines												
C0	0.94	0.84	1.04	0.2297	0.9909	0.70	0.51	0.96	0.0265	0.1793	0.4816	0.8282
C2	0.89	0.79	0.99	0.0384	0.9909	0.93	0.69	1.26	0.6433	0.7404	0.5552	0.8406
C3	1.00	0.87	1.15	0.9824	0.9909	1.06	0.64	1.78	0.8156	0.8505	0.4596	0.8282
C14:1	0.89	0.79	1.01	0.0635	0.9909	0.71	0.49	1.01	0.0587	0.2250	0.3688	0.8015
C16	0.94	0.83	1.06	0.2982	0.9909	1.10	0.75	1.63	0.6178	0.7318	0.1790	0.6824
C18:1	0.85	0.74	0.97	0.0138	0.9909	0.82	0.55	1.20	0.3040	0.5151	0.7879	0.9574
C18:2	0.96	0.84	1.10	0.5489	0.9909	0.67	0.41	1.11	0.1219	0.2887	0.2414	0.7012
Amino acids												
Alanine	0.98	0.87	1.10	0.7309	0.9909	1.18	0.84	1.66	0.3330	0.5410	0.2375	0.7012
Arginine	1.03	0.88	1.21	0.7037	0.9909	1.18	0.82	1.71	0.3697	0.5500	0.5843	0.8406
Asparagine	1.00	0.90	1.12	0.9795	0.9909	1.13	0.79	1.61	0.5074	0.6772	0.7791	0.9574
Citrulline	0.91	0.82	1.02	0.1206	0.9909	0.90	0.64	1.28	0.5657	0.7255	0.5959	0.8406
Glutamate	1.07	0.92	1.24	0.3820	0.9909	0.90	0.61	1.34	0.6114	0.7318	0.7759	0.9574
Glutamine	1.03	0.92	1.15	0.5896	0.9909	0.83	0.63	1.11	0.2171	0.4142	0.1579	0.6316
Glycine	1.04	0.93	1.17	0.4800	0.9909	1.00	0.75	1.33	0.9805	0.9886	0.4011	0.8015
Histidine	1.02	0.90	1.14	0.7718	0.9909	0.93	0.68	1.26	0.6246	0.7327	0.6237	0.8406
Isoleucine	1.10	0.98	1.23	0.1237	0.9909	0.96	0.72	1.29	0.8015	0.8431	0.8881	0.9734
Leucine	1.06	0.95	1.19	0.3091	0.9909	0.87	0.65	1.16	0.3450	0.5410	0.5204	0.8406
Lysine	0.98	0.87	1.09	0.6662	0.9909	0.92	0.68	1.25	0.5893	0.7318	0.9654	0.9734
Methionine	1.03	0.91	1.16	0.6751	0.9909	1.08	0.80	1.45	0.6173	0.7318	0.7304	0.9186
Ornithine	0.99	0.86	1.15	0.9345	0.9909	0.99	0.68	1.44	0.9652	0.9813	0.9615	0.9734
Phenylalanine	1.04	0.93	1.17	0.4958	0.9909	1.16	0.83	1.62	0.3739	0.5500	0.6285	0.8406
Proline	1.09	0.97	1.22	0.1403	0.9909	1.00	0.75	1.33	0.9952	0.9952	0.6339	0.8406
Serine	1.00	0.89	1.12	0.9737	0.9909	1.19	0.85	1.65	0.3112	0.5201	0.6610	0.8579
t4-OH-Pro	1.08	0.97	1.20	0.1484	0.9909	1.04	0.79	1.39	0.7684	0.8431	0.8615	0.9734
Threonine	1.01	0.90	1.12	0.9242	0.9909	1.19	0.86	1.64	0.2874	0.5009	0.4067	0.8015
Tryptophan	1.04	0.93	1.16	0.4971	0.9909	0.89	0.66	1.20	0.4475	0.6204	0.7223	0.9179
Tyrosine	0.97	0.87	1.09	0.6430	0.9909	1.05	0.79	1.40	0.7310	0.8107	0.4662	0.8282
Valine	1.03	0.91	1.16	0.6558	0.9909	0.76	0.56	1.04	0.0912	0.2527	0.5796	0.8406

Continues...

Table 6.10 continued

Metabolites	Low-intermediate grade[2] (cases, n=778)					High grade[2] (cases, n=124)					p-het[3]	p_{adj}-het
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Biogenic amines												
ADMA	1.05	0.93	1.17	0.4346	0.9909	1.12	0.81	1.56	0.4846	0.6569	0.9164	0.9734
Creatinine	1.07	0.96	1.19	0.2084	0.9909	0.97	0.72	1.31	0.8313	0.8595	0.4485	0.8282
Kynurenine	1.06	0.94	1.19	0.3332	0.9909	1.05	0.76	1.44	0.7883	0.8431	0.9458	0.9734
Sarcosine	0.97	0.87	1.08	0.5732	0.9909	1.06	0.78	1.44	0.7064	0.7928	0.3701	0.8015
SDMA	0.97	0.86	1.10	0.6277	0.9909	1.08	0.77	1.52	0.6416	0.7404	0.5962	0.8406
Taurine	0.95	0.80	1.12	0.5078	0.9909	0.85	0.49	1.48	0.5655	0.7255	0.6273	0.8406
Glycerophospholipids												
lysoPC a C16:0	0.95	0.84	1.08	0.4762	0.9909	0.76	0.51	1.13	0.1759	0.3577	0.4650	0.8282
lysoPC a C16:1	0.94	0.84	1.05	0.2800	0.9909	0.66	0.45	0.98	0.0405	0.1929	0.4073	0.8015
lysoPC a C17:0	1.01	0.90	1.14	0.8328	0.9909	0.87	0.62	1.24	0.4418	0.6204	0.2783	0.7312
lysoPC a C18:0	0.92	0.82	1.05	0.2222	0.9909	0.71	0.48	1.04	0.0806	0.2527	0.2771	0.7312
lysoPC a C18:1	0.89	0.79	1.01	0.0673	0.9909	0.84	0.59	1.20	0.3459	0.5410	0.9524	0.9734
lysoPC a C18:2	0.95	0.85	1.08	0.4483	0.9909	1.04	0.74	1.47	0.8016	0.8431	0.9019	0.9734
lysoPC a C20:3	0.91	0.81	1.03	0.1356	0.9909	0.83	0.57	1.22	0.3528	0.5448	0.8827	0.9734
lysoPC a C20:4	0.95	0.84	1.08	0.4252	0.9909	0.86	0.59	1.25	0.4287	0.6153	0.5186	0.8406
PC aa C28:1	1.06	0.94	1.19	0.3605	0.9909	0.95	0.66	1.38	0.8014	0.8431	0.8517	0.9734
PC aa C30:0	1.03	0.92	1.16	0.5816	0.9909	0.80	0.56	1.13	0.2019	0.4038	0.6714	0.8622
PC aa C32:0	0.99	0.89	1.11	0.9167	0.9909	0.78	0.56	1.08	0.1323	0.2989	0.6106	0.8406
PC aa C32:1	1.01	0.89	1.13	0.9249	0.9909	0.78	0.54	1.11	0.1692	0.3499	0.9181	0.9734
PC aa C32:2	1.04	0.94	1.16	0.4402	0.9909	0.59	0.40	0.87	0.0073	0.1098	0.1386	0.6294
PC aa C32:3	0.99	0.87	1.13	0.9239	0.9909	0.52	0.35	0.79	0.0020	0.1057	0.0749	0.6120
PC aa C34:1	0.95	0.85	1.07	0.4241	0.9909	0.86	0.62	1.20	0.3662	0.5500	0.6270	0.8406
PC aa C34:2	0.97	0.86	1.08	0.5583	0.9909	0.79	0.54	1.15	0.2173	0.4142	0.9519	0.9734
PC aa C34:3	0.99	0.87	1.12	0.8460	0.9909	0.56	0.36	0.86	0.0079	0.1098	0.2261	0.7012
PC aa C34:4	1.02	0.92	1.14	0.6732	0.9909	0.56	0.39	0.82	0.0026	0.1057	0.0604	0.6120
PC aa C36:0	1.00	0.89	1.12	0.9638	0.9909	0.82	0.62	1.08	0.1571	0.3362	0.3041	0.7664
PC aa C36:1	0.95	0.85	1.06	0.3719	0.9909	0.81	0.60	1.09	0.1685	0.3499	0.8238	0.9734
PC aa C36:2	0.91	0.82	1.02	0.0945	0.9909	0.74	0.52	1.05	0.0932	0.2527	0.8677	0.9734
PC aa C36:3	0.90	0.81	1.00	0.0582	0.9909	0.70	0.50	0.99	0.0411	0.1929	0.9539	0.9734

Continues...

Table 6.10 continued

Metabolites	Low-intermediate grade[2] (cases, n=778)					High grade[2] (cases, n=124)					p-het[3]	p_{adj}-het
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
PC aa C36:4	0.97	0.86	1.09	0.5679	0.9909	0.65	0.43	0.96	0.0316	0.1793	0.2159	0.7012
PC aa C36:5	1.04	0.93	1.17	0.5116	0.9909	0.85	0.63	1.15	0.3006	0.5151	0.4680	0.8282
PC aa C36:6	1.06	0.95	1.18	0.3363	0.9909	0.75	0.56	1.01	0.0589	0.2250	0.1390	0.6294
PC aa C38:0	1.00	0.88	1.12	0.9375	0.9909	0.77	0.56	1.06	0.1120	0.2847	0.2349	0.7012
PC aa C38:3	0.92	0.82	1.04	0.1810	0.9909	0.60	0.41	0.89	0.0114	0.1264	0.5628	0.8406
PC aa C38:4	0.94	0.84	1.05	0.2716	0.9909	0.59	0.40	0.87	0.0071	0.1098	0.1410	0.6294
PC aa C38:5	0.98	0.87	1.10	0.7022	0.9909	0.62	0.44	0.87	0.0063	0.1098	0.0903	0.6120
PC aa C38:6	1.01	0.90	1.13	0.8992	0.9909	0.85	0.64	1.14	0.2872	0.5009	0.5056	0.8406
PC aa C40:2	1.00	0.88	1.13	0.9828	0.9909	0.77	0.55	1.08	0.1254	0.2887	0.2536	0.7195
PC aa C40:3	0.97	0.85	1.10	0.6309	0.9909	0.71	0.50	1.02	0.0603	0.2250	0.2662	0.7312
PC aa C40:4	0.97	0.87	1.07	0.5167	0.9909	0.61	0.43	0.87	0.0065	0.1098	0.1355	0.6294
PC aa C40:5	0.98	0.87	1.11	0.8074	0.9909	0.57	0.40	0.82	0.0026	0.1057	0.0483	0.5905
PC aa C40:6	0.98	0.88	1.10	0.7544	0.9909	0.79	0.58	1.06	0.1196	0.2887	0.4820	0.8282
PC aa C42:0	1.06	0.94	1.19	0.3152	0.9909	0.83	0.62	1.12	0.2327	0.4368	0.0812	0.6120
PC aa C42:1	1.06	0.94	1.19	0.3388	0.9909	0.79	0.58	1.07	0.1231	0.2887	0.0726	0.6120
PC aa C42:2	0.97	0.87	1.09	0.6452	0.9909	0.92	0.69	1.22	0.5530	0.7254	0.8438	0.9734
PC aa C42:4	0.95	0.86	1.06	0.3530	0.9909	0.70	0.51	0.95	0.0222	0.1793	0.1058	0.6294
PC aa C42:5	1.00	0.87	1.14	0.9733	0.9909	0.66	0.45	0.97	0.0338	0.1793	0.1496	0.6294
PC aa C42:6	0.99	0.88	1.11	0.8299	0.9909	0.59	0.38	0.91	0.0176	0.1630	0.0787	0.6120
PC ae C30:0	1.16	1.02	1.32	0.0203	0.9909	0.91	0.61	1.36	0.6531	0.7447	0.4066	0.8015
PC ae C30:2	1.10	0.96	1.25	0.1653	0.9909	0.90	0.60	1.35	0.6132	0.7318	0.5245	0.8406
PC ae C32:1	1.04	0.93	1.16	0.5347	0.9909	0.76	0.55	1.05	0.0930	0.2527	0.3204	0.7664
PC ae C32:2	1.02	0.91	1.14	0.7296	0.9909	0.76	0.57	1.02	0.0692	0.2412	0.2002	0.7012
PC ae C34:0	1.03	0.92	1.15	0.6016	0.9909	0.92	0.67	1.26	0.5914	0.7318	0.7926	0.9574
PC ae C34:1	1.02	0.91	1.15	0.6861	0.9909	0.81	0.59	1.12	0.2112	0.4142	0.5933	0.8406
PC ae C34:2	1.04	0.94	1.16	0.4394	0.9909	0.73	0.53	1.01	0.0562	0.2250	0.1007	0.6294
PC ae C34:3	1.01	0.90	1.12	0.9044	0.9909	0.66	0.49	0.90	0.0083	0.1098	0.0169	0.5704
PC ae C36:0	1.00	0.90	1.12	0.9687	0.9909	0.96	0.72	1.28	0.7885	0.8431	0.9422	0.9734
PC ae C36:1	1.02	0.92	1.14	0.7032	0.9909	0.86	0.63	1.17	0.3305	0.5410	0.6464	0.8480
PC ae C36:2	1.01	0.90	1.14	0.8694	0.9909	0.86	0.61	1.22	0.3998	0.5807	0.5370	0.8406

Continues...

Table 6.10 continued

Metabolites	Low-intermediate grade[2] (cases, n=778)					High grade[2] (cases, n=124)						
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend	p-het[3]	p_{adj}-het
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
PC ae C36:3	1.01	0.90	1.13	0.9107	0.9909	0.71	0.52	0.96	0.0262	0.1793	0.1211	0.6294
PC ae C36:4	1.00	0.89	1.11	0.9754	0.9909	0.69	0.49	0.99	0.0434	0.1961	0.1605	0.6316
PC ae C36:5	1.00	0.89	1.12	0.9633	0.9909	0.69	0.50	0.96	0.0260	0.1793	0.0736	0.6120
PC ae C38:0	0.99	0.88	1.11	0.8303	0.9909	0.76	0.57	1.03	0.0777	0.2527	0.2375	0.7012
PC ae C38:2	1.01	0.89	1.14	0.8964	0.9909	0.74	0.51	1.06	0.0955	0.2533	0.2285	0.7012
PC ae C38:3	1.01	0.91	1.13	0.8262	0.9909	0.74	0.52	1.05	0.0911	0.2527	0.3141	0.7664
PC ae C38:4	1.00	0.90	1.11	0.9750	0.9909	0.72	0.51	1.03	0.0756	0.2527	0.1317	0.6294
PC ae C38:5	0.99	0.89	1.10	0.8509	0.9909	0.63	0.45	0.89	0.0090	0.1098	0.0484	0.5905
PC ae C38:6	1.02	0.90	1.14	0.7961	0.9909	0.70	0.51	0.97	0.0295	0.1793	0.0884	0.6120
PC ae C40:1	0.96	0.85	1.08	0.4731	0.9909	0.79	0.57	1.09	0.1478	0.3278	0.3379	0.7778
PC ae C40:2	0.99	0.88	1.11	0.8611	0.9909	0.89	0.65	1.23	0.4834	0.6569	0.5406	0.8406
PC ae C40:3	1.01	0.90	1.14	0.8474	0.9909	0.82	0.59	1.14	0.2383	0.4405	0.3172	0.7664
PC ae C40:4	1.04	0.94	1.16	0.4276	0.9909	0.74	0.53	1.02	0.0627	0.2250	0.0482	0.5905
PC ae C40:5	1.04	0.93	1.16	0.5195	0.9909	0.66	0.47	0.92	0.0146	0.1484	0.0187	0.5704
PC ae C40:6	1.03	0.92	1.16	0.5526	0.9909	0.80	0.59	1.09	0.1565	0.3362	0.1286	0.6294
PC ae C42:1	0.96	0.86	1.08	0.5231	0.9909	0.90	0.67	1.19	0.4442	0.6204	0.8188	0.9734
PC ae C42:2	0.94	0.83	1.05	0.2698	0.9909	0.79	0.59	1.06	0.1163	0.2887	0.4488	0.8282
PC ae C42:3	0.97	0.86	1.09	0.5828	0.9909	0.83	0.59	1.17	0.2798	0.5009	0.3650	0.8015
PC ae C42:4	1.06	0.95	1.19	0.2714	0.9909	0.76	0.56	1.04	0.0824	0.2527	0.0289	0.5876
PC ae C42:5	1.07	0.96	1.19	0.2472	0.9909	0.76	0.56	1.04	0.0852	0.2527	0.0238	0.5807
PC ae C44:4	0.99	0.89	1.11	0.9083	0.9909	0.73	0.50	1.07	0.1107	0.2847	0.2130	0.7012
PC ae C44:5	1.05	0.94	1.16	0.4096	0.9909	0.73	0.53	1.00	0.0485	0.2113	0.0374	0.5905
PC ae C44:6	1.07	0.96	1.19	0.1982	0.9909	0.74	0.56	0.98	0.0330	0.1793	0.0066	0.5704
Hexose												
Hexose	0.92	0.79	1.07	0.2652	0.9909	1.48	0.98	2.23	0.0610	0.2250	0.0126	0.5704
Sphingolipids												
SM (OH) C14:1	1.06	0.94	1.19	0.3474	0.9909	0.97	0.67	1.39	0.8485	0.8699	0.4974	0.8406
SM (OH) C16:1	1.04	0.93	1.16	0.4609	0.9909	0.94	0.66	1.32	0.7083	0.7928	0.3889	0.8015
SM (OH) C22:1	0.92	0.82	1.03	0.1539	0.9909	0.68	0.47	0.98	0.0379	0.1927	0.3791	0.8015
SM (OH) C22:2	0.98	0.87	1.09	0.6873	0.9909	0.85	0.62	1.18	0.3394	0.5410	0.6024	0.8406
SM (OH) C24:1	0.93	0.82	1.04	0.2128	0.9909	0.67	0.46	0.96	0.0314	0.1793	0.3309	0.7763

Continues...

Table 6.10 continued

Metabolites	Low-intermediate grade[2] (cases, n=778)					High grade[2] (cases, n=124)					p-het[3]	p_{adj}-het
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
SM C16:0	0.99	0.88	1.11	0.8299	0.9909	0.91	0.68	1.21	0.5107	0.6772	0.4812	0.8282
SM C16:1	0.96	0.86	1.07	0.5020	0.9909	0.86	0.62	1.19	0.3742	0.5500	0.9576	0.9734
SM C18:0	0.99	0.89	1.11	0.9081	0.9909	0.91	0.65	1.28	0.5992	0.7318	0.9995	0.9995
SM C18:1	0.95	0.86	1.06	0.3999	0.9909	0.91	0.66	1.26	0.5709	0.7255	0.8561	0.9734
SM C20:2	1.02	0.87	1.19	0.8016	0.9909	0.78	0.52	1.19	0.2544	0.4632	0.2817	0.7312
SM C24:0	0.96	0.86	1.08	0.5068	0.9909	0.64	0.44	0.93	0.0187	0.1630	0.1967	0.7012
SM C24:1	1.00	0.89	1.13	0.9952	0.9952	0.74	0.52	1.05	0.0910	0.2527	0.1487	0.6294

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (<10; $10-19$; $20-39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Low-intermediate: Gleason score <8 or coded as well, moderately or poorly differentiated tumours. High grade: Gleason score ≥ 8 or coded as undifferentiated tumours

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

Table 6.11: Metabolite concentrations and risk of prostate cancer by stage in 664 matched case-control sets in EPIC[1]

Metabolites	Localised[2] (cases, n=456)					Advanced[2] (cases, n=208)					p-het[3]	p_{adj}-het[3]
	1 SD increase			p-trend[3]	p_{adj}-trend[3]	1 SD increase			p-trend[3]	p_{adj}-trend[3]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Acylcarnitines												
C0	0.95	0.81	1.10	0.4852	0.9356	0.79	0.63	0.98	0.0361	0.1129	0.1314	0.2764
C2	0.90	0.78	1.05	0.1690	0.9356	0.79	0.62	1.00	0.0513	0.1464	0.3794	0.5645
C3	0.83	0.68	1.02	0.0766	0.9356	1.20	0.92	1.57	0.1683	0.3455	0.0306	0.1131
C14:1	0.96	0.81	1.14	0.6556	0.9356	0.77	0.62	0.96	0.0196	0.0841	0.1415	0.2830
C16	0.95	0.80	1.11	0.5086	0.9356	1.00	0.80	1.25	0.9797	0.9797	0.8725	0.9337
C18:1	0.93	0.78	1.11	0.4030	0.9356	0.67	0.49	0.91	0.0114	0.0732	0.0894	0.2226
C18:2	1.07	0.90	1.27	0.4675	0.9356	0.68	0.51	0.91	0.0085	0.0610	0.0098	0.0622
Amino acids												
Alanine	0.97	0.84	1.13	0.6855	0.9356	0.95	0.74	1.23	0.7017	0.8611	0.7247	0.8598
Arginine	1.07	0.84	1.37	0.5839	0.9356	0.88	0.64	1.22	0.4524	0.6936	0.2919	0.4625
Asparagine	0.96	0.83	1.12	0.6239	0.9356	1.01	0.80	1.27	0.9522	0.9756	0.9377	0.9603
Citrulline	0.92	0.79	1.08	0.2986	0.9356	0.75	0.59	0.95	0.0176	0.0841	0.1901	0.3304
Glutamate	0.98	0.80	1.20	0.8518	0.9603	0.97	0.72	1.29	0.8129	0.9003	0.7012	0.8555
Glutamine	1.04	0.90	1.21	0.5912	0.9356	0.83	0.66	1.04	0.1113	0.2576	0.0589	0.1711
Glycine	1.08	0.92	1.27	0.3476	0.9356	0.91	0.71	1.16	0.4548	0.6936	0.0959	0.2294
Histidine	1.08	0.93	1.25	0.3324	0.9356	0.78	0.60	1.02	0.0685	0.1804	0.0752	0.1994
Isoleucine	1.05	0.90	1.22	0.5315	0.9356	1.10	0.87	1.37	0.4312	0.6902	0.7945	0.9059
Leucine	0.99	0.86	1.14	0.9109	0.9710	1.04	0.83	1.30	0.7199	0.8696	0.7653	0.8808
Lysine	0.97	0.83	1.12	0.6397	0.9356	0.97	0.76	1.24	0.8180	0.9003	0.9442	0.9603
Methionine	1.01	0.87	1.18	0.8795	0.9627	1.08	0.85	1.38	0.5290	0.7477	0.7184	0.8598
Ornithine	1.08	0.90	1.31	0.4054	0.9356	0.91	0.67	1.24	0.5576	0.7598	0.1970	0.3304
Phenylalanine	1.03	0.89	1.20	0.6831	0.9356	1.03	0.82	1.29	0.7968	0.9001	0.8547	0.9337
Proline	1.06	0.92	1.22	0.4399	0.9356	1.31	1.03	1.67	0.0264	0.0924	0.2672	0.4346
Serine	1.03	0.88	1.20	0.7220	0.9356	1.02	0.78	1.34	0.8799	0.9335	0.9235	0.9603
t4-OH-Pro	1.07	0.93	1.23	0.3368	0.9356	1.11	0.88	1.40	0.3598	0.5853	0.9524	0.9603
Threonine	1.07	0.93	1.24	0.3475	0.9356	1.05	0.84	1.32	0.6462	0.8333	0.6818	0.8402
Tryptophan	0.99	0.86	1.14	0.8838	0.9627	0.85	0.67	1.09	0.1979	0.3515	0.4060	0.5897
Tyrosine	0.95	0.82	1.11	0.5251	0.9356	1.00	0.82	1.24	0.9686	0.9766	0.5920	0.7561
Valine	0.99	0.85	1.14	0.8580	0.9603	1.09	0.86	1.37	0.4944	0.7179	0.3667	0.5523

Continues...

Table 6.11 continued

Metabolites	Localised[2] (cases, n=456)					Advanced[2] (cases, n=208)					p-het[3]	p_{adj}-het[3]
	1 SD increase			p-trend[3]	p_{adj}-trend[3]	1 SD increase			p-trend[3]	p_{adj}-trend[3]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Biogenic amines												
ADMA	1.03	0.89	1.19	0.7083	0.9356	1.09	0.87	1.38	0.4413	0.6902	0.5261	0.6902
Creatinine	1.18	1.02	1.38	0.0297	0.9356	0.97	0.78	1.22	0.8191	0.9003	0.1914	0.3304
Kynurenine	1.06	0.91	1.25	0.4351	0.9356	1.02	0.79	1.32	0.8609	0.9295	0.9047	0.9515
Sarcosine	1.00	0.87	1.15	0.9651	0.9785	1.02	0.83	1.25	0.8777	0.9335	0.8681	0.9337
SDMA	1.05	0.89	1.24	0.5464	0.9356	0.88	0.68	1.13	0.3180	0.5243	0.4507	0.6109
Taurine	0.94	0.75	1.16	0.5490	0.9356	0.55	0.36	0.85	0.0074	0.0564	0.0381	0.1291
Glycerophospholipids												
lysoPC a C16:0	1.12	0.94	1.34	0.1950	0.9356	0.66	0.49	0.89	0.0062	0.0540	**0.0007**	0.0261
lysoPC a C16:1	0.98	0.84	1.15	0.8111	0.9603	0.75	0.58	0.95	0.0192	0.0841	0.0453	0.1494
lysoPC a C17:0	1.05	0.90	1.22	0.5689	0.9356	0.94	0.72	1.23	0.6489	0.8333	0.4493	0.6109
lysoPC a C18:0	1.06	0.90	1.24	0.5059	0.9356	0.64	0.48	0.86	**0.0028**	0.0380	**0.0034**	0.0461
lysoPC a C18:1	1.02	0.87	1.19	0.8401	0.9603	0.73	0.55	0.96	0.0265	0.0924	0.0188	0.0882
lysoPC a C18:2	0.98	0.84	1.15	0.8299	0.9603	0.99	0.77	1.27	0.9290	0.9756	0.7467	0.8676
lysoPC a C20:3	0.96	0.81	1.13	0.6083	0.9356	0.77	0.59	1.02	0.0645	0.1749	0.1388	0.2822
lysoPC a C20:4	0.99	0.84	1.17	0.8976	0.9691	0.92	0.69	1.22	0.5667	0.7598	0.3278	0.4999
PC aa C28:1	1.10	0.94	1.30	0.2231	0.9356	0.97	0.76	1.23	0.7916	0.9001	0.4897	0.6494
PC aa C30:0	1.05	0.90	1.23	0.5129	0.9356	1.03	0.81	1.33	0.7944	0.9001	0.6559	0.8165
PC aa C32:0	1.06	0.91	1.23	0.4484	0.9356	0.85	0.66	1.08	0.1763	0.3455	0.0666	0.1847
PC aa C32:1	1.04	0.89	1.23	0.5974	0.9356	0.85	0.66	1.09	0.1959	0.3515	0.0856	0.2222
PC aa C32:2	1.03	0.90	1.19	0.6612	0.9356	0.86	0.67	1.10	0.2178	0.3796	0.3112	0.4867
PC aa C32:3	1.04	0.88	1.23	0.6543	0.9356	0.73	0.55	0.96	0.0248	0.0924	0.0660	0.1847
PC aa C34:1	1.09	0.93	1.28	0.2798	0.9356	0.74	0.57	0.97	0.0283	0.0959	0.0052	0.0569
PC aa C34:2	1.03	0.88	1.21	0.6835	0.9356	0.72	0.56	0.93	0.0130	0.0784	0.0220	0.0889
PC aa C34:3	0.96	0.81	1.14	0.6645	0.9356	0.73	0.55	0.96	0.0239	0.0924	0.0955	0.2294
PC aa C34:4	1.03	0.89	1.18	0.7230	0.9356	0.84	0.65	1.08	0.1801	0.3455	0.1905	0.3304
PC aa C36:0	1.05	0.90	1.22	0.5530	0.9356	0.77	0.61	0.99	0.0374	0.1141	0.0374	0.1291
PC aa C36:1	1.08	0.93	1.26	0.2925	0.9356	0.76	0.60	0.95	0.0186	0.0841	0.0075	0.0572
PC aa C36:2	0.98	0.85	1.14	0.8224	0.9603	0.71	0.56	0.90	**0.0043**	0.0437	0.0553	0.1646
PC aa C36:3	0.95	0.82	1.10	0.5080	0.9356	0.65	0.51	0.82	**0.0004**	0.0098	0.0225	0.0889

Continues...

Table 6.11 continued

Metabolites	Localised[2] (cases, n=456)					Advanced[2] (cases, n=208)					p-het[3]	p_{adj}-het[3]
	1 SD increase			p-trend[3]	p_{adj}-trend[3]	1 SD increase			p-trend[3]	p_{adj}-trend[3]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
PC aa C36:4	1.01	0.87	1.17	0.9153	0.9710	0.78	0.59	1.02	0.0695	0.1804	0.0549	0.1646
PC aa C36:5	1.15	0.98	1.33	0.0790	0.9356	0.85	0.67	1.07	0.1566	0.3412	0.0161	0.0818
PC aa C36:6	1.13	0.98	1.30	0.1044	0.9356	0.83	0.67	1.04	0.1041	0.2540	0.0262	0.0999
PC aa C38:0	1.15	0.98	1.36	0.0910	0.9356	0.74	0.57	0.97	0.0303	0.0999	0.0107	0.0622
PC aa C38:3	0.98	0.84	1.14	0.7825	0.9603	0.60	0.46	0.80	**0.0004**	0.0098	0.0182	0.0882
PC aa C38:4	0.97	0.84	1.11	0.6248	0.9356	0.72	0.55	0.93	0.0135	0.0784	0.1022	0.2398
PC aa C38:5	1.06	0.92	1.23	0.4114	0.9356	0.66	0.51	0.85	**0.0013**	0.0198	**0.0013**	0.0261
PC aa C38:6	1.14	0.98	1.33	0.0944	0.9356	0.77	0.61	0.98	0.0317	0.1018	0.0056	0.0569
PC aa C40:2	1.11	0.94	1.32	0.2075	0.9356	0.63	0.48	0.82	**0.0008**	0.0163	**0.0014**	0.0261
PC aa C40:3	1.05	0.89	1.24	0.5869	0.9356	0.54	0.40	0.72	**0.0000**	0.0026	**0.0005**	0.0261
PC aa C40:4	1.00	0.87	1.15	0.9773	0.9785	0.70	0.56	0.89	**0.0037**	0.0410	0.0218	0.0889
PC aa C40:5	1.00	0.86	1.17	0.9785	0.9785	0.63	0.48	0.83	**0.0010**	0.0174	0.0070	0.0569
PC aa C40:6	1.08	0.93	1.26	0.3158	0.9356	0.74	0.58	0.95	0.0170	0.0841	0.0226	0.0889
PC aa C42:0	1.14	0.98	1.33	0.0872	0.9356	0.96	0.75	1.23	0.7417	0.8701	0.1535	0.3020
PC aa C42:1	1.10	0.94	1.28	0.2315	0.9356	0.91	0.72	1.15	0.4356	0.6902	0.1601	0.3052
PC aa C42:2	1.08	0.94	1.26	0.2797	0.9356	0.85	0.68	1.07	0.1719	0.3455	0.1195	0.2621
PC aa C42:4	0.97	0.84	1.12	0.6962	0.9356	0.57	0.44	0.74	**0.0000**	0.0022	**0.0013**	0.0261
PC aa C42:5	1.04	0.87	1.24	0.6613	0.9356	0.58	0.43	0.78	**0.0003**	0.0098	0.0070	0.0569
PC aa C42:6	1.01	0.87	1.17	0.8679	0.9626	0.67	0.51	0.89	0.0054	0.0507	0.0101	0.0622
PC ae C30:0	1.15	0.97	1.37	0.1132	0.9356	1.23	0.92	1.64	0.1542	0.3412	0.9484	0.9603
PC ae C30:2	1.10	0.93	1.31	0.2714	0.9356	0.76	0.57	1.01	0.0588	0.1630	0.0892	0.2226
PC ae C32:1	1.08	0.93	1.25	0.2957	0.9356	0.85	0.68	1.08	0.1819	0.3455	0.0524	0.1639
PC ae C32:2	1.14	0.98	1.33	0.0950	0.9356	0.73	0.57	0.93	0.0097	0.0657	**0.0015**	0.0261
PC ae C34:0	1.03	0.89	1.20	0.6508	0.9356	1.07	0.84	1.36	0.5937	0.7873	0.8201	0.9264
PC ae C34:1	1.07	0.92	1.24	0.3953	0.9356	0.92	0.72	1.17	0.4923	0.7179	0.1977	0.3304
PC ae C34:2	1.04	0.90	1.20	0.5811	0.9356	1.01	0.80	1.27	0.9385	0.9756	0.8565	0.9337
PC ae C34:3	1.00	0.87	1.16	0.9714	0.9785	0.79	0.63	1.00	0.0488	0.1452	0.0470	0.1509
PC ae C36:0	1.12	0.97	1.30	0.1239	0.9356	0.87	0.70	1.08	0.1988	0.3515	0.0197	0.0889
PC ae C36:1	1.05	0.91	1.22	0.4893	0.9356	0.92	0.73	1.16	0.4639	0.6987	0.2723	0.4371
PC ae C36:2	1.02	0.87	1.19	0.8414	0.9603	0.95	0.75	1.21	0.7058	0.8611	0.7422	0.8676

Continues...

Table 6.11 continued

Metabolites	Localised[2] (cases, n=456)					Advanced[2] (cases, n=208)					p-het[3]	p_{adj}-het[3]
	1 SD increase			p-trend[3]	p_{adj}-trend[3]	1 SD increase			p-trend[3]	p_{adj}-trend[3]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
PC ae C36:3	1.03	0.89	1.19	0.7008	0.9356	0.85	0.68	1.07	0.1619	0.3455	0.1946	0.3304
PC ae C36:4	1.00	0.86	1.17	0.9582	0.9785	0.94	0.73	1.20	0.6103	0.8006	0.5941	0.7561
PC ae C36:5	1.03	0.89	1.21	0.6687	0.9356	0.75	0.58	0.97	0.0265	0.0924	0.0150	0.0796
PC ae C38:0	1.11	0.95	1.29	0.1916	0.9356	0.72	0.57	0.92	0.0073	0.0564	**0.0034**	0.0461
PC ae C38:2	1.04	0.88	1.22	0.6544	0.9356	0.72	0.56	0.94	0.0168	0.0841	0.0340	0.1220
PC ae C38:3	1.01	0.88	1.17	0.8505	0.9603	0.85	0.67	1.07	0.1547	0.3412	0.3229	0.4987
PC ae C38:4	0.97	0.84	1.12	0.6870	0.9356	1.01	0.80	1.27	0.9596	0.9756	0.8718	0.9337
PC ae C38:5	1.03	0.88	1.19	0.7330	0.9356	0.84	0.66	1.07	0.1654	0.3455	0.1653	0.3103
PC ae C38:6	1.12	0.96	1.31	0.1574	0.9356	0.73	0.56	0.95	0.0200	0.0841	0.0068	0.0569
PC ae C40:1	1.09	0.94	1.27	0.2652	0.9356	0.68	0.52	0.88	**0.0034**	0.0410	**0.0005**	0.0261
PC ae C40:2	1.02	0.88	1.19	0.7988	0.9603	0.97	0.76	1.24	0.8282	0.9021	0.6260	0.7873
PC ae C40:3	1.03	0.88	1.21	0.6811	0.9356	0.84	0.65	1.08	0.1788	0.3455	0.1288	0.2757
PC ae C40:4	1.02	0.89	1.17	0.7894	0.9603	0.96	0.77	1.20	0.7351	0.8701	0.5950	0.7561
PC ae C40:5	1.08	0.94	1.25	0.2785	0.9356	0.88	0.68	1.13	0.3037	0.5077	0.1154	0.2621
PC ae C40:6	1.09	0.94	1.26	0.2664	0.9356	0.87	0.68	1.12	0.2879	0.4947	0.1383	0.2822
PC ae C42:1	1.00	0.86	1.15	0.9568	0.9785	0.85	0.69	1.04	0.1086	0.2576	0.1814	0.3304
PC ae C42:2	1.02	0.88	1.19	0.7502	0.9436	0.82	0.65	1.03	0.0822	0.2047	0.0733	0.1987
PC ae C42:3	1.10	0.94	1.29	0.2284	0.9356	0.77	0.59	1.00	0.0516	0.1464	0.0045	0.0549
PC ae C42:4	1.05	0.91	1.22	0.4866	0.9356	1.01	0.81	1.26	0.9441	0.9756	0.4004	0.5885
PC ae C42:5	1.11	0.96	1.28	0.1587	0.9356	0.96	0.76	1.22	0.7633	0.8869	0.1916	0.3304
PC ae C44:4	1.05	0.90	1.23	0.5105	0.9356	0.96	0.74	1.24	0.7373	0.8701	0.1591	0.3052
PC ae C44:5	1.10	0.96	1.26	0.1825	0.9356	1.07	0.85	1.35	0.5402	0.7489	0.4462	0.6109
PC ae C44:6	1.09	0.94	1.25	0.2512	0.9356	1.04	0.84	1.31	0.7011	0.8611	0.4215	0.6029
Hexose												
Hexose	0.82	0.65	1.04	0.0972	0.9356	1.33	0.97	1.82	0.0771	0.1960	0.0115	0.0638
Sphingolipids												
SM (OH) C14:1	1.04	0.88	1.22	0.6630	0.9356	1.08	0.84	1.38	0.5658	0.7598	0.9024	0.9515
SM (OH) C16:1	1.00	0.86	1.16	0.9769	0.9785	1.18	0.92	1.52	0.1841	0.3455	0.4250	0.6029
SM (OH) C22:1	0.95	0.81	1.11	0.4911	0.9356	0.95	0.75	1.20	0.6769	0.8602	0.8698	0.9337
SM (OH) C22:2	1.03	0.89	1.20	0.6883	0.9356	0.92	0.72	1.18	0.5332	0.7477	0.4346	0.6094
SM (OH) C24:1	0.94	0.80	1.11	0.4668	0.9356	0.92	0.72	1.17	0.4808	0.7153	0.9950	0.9950

Continues...

Table 6.11 continued

Metabolites	Localised[2] (cases, n=456)					Advanced[2] (cases, n=208)					p-het[3]	p_{adj}-het[3]
	1 SD increase			p-trend[3]	p_{adj}-trend[3]	1 SD increase			p-trend[3]	p_{adj}-trend[3]		
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
SM C16:0	1.07	0.92	1.25	0.4055	0.9356	0.93	0.74	1.16	0.5002	0.7179	0.1190	0.2621
SM C16:1	1.10	0.95	1.28	0.1858	0.9356	0.77	0.62	0.96	0.0228	0.0924	0.0090	0.0622
SM C18:0	1.02	0.87	1.18	0.8271	0.9603	1.14	0.89	1.46	0.3038	0.5077	0.7259	0.8598
SM C18:1	1.03	0.89	1.19	0.6539	0.9356	0.96	0.76	1.21	0.7054	0.8611	0.4623	0.6198
SM C20:2	0.96	0.78	1.19	0.7362	0.9356	0.66	0.47	0.92	0.0151	0.0837	0.2263	0.3731
SM C24:0	1.05	0.90	1.22	0.5746	0.9356	0.82	0.65	1.05	0.1119	0.2576	0.1203	0.2621
SM C24:1	1.17	1.00	1.37	0.0508	0.9356	0.84	0.64	1.09	0.1912	0.3515	0.0107	0.0622

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] The tumour-node-metastasis (TNM) system was used to define localised ($\leq$ T2 and N0/X and M0, or stage coded as localised) and advanced (T3 $-$ 4 and/or N1 $-$ 3 and/or M1, or coded as advanced)

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

Table 6.12: Metabolite concentrations and risk of aggressive prostate cancer in 654 matched case-control sets in EPIC[1]

	Non-aggressive (cases, n=549)					Aggressive[2] (cases, n=115)						
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend	p-het[3]	p_{adj}-het
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Acylcarnitines												
C0	0.89	0.78	1.03	0.1091	0.9970	0.88	0.66	1.18	0.3987	0.8029	0.9896	0.9962
C2	0.85	0.74	0.98	0.0211	0.9970	1.04	0.77	1.42	0.7941	0.9427	0.2934	0.9652
C3	0.86	0.72	1.02	0.0923	0.9970	1.76	1.05	2.94	0.0318	0.4354	0.0018	0.2196
C14:1	0.87	0.75	1.00	0.0547	0.9970	1.01	0.74	1.38	0.9310	0.9837	0.4175	0.9652
C16	0.94	0.81	1.10	0.4402	0.9970	1.03	0.78	1.36	0.8452	0.9548	0.5772	0.9652
C18:1	0.88	0.75	1.04	0.1384	0.9970	0.77	0.52	1.13	0.1771	0.7426	0.5140	0.9652
C18:2	0.98	0.84	1.15	0.7951	0.9970	0.72	0.49	1.05	0.0902	0.5795	0.1163	0.9652
Amino acids												
Alanine	0.97	0.85	1.12	0.6792	0.9970	0.92	0.65	1.29	0.6130	0.9411	0.9051	0.9835
Arginine	1.00	0.81	1.25	0.9843	0.9970	0.89	0.58	1.37	0.5845	0.9411	0.5866	0.9652
Asparagine	0.97	0.85	1.11	0.7058	0.9970	0.98	0.69	1.39	0.9032	0.9837	0.7090	0.9652
Citrulline	0.89	0.77	1.02	0.0994	0.9970	0.87	0.65	1.18	0.3827	0.8029	0.3950	0.9652
Glutamate	0.99	0.83	1.18	0.8707	0.9970	0.96	0.64	1.44	0.8348	0.9541	0.5745	0.9652
Glutamine	0.98	0.86	1.12	0.8134	0.9970	0.84	0.62	1.15	0.2786	0.8029	0.3052	0.9652
Glycine	1.03	0.89	1.19	0.6720	0.9970	0.99	0.71	1.37	0.9315	0.9837	0.2607	0.9652
Histidine	1.01	0.88	1.15	0.9397	0.9970	0.82	0.56	1.21	0.3203	0.8029	0.5091	0.9652
Isoleucine	1.04	0.91	1.19	0.5361	0.9970	1.12	0.81	1.53	0.4930	0.8471	0.5818	0.9652
Leucine	0.99	0.87	1.12	0.8755	0.9970	1.03	0.75	1.40	0.8720	0.9760	0.5887	0.9652
Lysine	0.95	0.83	1.09	0.4805	0.9970	0.99	0.70	1.40	0.9431	0.9837	0.7774	0.9778
Methionine	1.00	0.87	1.15	0.9880	0.9970	1.15	0.82	1.63	0.4220	0.8044	0.5098	0.9652
Ornithine	1.06	0.90	1.26	0.4760	0.9970	0.83	0.53	1.30	0.4146	0.8029	0.1361	0.9652
Phenylalanine	1.02	0.89	1.16	0.7956	0.9970	1.08	0.76	1.55	0.6618	0.9411	0.7225	0.9652
Proline	1.08	0.94	1.23	0.2653	0.9970	1.42	1.01	1.99	0.0451	0.4354	0.6650	0.9652
Serine	1.04	0.90	1.19	0.5910	0.9970	1.01	0.68	1.50	0.9595	0.9837	0.5547	0.9652
t4-OH-Pro	1.05	0.93	1.19	0.4100	0.9970	1.36	0.96	1.93	0.0847	0.5795	0.4169	0.9652
Threonine	1.07	0.94	1.21	0.3106	0.9970	1.00	0.67	1.50	0.9998	0.9998	0.6181	0.9652
Tryptophan	0.96	0.84	1.09	0.5055	0.9970	0.85	0.61	1.18	0.3332	0.8029	0.8107	0.9779
Tyrosine	0.94	0.82	1.07	0.3261	0.9970	1.06	0.79	1.43	0.6910	0.9411	0.3625	0.9652
Valine	0.97	0.85	1.11	0.6640	0.9970	1.15	0.83	1.60	0.4040	0.8029	0.1302	0.9652

Continues...

Table 6.12 continued

	Non-aggressive (cases, n=549)					Aggressive[2] (cases, n=115)						
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend	p-het[3]	p_{adj}-het
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
Biogenic amines												
ADMA	1.06	0.93	1.20	0.4098	0.9970	1.06	0.73	1.54	0.7665	0.9427	0.9962	0.9962
Creatinine	1.11	0.97	1.27	0.1175	0.9970	1.12	0.82	1.53	0.4791	0.8418	0.7393	0.9652
Kynurenine	1.04	0.90	1.20	0.6161	0.9970	1.05	0.75	1.48	0.7591	0.9427	0.9257	0.9835
Sarcosine	0.99	0.87	1.12	0.8779	0.9970	1.06	0.79	1.42	0.6987	0.9411	0.8463	0.9833
SDMA	1.04	0.90	1.21	0.5800	0.9970	0.93	0.67	1.29	0.6641	0.9411	0.4482	0.9652
Taurine	0.86	0.71	1.06	0.1574	0.9970	0.70	0.40	1.23	0.2173	0.7426	0.6444	0.9652
Glycerophospholipids												
lysoPC a C16:0	1.04	0.89	1.21	0.6325	0.9970	0.52	0.31	0.86	0.0114	0.4354	0.0043	0.2623
lysoPC a C16:1	0.93	0.81	1.07	0.3201	0.9970	0.61	0.40	0.91	0.0162	0.4354	0.0951	0.9652
lysoPC a C17:0	1.02	0.89	1.18	0.7478	0.9970	1.01	0.69	1.49	0.9493	0.9837	0.5440	0.9652
lysoPC a C18:0	0.98	0.85	1.14	0.8395	0.9970	0.64	0.42	0.98	0.0378	0.4354	0.0341	0.9652
lysoPC a C18:1	0.96	0.83	1.11	0.5927	0.9970	0.74	0.50	1.10	0.1356	0.7193	0.1021	0.9652
lysoPC a C18:2	0.97	0.84	1.12	0.6743	0.9970	1.07	0.76	1.50	0.7011	0.9411	0.7716	0.9778
lysoPC a C20:3	0.91	0.79	1.06	0.2202	0.9970	0.80	0.54	1.19	0.2744	0.8029	0.4945	0.9652
lysoPC a C20:4	0.97	0.84	1.13	0.7217	0.9970	0.94	0.60	1.45	0.7672	0.9427	0.3543	0.9652
PC aa C28:1	1.04	0.90	1.20	0.6248	0.9970	1.21	0.86	1.71	0.2659	0.8029	0.4734	0.9652
PC aa C30:0	1.00	0.87	1.15	0.9970	0.9970	1.29	0.90	1.85	0.1640	0.7426	0.2019	0.9652
PC aa C32:0	0.98	0.86	1.12	0.7758	0.9970	1.07	0.75	1.51	0.7174	0.9411	0.7306	0.9652
PC aa C32:1	0.99	0.86	1.14	0.8886	0.9970	0.80	0.54	1.20	0.2862	0.8029	0.6134	0.9652
PC aa C32:2	0.97	0.86	1.11	0.6694	0.9970	0.90	0.61	1.32	0.5742	0.9411	0.8336	0.9779
PC aa C32:3	0.95	0.81	1.10	0.4736	0.9970	0.77	0.53	1.13	0.1840	0.7426	0.5980	0.9652
PC aa C34:1	1.01	0.88	1.17	0.8531	0.9970	0.74	0.50	1.11	0.1509	0.7426	0.2414	0.9652
PC aa C34:2	0.95	0.83	1.10	0.5034	0.9970	0.84	0.59	1.19	0.3225	0.8029	0.5873	0.9652
PC aa C34:3	0.89	0.77	1.04	0.1548	0.9970	0.75	0.49	1.13	0.1706	0.7426	0.6904	0.9652
PC aa C34:4	0.96	0.85	1.09	0.5335	0.9970	0.91	0.62	1.34	0.6476	0.9411	0.9047	0.9835
PC aa C36:0	0.96	0.84	1.11	0.5787	0.9970	0.87	0.63	1.19	0.3783	0.8029	0.6584	0.9652
PC aa C36:1	0.98	0.86	1.12	0.7899	0.9970	0.87	0.64	1.18	0.3769	0.8029	0.6211	0.9652
PC aa C36:2	0.91	0.79	1.04	0.1544	0.9970	0.82	0.60	1.12	0.2076	0.7426	0.8091	0.9779
PC aa C36:3	0.87	0.77	1.00	0.0449	0.9970	0.72	0.52	0.99	0.0458	0.4354	0.5917	0.9652

Continues...

Table 6.12 continued

	Non-aggressive (cases, n=549)					Aggressive[2] (cases, n=115)						
	1 SD increase					1 SD increase						
	OR	95% CI		p-trend[3]	p_{adj}-trend	OR	95% CI		p-trend[3]	p_{adj}-trend	p-het[3]	p_{adj}-het
		LL	UL				LL	UL				
PC aa C36:4	0.95	0.83	1.10	0.4924	0.9970	0.87	0.58	1.29	0.4830	0.8418	0.6103	0.9652
PC aa C36:5	1.03	0.90	1.18	0.6436	0.9970	1.04	0.75	1.45	0.8004	0.9427	0.9240	0.9835
PC aa C36:6	1.02	0.90	1.16	0.7501	0.9970	0.94	0.69	1.28	0.6918	0.9411	0.8312	0.9779
PC aa C38:0	1.03	0.89	1.19	0.7166	0.9970	0.91	0.63	1.31	0.6132	0.9411	0.6156	0.9652
PC aa C38:3	0.88	0.76	1.01	0.0738	0.9970	0.71	0.50	1.02	0.0638	0.5189	0.8257	0.9779
PC aa C38:4	0.91	0.80	1.04	0.1748	0.9970	0.80	0.56	1.14	0.2229	0.7426	0.7323	0.9652
PC aa C38:5	0.97	0.85	1.11	0.6283	0.9970	0.74	0.52	1.05	0.0950	0.5795	0.2430	0.9652
PC aa C38:6	1.04	0.90	1.18	0.6131	0.9970	0.84	0.60	1.19	0.3262	0.8029	0.3580	0.9652
PC aa C40:2	0.95	0.82	1.10	0.4735	0.9970	0.81	0.58	1.13	0.2150	0.7426	0.6437	0.9652
PC aa C40:3	0.89	0.76	1.04	0.1363	0.9970	0.73	0.52	1.02	0.0611	0.5189	0.5406	0.9652
PC aa C40:4	0.93	0.82	1.05	0.2466	0.9970	0.76	0.55	1.05	0.0910	0.5795	0.4564	0.9652
PC aa C40:5	0.92	0.80	1.07	0.2814	0.9970	0.68	0.47	0.97	0.0355	0.4354	0.2532	0.9652
PC aa C40:6	0.99	0.86	1.13	0.8541	0.9970	0.80	0.56	1.14	0.2114	0.7426	0.5752	0.9652
PC aa C42:0	1.07	0.93	1.22	0.3624	0.9970	1.18	0.83	1.67	0.3542	0.8029	0.8867	0.9835
PC aa C42:1	1.05	0.91	1.20	0.4996	0.9970	0.98	0.70	1.36	0.8870	0.9836	0.4043	0.9652
PC aa C42:2	0.98	0.86	1.12	0.7897	0.9970	1.06	0.75	1.50	0.7523	0.9427	0.6241	0.9652
PC aa C42:4	0.87	0.76	0.99	0.0386	0.9970	0.66	0.48	0.91	0.0099	0.4354	0.2160	0.9652
PC aa C42:5	0.92	0.79	1.08	0.3065	0.9970	0.65	0.44	0.97	0.0350	0.4354	0.3634	0.9652
PC aa C42:6	0.91	0.79	1.04	0.1465	0.9970	0.84	0.61	1.17	0.3046	0.8029	0.9246	0.9835
PC ae C30:0	1.09	0.93	1.27	0.3048	0.9970	1.71	1.13	2.59	0.0118	0.4354	0.0530	0.9652
PC ae C30:2	1.01	0.86	1.18	0.9341	0.9970	0.87	0.58	1.29	0.4808	0.8418	0.5753	0.9652
PC ae C32:1	1.01	0.89	1.16	0.8525	0.9970	0.94	0.69	1.28	0.7059	0.9411	0.6672	0.9652
PC ae C32:2	1.01	0.88	1.16	0.9100	0.9970	0.86	0.62	1.18	0.3475	0.8029	0.3697	0.9652
PC ae C34:0	1.00	0.87	1.14	0.9675	0.9970	1.33	0.95	1.86	0.1004	0.5833	0.2740	0.9652
PC ae C34:1	1.01	0.88	1.15	0.9064	0.9970	1.08	0.78	1.48	0.6494	0.9411	0.7437	0.9652
PC ae C34:2	1.00	0.87	1.13	0.9493	0.9970	1.23	0.90	1.68	0.1844	0.7426	0.2701	0.9652
PC ae C34:3	0.95	0.83	1.08	0.4092	0.9970	0.88	0.64	1.20	0.4073	0.8029	0.4498	0.9652
PC ae C36:0	1.02	0.89	1.16	0.8102	0.9970	1.05	0.79	1.38	0.7422	0.9427	0.9400	0.9886
PC ae C36:1	1.00	0.87	1.14	0.9736	0.9970	1.09	0.80	1.50	0.5803	0.9411	0.7206	0.9652
PC ae C36:2	0.97	0.84	1.12	0.6742	0.9970	1.15	0.85	1.58	0.3673	0.8029	0.5213	0.9652

Continues...

Table 6.12 continued

	Non-aggressive (cases, n=549)					Aggressive[2] (cases, n=115)						
	1 SD increase					1 SD increase						
	OR	95% CI		p-trend[3]	p_{adj}-trend	OR	95% CI		p-trend[3]	p_{adj}-trend	p-het[3]	p_{adj}-het
		LL	UL				LL	UL				
PC ae C36:3	0.96	0.84	1.10	0.5492	0.9970	0.99	0.75	1.32	0.9538	0.9837	0.8285	0.9779
PC ae C36:4	0.97	0.84	1.11	0.6152	0.9970	1.07	0.78	1.47	0.6731	0.9411	0.5716	0.9652
PC ae C36:5	0.95	0.83	1.10	0.5107	0.9970	0.86	0.62	1.20	0.3828	0.8029	0.5023	0.9652
PC ae C38:0	0.97	0.85	1.12	0.7064	0.9970	0.87	0.63	1.22	0.4292	0.8056	0.6263	0.9652
PC ae C38:2	0.94	0.81	1.08	0.3708	0.9970	0.96	0.70	1.33	0.8205	0.9533	0.9271	0.9835
PC ae C38:3	0.95	0.83	1.09	0.4845	0.9970	0.99	0.74	1.33	0.9516	0.9837	0.7561	0.9710
PC ae C38:4	0.95	0.84	1.09	0.4804	0.9970	1.16	0.84	1.59	0.3684	0.8029	0.4190	0.9652
PC ae C38:5	0.96	0.84	1.10	0.6003	0.9970	0.98	0.71	1.35	0.8949	0.9836	0.9598	0.9888
PC ae C38:6	1.00	0.86	1.15	0.9643	0.9970	0.93	0.66	1.30	0.6552	0.9411	0.7870	0.9779
PC ae C40:1	0.96	0.83	1.10	0.5463	0.9970	0.87	0.62	1.23	0.4380	0.8096	0.2957	0.9652
PC ae C40:2	0.97	0.84	1.12	0.6951	0.9970	1.23	0.89	1.68	0.2094	0.7426	0.3925	0.9652
PC ae C40:3	0.95	0.83	1.10	0.4980	0.9970	1.04	0.75	1.44	0.8036	0.9427	0.8909	0.9835
PC ae C40:4	0.99	0.88	1.12	0.9173	0.9970	1.09	0.80	1.47	0.5976	0.9411	0.9114	0.9835
PC ae C40:5	1.03	0.90	1.17	0.6765	0.9970	1.00	0.71	1.39	0.9775	0.9865	0.6241	0.9652
PC ae C40:6	1.01	0.88	1.16	0.8820	0.9970	1.09	0.75	1.56	0.6598	0.9411	0.9219	0.9835
PC ae C42:1	0.92	0.81	1.05	0.2043	0.9970	1.00	0.76	1.30	0.9784	0.9865	0.6487	0.9652
PC ae C42:2	0.93	0.82	1.07	0.3347	0.9970	1.04	0.78	1.39	0.7964	0.9427	0.7055	0.9652
PC ae C42:3	0.98	0.85	1.13	0.7767	0.9970	1.05	0.74	1.48	0.8002	0.9427	0.6733	0.9652
PC ae C42:4	1.00	0.88	1.14	0.9720	0.9970	1.24	0.91	1.71	0.1782	0.7426	0.6519	0.9652
PC ae C42:5	1.07	0.94	1.21	0.3093	0.9970	1.04	0.74	1.45	0.8368	0.9541	0.4345	0.9652
PC ae C44:4	1.00	0.87	1.15	0.9653	0.9970	1.23	0.83	1.82	0.3106	0.8029	0.9952	0.9962
PC ae C44:5	1.08	0.96	1.23	0.2148	0.9970	1.16	0.83	1.61	0.3904	0.8029	0.6389	0.9652
PC ae C44:6	1.05	0.93	1.19	0.4506	0.9970	1.23	0.88	1.73	0.2252	0.7426	0.9489	0.9888
Hexose												
Hexose	0.87	0.71	1.06	0.1660	0.9970	1.62	1.01	2.60	0.0464	0.4354	0.0209	0.8499
Sphingolipids												
SM (OH) C14:1	1.01	0.87	1.17	0.8897	0.9970	1.32	0.94	1.87	0.1102	0.6111	0.4648	0.9652
SM (OH) C16:1	1.00	0.87	1.15	0.9773	0.9970	1.44	1.01	2.05	0.0437	0.4354	0.2656	0.9652
SM (OH) C22:1	0.89	0.77	1.03	0.1270	0.9970	1.20	0.88	1.64	0.2388	0.7667	0.1103	0.9652
SM (OH) C22:2	0.97	0.85	1.12	0.6930	0.9970	1.13	0.82	1.56	0.4599	0.8374	0.6062	0.9652
SM (OH) C24:1	0.90	0.78	1.05	0.1765	0.9970	1.07	0.78	1.47	0.6610	0.9411	0.3105	0.9652

Continues...

Table 6.12 continued

	Non-aggressive (cases, n=549)					Aggressive[2] (cases, n=115)						
	1 SD increase			p-trend[3]	p_{adj}-trend	1 SD increase			p-trend[3]	p_{adj}-trend	p-het[3]	p_{adj}-het
	OR	95% CI				OR	95% CI					
		LL	UL				LL	UL				
SM C16:0	1.00	0.87	1.15	0.9913	0.9970	1.13	0.84	1.52	0.4096	0.8029	0.9645	0.9888
SM C16:1	1.00	0.88	1.14	0.9731	0.9970	0.86	0.65	1.16	0.3250	0.8029	0.3872	0.9652
SM C18:0	1.00	0.87	1.15	0.9563	0.9970	1.33	0.95	1.86	0.0914	0.5795	0.2153	0.9652
SM C18:1	0.99	0.87	1.13	0.9131	0.9970	1.06	0.77	1.44	0.7335	0.9427	0.7333	0.9652
SM C20:2	0.90	0.74	1.09	0.2635	0.9970	0.62	0.39	0.98	0.0388	0.4354	0.5420	0.9652
SM C24:0	0.94	0.82	1.08	0.3971	0.9970	1.09	0.80	1.47	0.5855	0.9411	0.3260	0.9652
SM C24:1	1.07	0.93	1.24	0.3458	0.9970	1.07	0.74	1.56	0.7131	0.9411	0.6811	0.9652

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (<10; $10-19$; $20-39$; ≥ 40 g of alcohol per day; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] The tumour-node-metastasis (TNM) system was used. Aggressive: T4 and/or N1 $-$ 3 and/or M1

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

Table 6.13: Metabolite concentrations and risk of death from prostate cancer in 127 matched case-control sets in EPIC[1,2]

Metabolites	1 SD increase			p-trend[3]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
Acylcarnitines					
C0	1.16	0.88	1.53	0.2853	0.9578
C2	1.13	0.81	1.58	0.4641	0.9578
C3	1.68	1.14	2.49	0.0093	0.7286
C14:1	1.01	0.75	1.35	0.9682	0.9832
C16	1.29	0.92	1.81	0.1410	0.9578
C18:1	1.19	0.84	1.68	0.3388	0.9578
C18:2	1.09	0.74	1.59	0.6642	0.9578
Amino acids					
Alanine	1.19	0.89	1.59	0.2344	0.9578
Arginine	1.22	0.81	1.83	0.3341	0.9578
Asparagine	1.30	0.94	1.81	0.1129	0.9578
Citrulline	0.82	0.60	1.12	0.2158	0.9578
Glutamate	0.78	0.50	1.21	0.2670	0.9578
Glutamine	1.17	0.87	1.56	0.2905	0.9578
Glycine	1.05	0.76	1.45	0.7619	0.9578
Histidine	1.08	0.79	1.49	0.6135	0.9578
Isoleucine	1.21	0.93	1.58	0.1506	0.9578
Leucine	1.08	0.84	1.39	0.5296	0.9578
Lysine	1.08	0.82	1.43	0.5633	0.9578
Methionine	1.42	1.03	1.96	0.0349	0.7286
Ornithine	1.14	0.78	1.66	0.5015	0.9578
Phenylalanine	1.04	0.77	1.39	0.8131	0.9578
Proline	1.24	0.94	1.63	0.1273	0.9578
Serine	1.34	0.98	1.82	0.0630	0.8538
t4-OH-Pro	1.43	1.07	1.91	0.0166	0.7286
Threonine	1.17	0.87	1.57	0.3019	0.9578
Tryptophan	1.07	0.81	1.41	0.6447	0.9578
Tyrosine	1.13	0.84	1.51	0.4182	0.9578
Valine	1.20	0.91	1.58	0.2046	0.9578
Biogenic amines					
ADMA	1.49	1.04	2.13	0.0303	0.7286
Creatinine	1.06	0.80	1.41	0.6708	0.9578
Kynurenine	1.19	0.87	1.62	0.2834	0.9578
Sarcosine	0.89	0.66	1.19	0.4252	0.9578
SDMA	1.04	0.74	1.47	0.8137	0.9578
Taurine	0.82	0.51	1.34	0.4339	0.9578
Glycerophospholipids					
lysoPC a C16:0	0.75	0.50	1.12	0.1625	0.9578
lysoPC a C16:1	0.82	0.58	1.16	0.2571	0.9578
lysoPC a C17:0	0.95	0.68	1.34	0.7822	0.9578
lysoPC a C18:0	0.84	0.61	1.16	0.2947	0.9578
lysoPC a C18:1	0.78	0.55	1.11	0.1730	0.9578
lysoPC a C18:2	1.08	0.78	1.49	0.6452	0.9578
lysoPC a C20:3	0.84	0.58	1.22	0.3647	0.9578
lysoPC a C20:4	0.71	0.48	1.06	0.0910	0.9578
PC aa C28:1	1.37	1.00	1.90	0.0536	0.8177
PC aa C30:0	1.01	0.74	1.38	0.9479	0.9832
PC aa C32:0	1.08	0.79	1.47	0.6438	0.9578
PC aa C32:1	0.96	0.68	1.36	0.8260	0.9578
PC aa C32:2	0.95	0.68	1.32	0.7382	0.9578
PC aa C32:3	1.04	0.75	1.44	0.8037	0.9578

Continues...

Table 6.13 continued

Metabolites	1 SD increase			p-trend[3]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
PC aa C34:1	0.95	0.69	1.31	0.7501	0.9578
PC aa C34:2	1.24	0.90	1.71	0.1802	0.9578
PC aa C34:3	1.02	0.70	1.48	0.9346	0.9832
PC aa C34:4	0.90	0.65	1.24	0.5222	0.9578
PC aa C36:0	1.07	0.79	1.45	0.6668	0.9578
PC aa C36:1	0.94	0.71	1.25	0.6695	0.9578
PC aa C36:2	1.14	0.86	1.50	0.3614	0.9578
PC aa C36:3	1.03	0.76	1.39	0.8636	0.9578
PC aa C36:4	0.90	0.64	1.27	0.5437	0.9578
PC aa C36:5	0.96	0.71	1.29	0.7693	0.9578
PC aa C36:6	0.96	0.72	1.28	0.7757	0.9578
PC aa C38:0	1.04	0.74	1.46	0.8131	0.9578
PC aa C38:3	1.01	0.71	1.42	0.9763	0.9832
PC aa C38:4	0.87	0.62	1.22	0.4191	0.9578
PC aa C38:5	0.86	0.63	1.17	0.3271	0.9578
PC aa C38:6	1.08	0.80	1.48	0.6074	0.9578
PC aa C40:2	0.80	0.56	1.14	0.2154	0.9578
PC aa C40:3	0.80	0.57	1.12	0.1937	0.9578
PC aa C40:4	0.85	0.61	1.20	0.3558	0.9578
PC aa C40:5	0.83	0.59	1.18	0.3056	0.9578
PC aa C40:6	1.06	0.78	1.45	0.7055	0.9578
PC aa C42:0	1.12	0.80	1.57	0.5119	0.9578
PC aa C42:1	1.10	0.79	1.52	0.5756	0.9578
PC aa C42:2	1.03	0.74	1.44	0.8532	0.9578
PC aa C42:4	0.73	0.54	0.99	0.0446	0.7774
PC aa C42:5	1.00	0.70	1.42	0.9832	0.9832
PC aa C42:6	0.94	0.67	1.33	0.7452	0.9578
PC ae C30:0	1.17	0.81	1.68	0.3976	0.9578
PC ae C30:2	1.41	0.97	2.05	0.0745	0.9093
PC ae C32:1	0.92	0.67	1.25	0.5797	0.9578
PC ae C32:2	0.99	0.75	1.32	0.9658	0.9832
PC ae C34:0	1.13	0.86	1.50	0.3713	0.9578
PC ae C34:1	1.03	0.78	1.36	0.8342	0.9578
PC ae C34:2	1.08	0.82	1.42	0.6018	0.9578
PC ae C34:3	0.90	0.67	1.21	0.4899	0.9578
PC ae C36:0	1.03	0.79	1.34	0.8347	0.9578
PC ae C36:1	1.09	0.83	1.43	0.5289	0.9578
PC ae C36:2	1.24	0.92	1.68	0.1544	0.9578
PC ae C36:3	1.01	0.76	1.34	0.9294	0.9832
PC ae C36:4	0.99	0.73	1.34	0.9501	0.9832
PC ae C36:5	0.86	0.62	1.19	0.3679	0.9578
PC ae C38:0	1.00	0.74	1.34	0.9760	0.9832
PC ae C38:2	0.95	0.70	1.28	0.7297	0.9578
PC ae C38:3	1.16	0.87	1.54	0.3211	0.9578
PC ae C38:4	1.03	0.78	1.36	0.8292	0.9578
PC ae C38:5	0.91	0.67	1.24	0.5425	0.9578
PC ae C38:6	0.92	0.68	1.25	0.6127	0.9578
PC ae C40:1	0.95	0.71	1.29	0.7614	0.9578
PC ae C40:2	1.02	0.76	1.37	0.8828	0.9703
PC ae C40:3	1.13	0.82	1.55	0.4636	0.9578
PC ae C40:4	1.06	0.80	1.39	0.6920	0.9578
PC ae C40:5	0.92	0.68	1.25	0.5915	0.9578
PC ae C40:6	1.13	0.82	1.55	0.4478	0.9578

Continues...

Table 6.13 continued

Metabolites	1 SD increase			p-trend[3]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
PC ae C42:1	0.94	0.70	1.26	0.6765	0.9578
PC ae C42:2	0.97	0.73	1.30	0.8584	0.9578
PC ae C42:3	0.94	0.67	1.31	0.7180	0.9578
PC ae C42:4	1.09	0.81	1.48	0.5605	0.9578
PC ae C42:5	0.97	0.72	1.30	0.8199	0.9578
PC ae C44:4	1.04	0.72	1.51	0.8249	0.9578
PC ae C44:5	1.02	0.76	1.36	0.9065	0.9832
PC ae C44:6	1.10	0.81	1.49	0.5408	0.9578
Hexose					
Hexose	1.57	1.07	2.31	0.0208	0.7286
Sphingolipids					
SM (OH) C14:1	1.41	1.02	1.94	0.0358	0.7286
SM (OH) C16:1	1.29	0.94	1.75	0.1114	0.9578
SM (OH) C22:1	1.05	0.79	1.39	0.7420	0.9578
SM (OH) C22:2	1.10	0.82	1.48	0.5204	0.9578
SM (OH) C24:1	1.09	0.82	1.44	0.5455	0.9578
SM C16:0	1.03	0.77	1.37	0.8506	0.9578
SM C16:1	1.08	0.79	1.46	0.6349	0.9578
SM C18:0	1.23	0.91	1.67	0.1716	0.9578
SM C18:1	1.21	0.89	1.65	0.2332	0.9578
SM C20:2	0.92	0.63	1.35	0.6758	0.9578
SM C24:0	1.05	0.78	1.41	0.7477	0.9578
SM C24:1	0.99	0.72	1.36	0.9586	0.9832

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Death from prostate cancer: prostate cancer listed as the underlying cause of death on the death certificate during follow-up. 144 died of prostate cancer but 17 were excluded from further analysis as their matched control had died before them (n $= 13$) or vital status was not known for their control (n $= 4$)

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

Table 6.14: Risk of overall prostate cancer in 985 matched cases control sets after excluding the overlap in men with the analysis in EPIC-Heidelberg[1,2]

Metabolites	1 SD increase			p-trend[3]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
Acylcarnitines					
C0	0.94	0.86	1.03	0.1951	0.9274
C2	0.97	0.88	1.07	0.5043	0.9274
C3	1.05	0.92	1.20	0.4633	0.9274
C14:1	0.94	0.85	1.05	0.2861	0.9274
C16	1.01	0.91	1.12	0.8520	0.9967
C18:1	0.92	0.82	1.03	0.1390	0.9274
C18:2	0.95	0.84	1.07	0.4170	0.9274
Amino acids					
Alanine	0.97	0.87	1.07	0.5136	0.9274
Arginine	1.04	0.91	1.19	0.5549	0.9342
Asparagine	1.03	0.93	1.14	0.5590	0.9342
Citrulline	0.92	0.83	1.02	0.1180	0.9274
Glutamate	1.10	0.96	1.26	0.1770	0.9274
Glutamine	0.99	0.90	1.09	0.8089	0.9967
Glycine	1.01	0.91	1.12	0.9109	0.9967
Histidine	1.00	0.90	1.11	0.9882	0.9967
Isoleucine	1.10	1.00	1.22	0.0570	0.8690
Leucine	1.06	0.96	1.17	0.2372	0.9274
Lysine	0.98	0.89	1.08	0.7160	0.9967
Methionine	1.01	0.91	1.12	0.8428	0.9967
Ornithine	1.02	0.90	1.16	0.7811	0.9967
Phenylalanine	1.08	0.98	1.20	0.1357	0.9274
Proline	1.06	0.96	1.16	0.2700	0.9274
Serine	1.04	0.94	1.15	0.4872	0.9274
t4-OH-Pro	1.13	1.03	1.24	0.0125	0.3138
Threonine	1.04	0.94	1.14	0.4455	0.9274
Tryptophan	1.02	0.93	1.12	0.6721	0.9798
Tyrosine	1.00	0.91	1.11	0.9798	0.9967
Valine	1.06	0.95	1.17	0.2990	0.9274
Biogenic amines					
ADMA	1.04	0.94	1.15	0.4459	0.9274
Creatinine	1.05	0.95	1.15	0.3534	0.9274
Kynurenine	1.06	0.96	1.18	0.2360	0.9274
Sarcosine	0.99	0.90	1.09	0.9085	0.9967
SDMA	0.98	0.88	1.09	0.7240	0.9967
Taurine	0.98	0.84	1.13	0.7438	0.9967
Glycerophospholipids					
lysoPC a C16:0	0.96	0.86	1.08	0.5187	0.9274
lysoPC a C16:1	0.93	0.84	1.03	0.1563	0.9274
lysoPC a C17:0	1.01	0.91	1.12	0.7987	0.9967
lysoPC a C18:0	0.93	0.84	1.04	0.2269	0.9274
lysoPC a C18:1	0.92	0.82	1.02	0.1036	0.9274
lysoPC a C18:2	0.98	0.89	1.09	0.7596	0.9967
lysoPC a C20:3	0.94	0.84	1.04	0.2280	0.9274
lysoPC a C20:4	0.96	0.85	1.07	0.4421	0.9274
PC aa C28:1	1.15	1.03	1.28	0.0094	0.3138
PC aa C30:0	1.06	0.96	1.17	0.2715	0.9274
PC aa C32:0	1.05	0.95	1.16	0.3597	0.9274
PC aa C32:1	1.01	0.91	1.12	0.8668	0.9967
PC aa C32:2	1.02	0.93	1.13	0.6826	0.9798
PC aa C32:3	0.99	0.88	1.10	0.8195	0.9967

Continues...

Table 6.14 continued

Metabolites	1 SD increase			p-trend[3]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
PC aa C34:1	0.99	0.90	1.10	0.9219	0.9967
PC aa C34:2	1.03	0.94	1.14	0.5044	0.9274
PC aa C34:3	0.98	0.87	1.10	0.7541	0.9967
PC aa C34:4	1.00	0.91	1.11	0.9232	0.9967
PC aa C36:0	1.04	0.94	1.15	0.4424	0.9274
PC aa C36:1	0.99	0.90	1.08	0.7636	0.9967
PC aa C36:2	0.98	0.89	1.08	0.6433	0.9798
PC aa C36:3	0.95	0.87	1.05	0.2983	0.9274
PC aa C36:4	1.00	0.90	1.11	0.9791	0.9967
PC aa C36:5	1.05	0.95	1.16	0.2961	0.9274
PC aa C36:6	1.05	0.95	1.15	0.3418	0.9274
PC aa C38:0	1.02	0.92	1.14	0.6762	0.9798
PC aa C38:3	0.95	0.86	1.05	0.3339	0.9274
PC aa C38:4	0.95	0.86	1.05	0.3343	0.9274
PC aa C38:5	0.99	0.89	1.09	0.7975	0.9967
PC aa C38:6	1.04	0.94	1.14	0.4588	0.9274
PC aa C40:2	0.98	0.88	1.10	0.7744	0.9967
PC aa C40:3	0.99	0.89	1.11	0.8529	0.9967
PC aa C40:4	0.98	0.89	1.07	0.6135	0.9680
PC aa C40:5	0.96	0.86	1.06	0.3973	0.9274
PC aa C40:6	1.00	0.90	1.10	0.9406	0.9967
PC aa C42:0	1.04	0.94	1.15	0.4543	0.9274
PC aa C42:1	1.03	0.94	1.14	0.5047	0.9274
PC aa C42:2	1.05	0.95	1.16	0.3568	0.9274
PC aa C42:4	0.96	0.88	1.06	0.4315	0.9274
PC aa C42:5	0.97	0.86	1.09	0.6211	0.9680
PC aa C42:6	0.99	0.89	1.10	0.8381	0.9967
PC ae C30:0	1.17	1.04	1.31	0.0083	0.3138
PC ae C30:2	1.16	1.03	1.30	0.0129	0.3138
PC ae C32:1	1.04	0.94	1.15	0.4139	0.9274
PC ae C32:2	1.04	0.94	1.15	0.4223	0.9274
PC ae C34:0	1.10	1.00	1.21	0.0480	0.8690
PC ae C34:1	1.05	0.95	1.16	0.3261	0.9274
PC ae C34:2	1.07	0.97	1.17	0.1602	0.9274
PC ae C34:3	1.01	0.92	1.11	0.8819	0.9967
PC ae C36:0	1.10	1.00	1.21	0.0516	0.8690
PC ae C36:1	1.06	0.96	1.17	0.2205	0.9274
PC ae C36:2	1.06	0.96	1.18	0.2410	0.9274
PC ae C36:3	1.03	0.93	1.13	0.5752	0.9348
PC ae C36:4	1.03	0.94	1.14	0.5129	0.9274
PC ae C36:5	1.01	0.91	1.12	0.8149	0.9967
PC ae C38:0	1.01	0.91	1.12	0.8933	0.9967
PC ae C38:2	1.04	0.93	1.15	0.5127	0.9274
PC ae C38:3	1.05	0.96	1.16	0.2977	0.9274
PC ae C38:4	1.03	0.93	1.13	0.5823	0.9348
PC ae C38:5	1.01	0.91	1.11	0.8892	0.9967
PC ae C38:6	1.04	0.94	1.15	0.4994	0.9274
PC ae C40:1	0.99	0.90	1.10	0.9157	0.9967
PC ae C40:2	1.06	0.96	1.18	0.2422	0.9274
PC ae C40:3	1.05	0.95	1.17	0.3216	0.9274
PC ae C40:4	1.05	0.96	1.15	0.2819	0.9274
PC ae C40:5	1.03	0.94	1.14	0.5245	0.9274
PC ae C40:6	1.05	0.95	1.16	0.3478	0.9274

Continues...

Table 6.14 continued

Metabolites	1 SD increase			p-trend[3]	p_{adj}-trend
	OR	95% CI			
		LL	UL		
PC ae C42:1	0.98	0.89	1.08	0.6577	0.9798
PC ae C42:2	0.97	0.88	1.07	0.5705	0.9348
PC ae C42:3	1.00	0.90	1.11	0.9836	0.9967
PC ae C42:4	1.04	0.94	1.15	0.4121	0.9274
PC ae C42:5	1.03	0.94	1.13	0.5454	0.9342
PC ae C44:4	1.00	0.91	1.10	0.9967	0.9967
PC ae C44:5	1.00	0.91	1.09	0.9404	0.9967
PC ae C44:6	1.02	0.93	1.12	0.6749	0.9798
Hexose					
Hexose	1.05	0.93	1.20	0.4158	0.9274
Sphingolipids					
SM (OH) C14:1	1.15	1.03	1.28	0.0112	0.3138
SM (OH) C16:1	1.09	0.99	1.21	0.0802	0.9274
SM (OH) C22:1	1.01	0.91	1.12	0.8454	0.9967
SM (OH) C22:2	1.06	0.96	1.18	0.2300	0.9274
SM (OH) C24:1	1.02	0.92	1.13	0.7144	0.9967
SM C16:0	1.05	0.95	1.16	0.3172	0.9274
SM C16:1	1.03	0.94	1.13	0.5325	0.9281
SM C18:0	1.04	0.94	1.15	0.4410	0.9274
SM C18:1	1.00	0.91	1.10	0.9669	0.9967
SM C20:2	1.00	0.87	1.14	0.9906	0.9967
SM C24:0	1.03	0.93	1.13	0.6268	0.9680
SM C24:1	1.06	0.95	1.17	0.3216	0.9274

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (<10; $10-19$; $20-39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Kühn et al. [84]

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

6.8 Figures

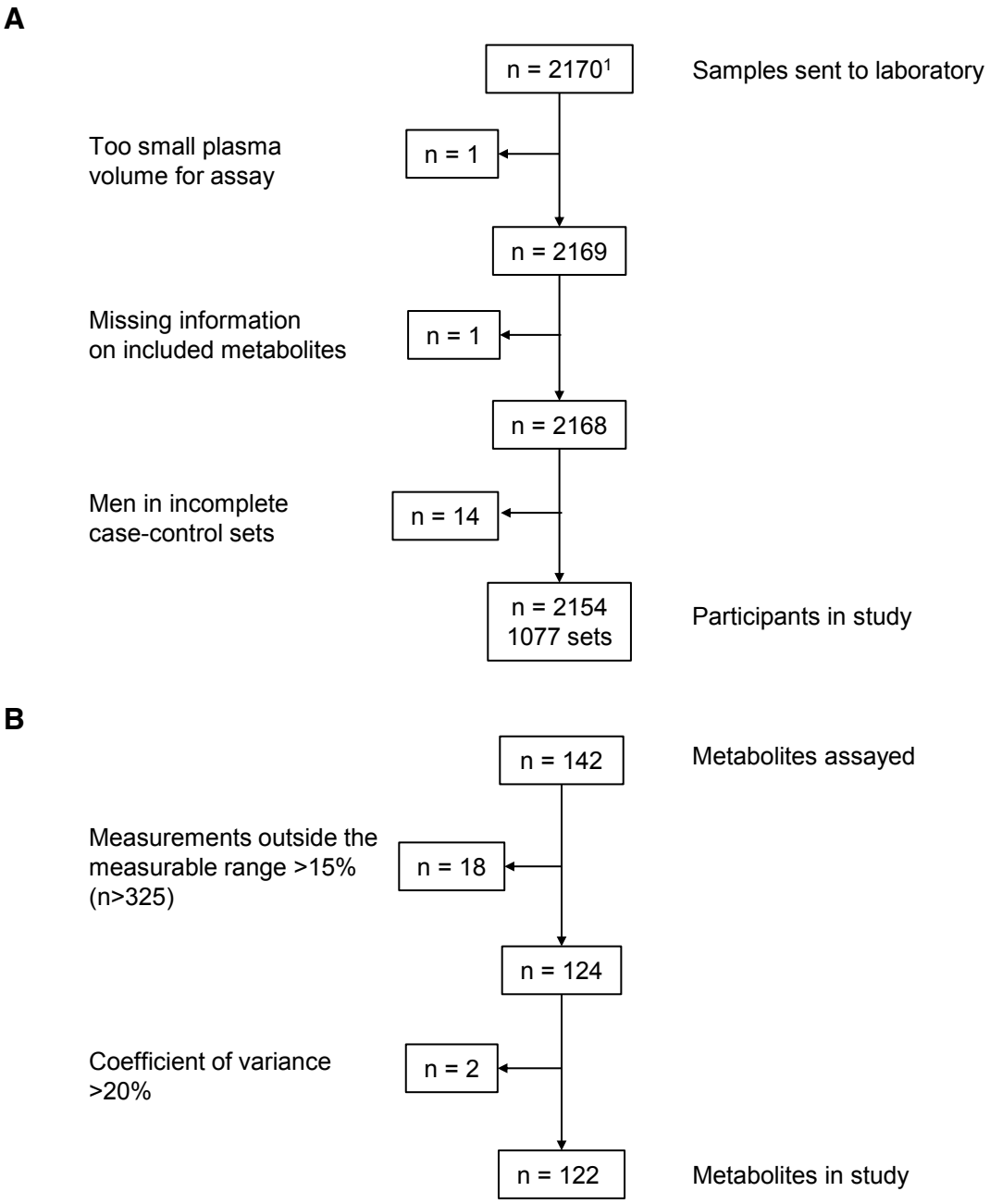


Figure 6.1: Flowchart of exclusions of metabolites and participants

A. Participants. B. Metabolites. [1] There were 2,162 samples but eight samples contributed to the dataset twice because three controls became cases and five controls were controls for two cases each.

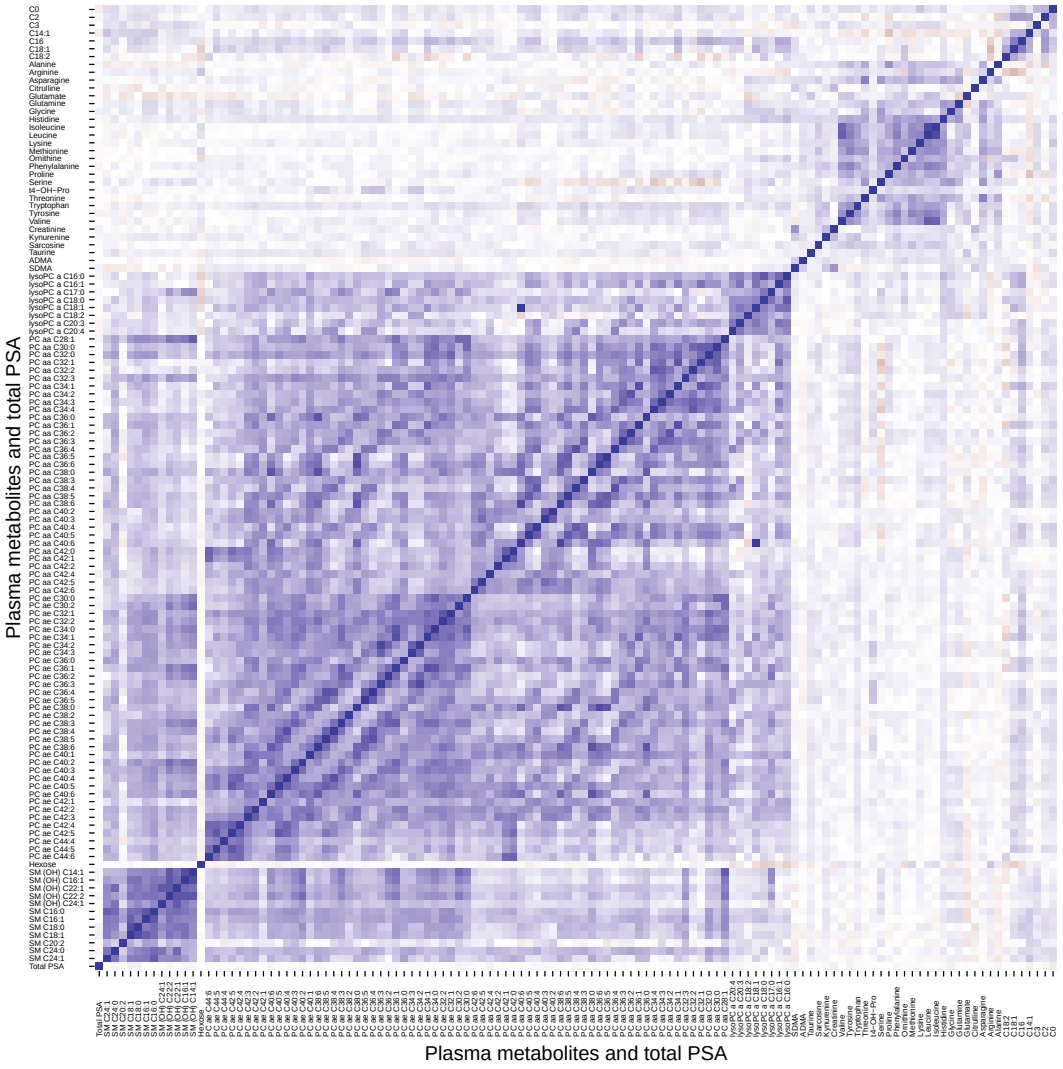


Figure 6.2: Partial correlation coefficients between metabolite and total PSA concentrations

Correlations between total PSA and metabolites were based on 764 controls for which total PSA was available, while correlations between metabolites included 1,077 controls. The analyses were adjusted for age at blood collection (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years), BMI (fourths; unknown) and study centre.

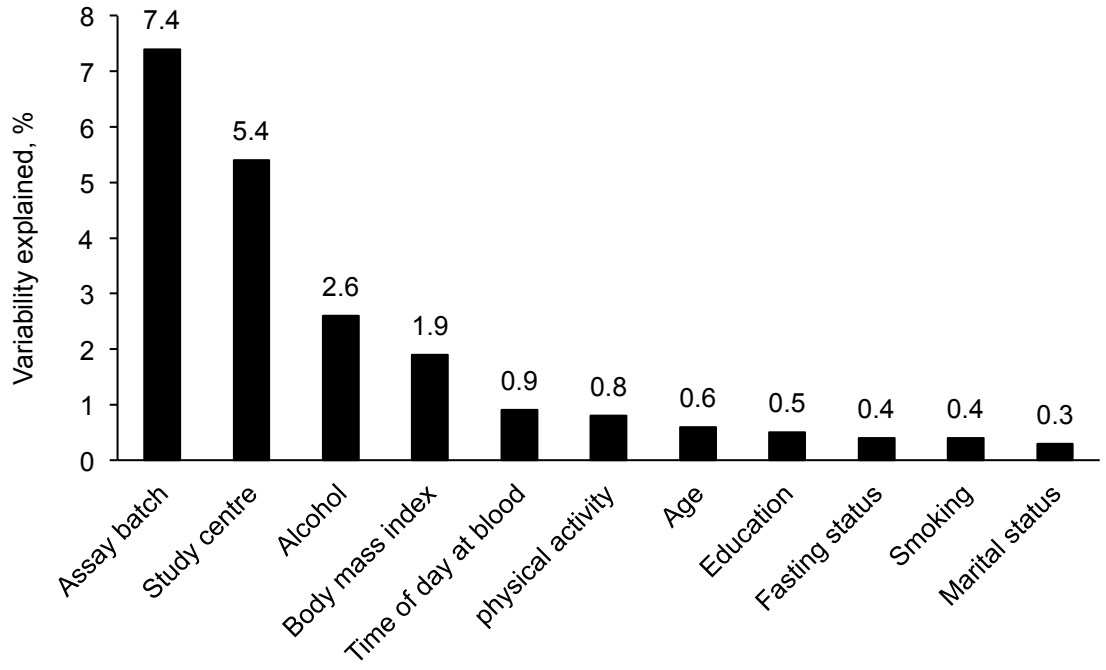


Figure 6.3: Variability in metabolite concentrations explained by covariates

The percentage of the total variability in metabolite concentrations explained by participant and blood sample characteristics was calculated using the PC-PR2 method [161].

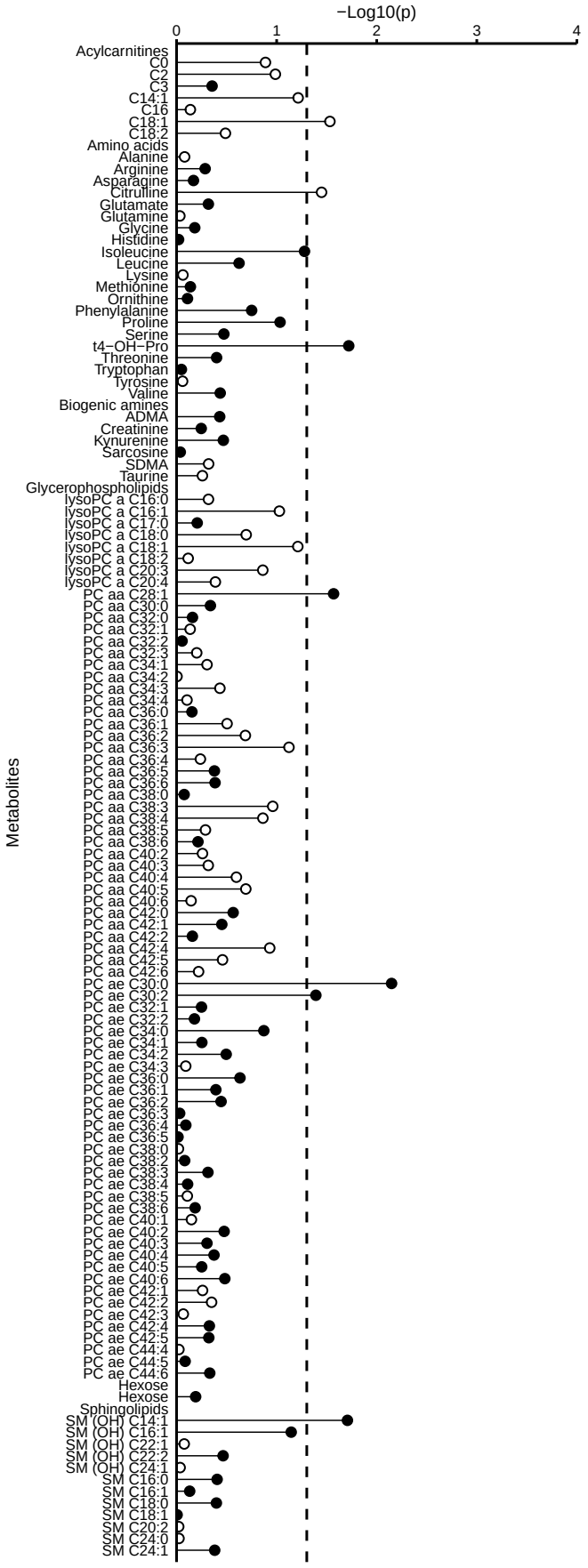


Figure 6.4: Statistical significance of associations between metabolite concentrations and risk of overall prostate cancer in EPIC

The analysis included 1,077 matched case-control sets. Statistical significance was plotted as $-\log10$(p-values). The dashed line represents conventionally statistical significance at $\alpha = 0.05$. Filled circles represent positive associations and unfilled circles represent inverse associations. The p-values were derived from a conditional logistic regression using log-metabolite concentration as a continuous variable and adjusting for exact age, BMI, smoking, alcohol intake, education and marital status.

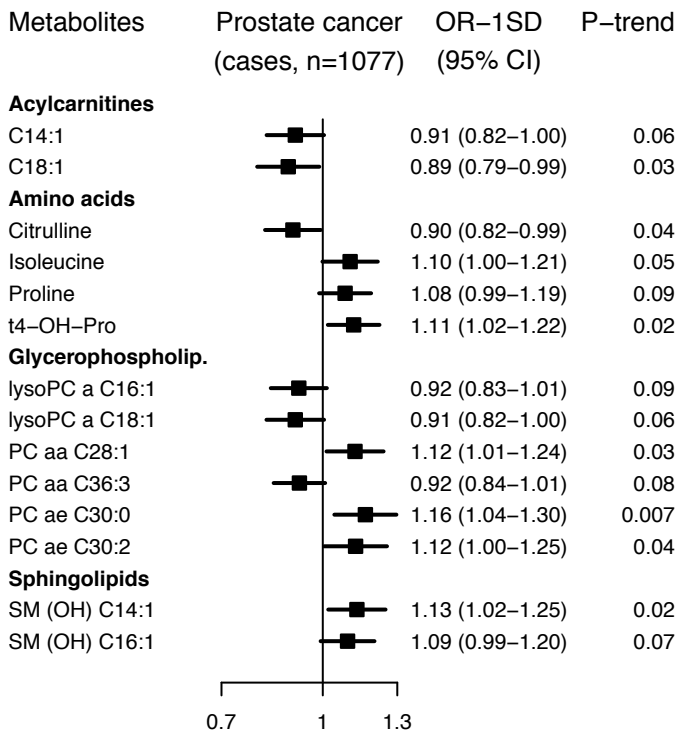


Figure 6.5: Odds ratios for overall prostate cancer risk by concentration of selected metabolites in EPIC
Metabolites with $p_{trend} < 0.1$ were included in the figure; no associations were statistically significant after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). Odds ratios for one standard deviation increase in metabolite concentrations, 95%CI and p_{trend} were derived from a conditional logistic regression using log-metabolite concentration divided by the standard deviation of log-metabolite concentration as a continuous variable and and adjusting for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown).

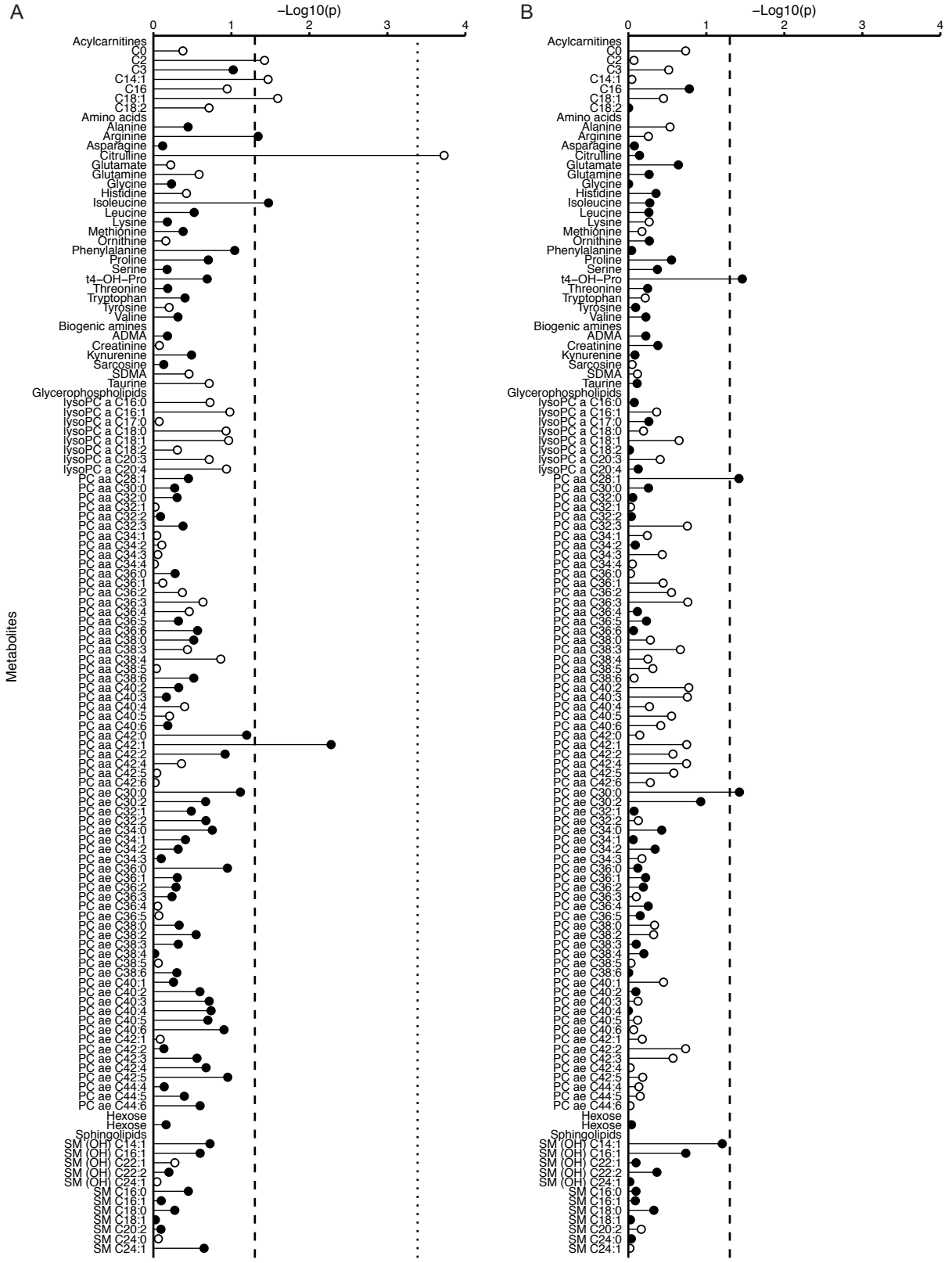


Figure 6.6: Statistical significance of associations between metabolite concentrations and prostate cancer risk by time to diagnosis in EPIC

A. Five years or less between blood collection and diagnosis; n = 428 matched case-control sets. B. More than five years between blood collection and diagnosis; n = 649 matched sets. Statistical significance was plotted as $-$log10(p-values). The dashed and the dotted lines represent conventionally statistical significance and statistical significance after controlling the false discovery rate (Benjamini–Hochberg), respectively, both at $\alpha = 0.05$. Filled circles represent positive associations and unfilled circles represent inverse associations. The p-values were derived from a conditional logistic regression using log-metabolite concentration as a continuous variable and adjusting for and adjusting for exact age, BMI, smoking, alcohol intake, education and marital status .

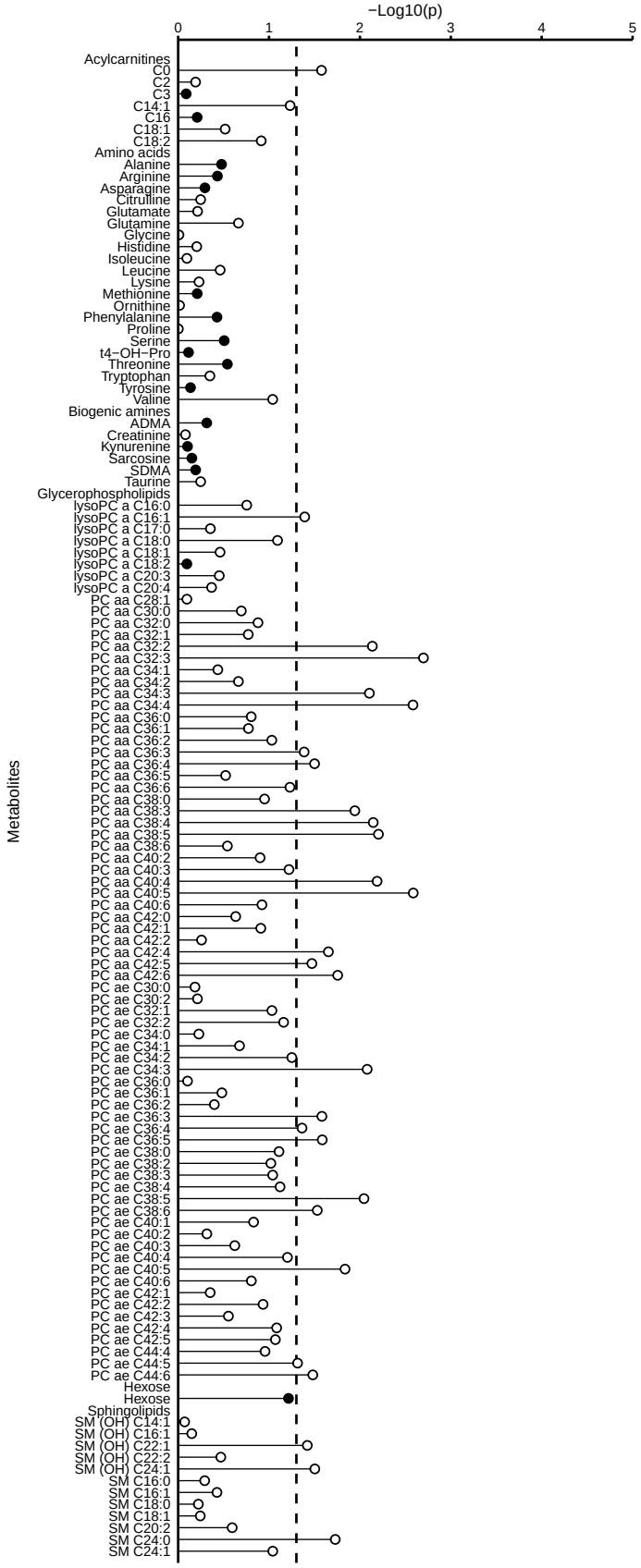


Figure 6.7: Statistical significance of the associations between metabolite concentrations and risk of high grade prostate cancer in EPIC

Tumours were high grade if Gleason score $\geq$ 8 or if coded as undifferentiated. The analysis included 124 matched case-control sets. Statistical significance was plotted as $-$log10(p-values). The dashed line represents conventionally statistical significance at $\alpha = 0.05$; no associations were statistical significance after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). Filled circles represent positive associations and unfilled circles represent inverse associations. The p-values were derived from a conditional logistic regression using log-metabolite concentration as a continuous variable and adjusting for exact age, BMI, smoking, alcohol intake, educationand marital status.

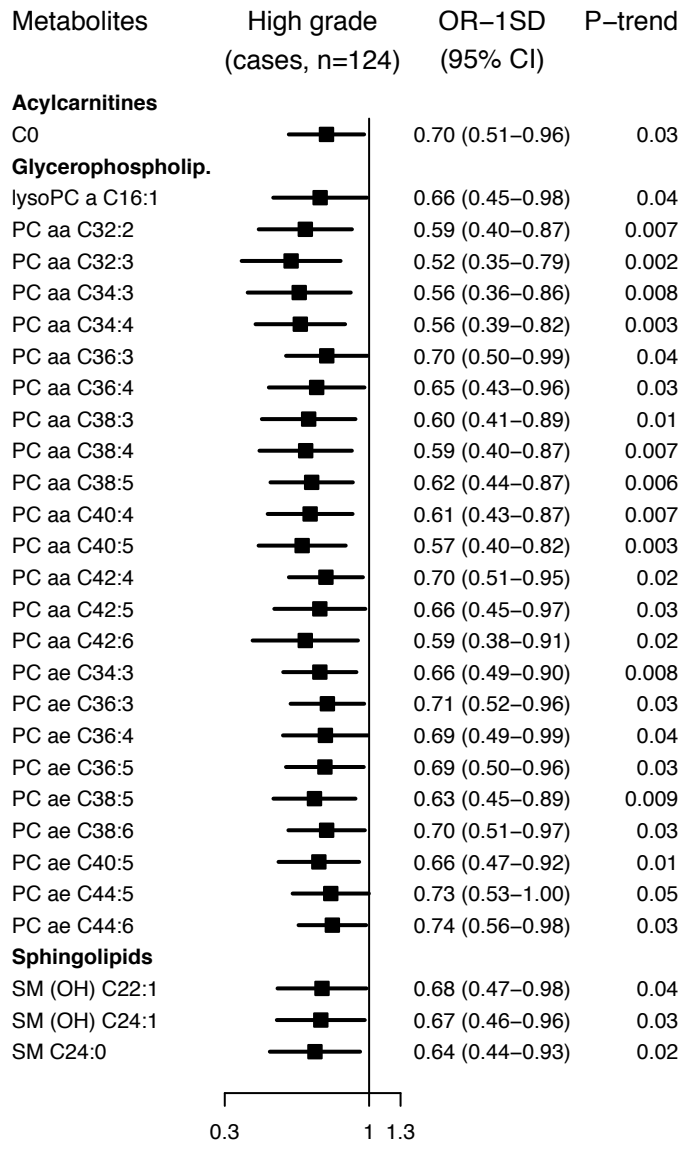


Figure 6.8: Odds ratios for high grade prostate cancer by concentration of selected metabolites in EPIC Tumours were high grade if Gleason score $\geq$ 8 or if coded as undifferentiated. Metabolites with $p_{trend} < 0.05$ were included in the figure; no associations were statistically significant after controlling the false discovery rate at $\alpha = 0.05$ (Benjemini-Hochberg). Odds ratios for one standard deviation increase in metabolite concentrations, 95%CI and p_{trend} were derived from a conditional logistic regression using log-metabolite concentration divided by the standard deviation of log-metabolite concentration as a continuous variable and adjusting for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown).

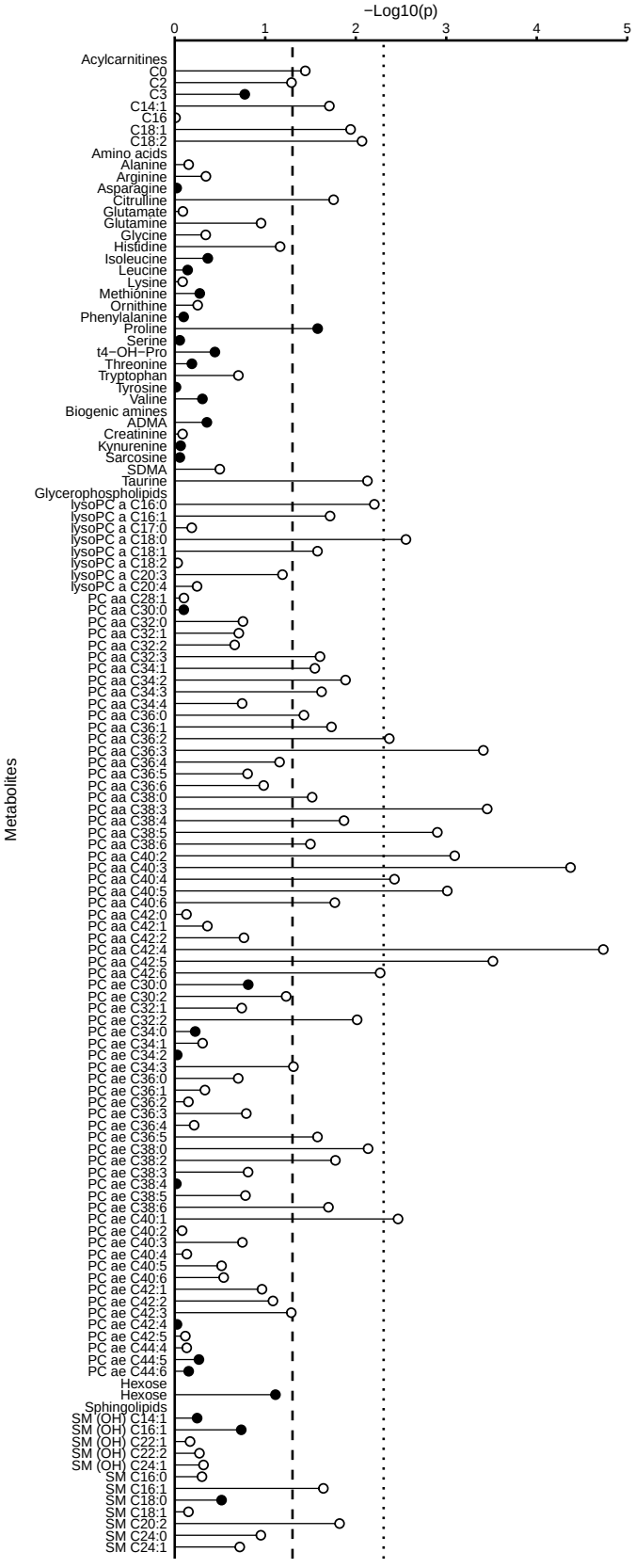


Figure 6.9: Statistical significance of associations between metabolite concentrations and risk of advanced stage prostate cancer in EPIC

Advanced stage prostate cancer was TMN-score of T3 − T4 and/or N1 − 3 and/or M1. The analysis included 208 matched case-control sets. Statistical significance was plotted as −log10(p-values). The dashed and the dotted lines represent conventionally statistical significance and statistical significance after allowing for multiple testing using a false discovery rate-controlling procedure (Benjamini–Hochberg), respectively, both at $\alpha = 0.05$. Filled circles represent positive associations and unfilled circles represent inverse associations. The p-values were derived from a conditional logistic regression using log-metabolite concentration as a continuous variable and adjusting for exact age, BMI, smoking, alcohol intake, education and marital status.

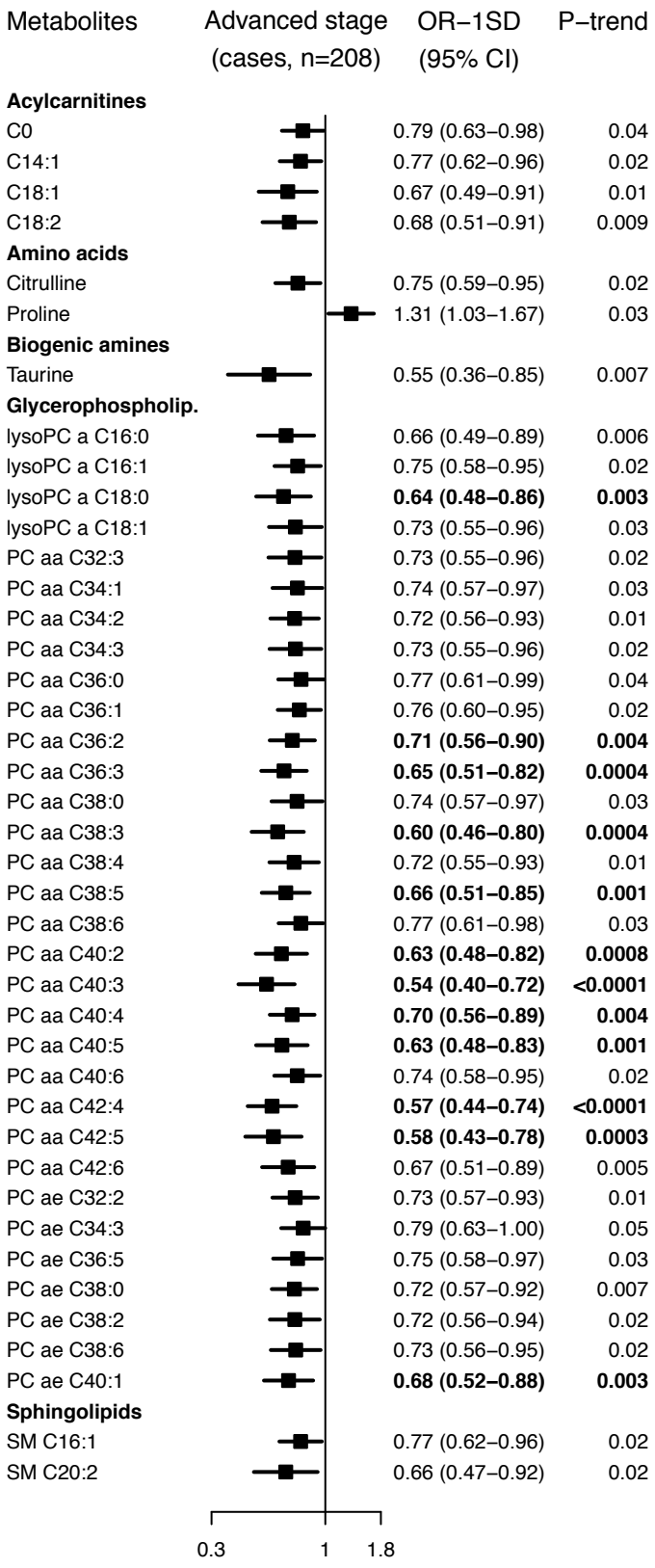


Figure 6.10: Odds ratios for advanced stage prostate cancer by concentration of selected metabolites in EPIC

Advanced stage prostate cancer was TMN-score of T3 − T4 and/or N1 − 3 and/or M1. Metabolites with $p_{trend} < 0.05$ were included in the figure and values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg). Odds ratios for one standard deviation increase in metabolite concentrations, 95%CI and p_{trend} were derived from a conditional logistic regression using log-metabolite concentration divided by the standard deviation of log-metabolite concentration as a continuous variable and adjusting for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown).

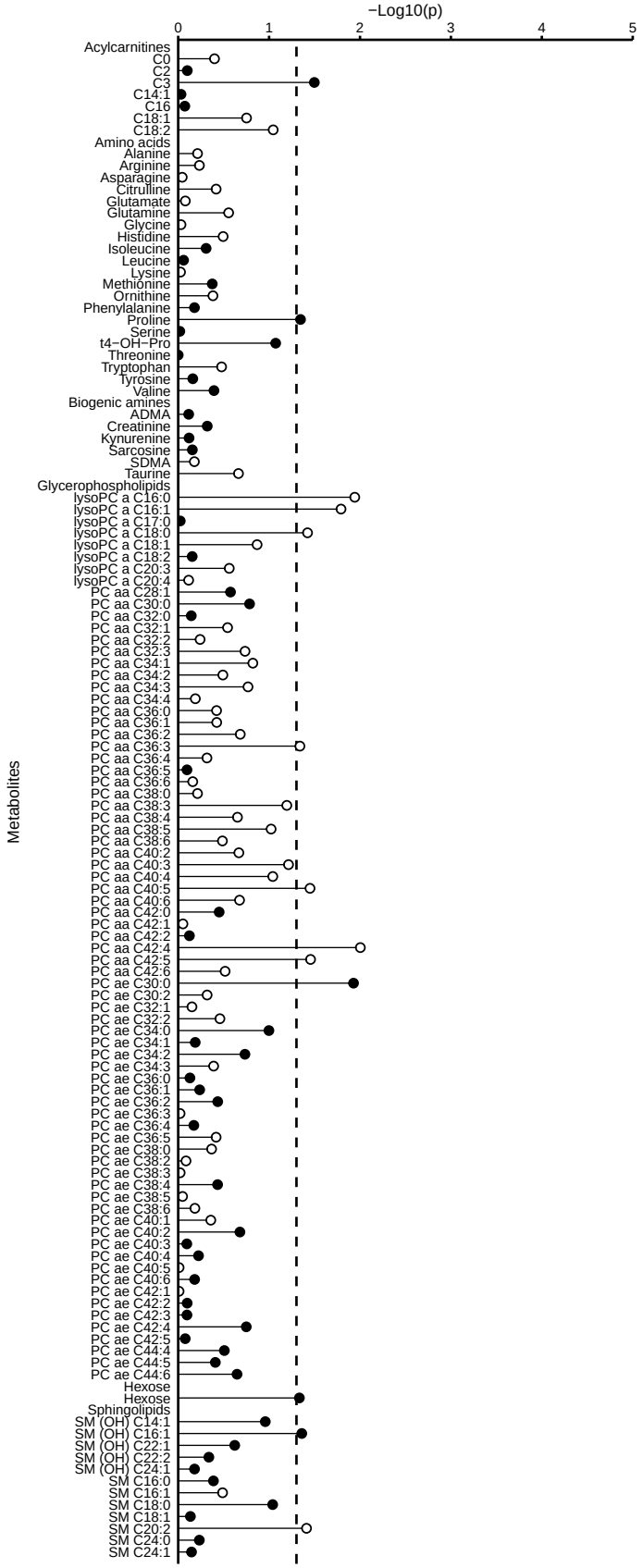


Figure 6.11: Statistical significance of the associations between metabolite concentrations and risk of aggressive prostate cancer in EPIC

Aggressive prostate cancer was tumours with TNM-score of T4 and/or N1 $-$ 3 and/or M1. The analysis included 115 matched case-control. Statistical significance was plotted as $-$log10(p-values). The dashed line represents conventionally statistical significance at $\alpha = 0.05$; no associations were statistical significance after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). Filled circles represent positive associations and unfilled circles represent inverse associations. The p-values were derived from a conditional logistic regression using log-metabolite concentration as a continuous variable and adjusting for exact age, BMI, smoking, alcohol intake , education and marital status.

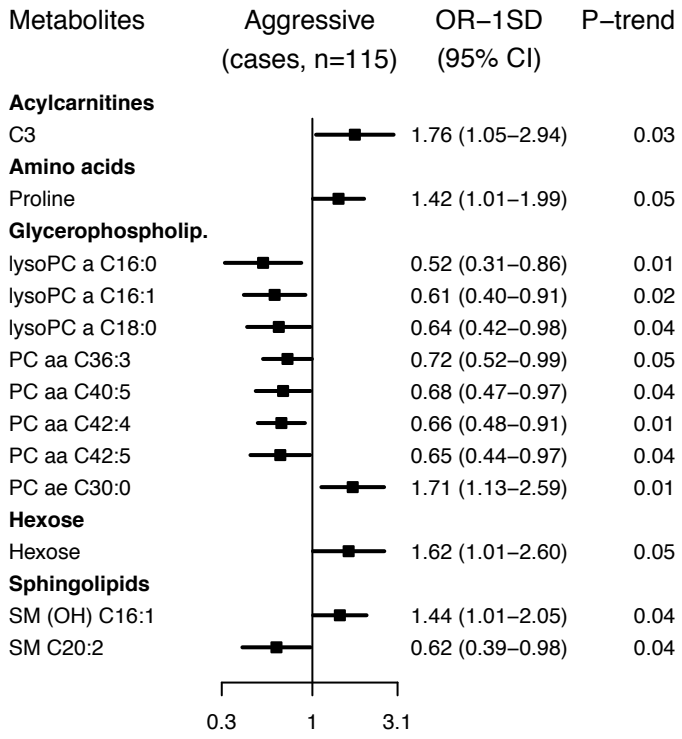


Figure 6.12: Odds ratios for aggressive prostate cancer by concentration of selected metabolites in EPIC

Aggressive prostate cancer was tumours with TNM-score of T4 and/or N1 $-$ 3 and/or M1. Metabolites with $p_{trend} <$ 0.05 were included in the figure; no associations were statistically significant after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). Odds ratios for one standard deviation increase in metabolite concentrations, 95%CI and p_{trend} were derived from a conditional logistic regression using log-metabolite concentration divided by the standard deviation of log-metabolite concentration as a continuous variable and adjusting for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown).

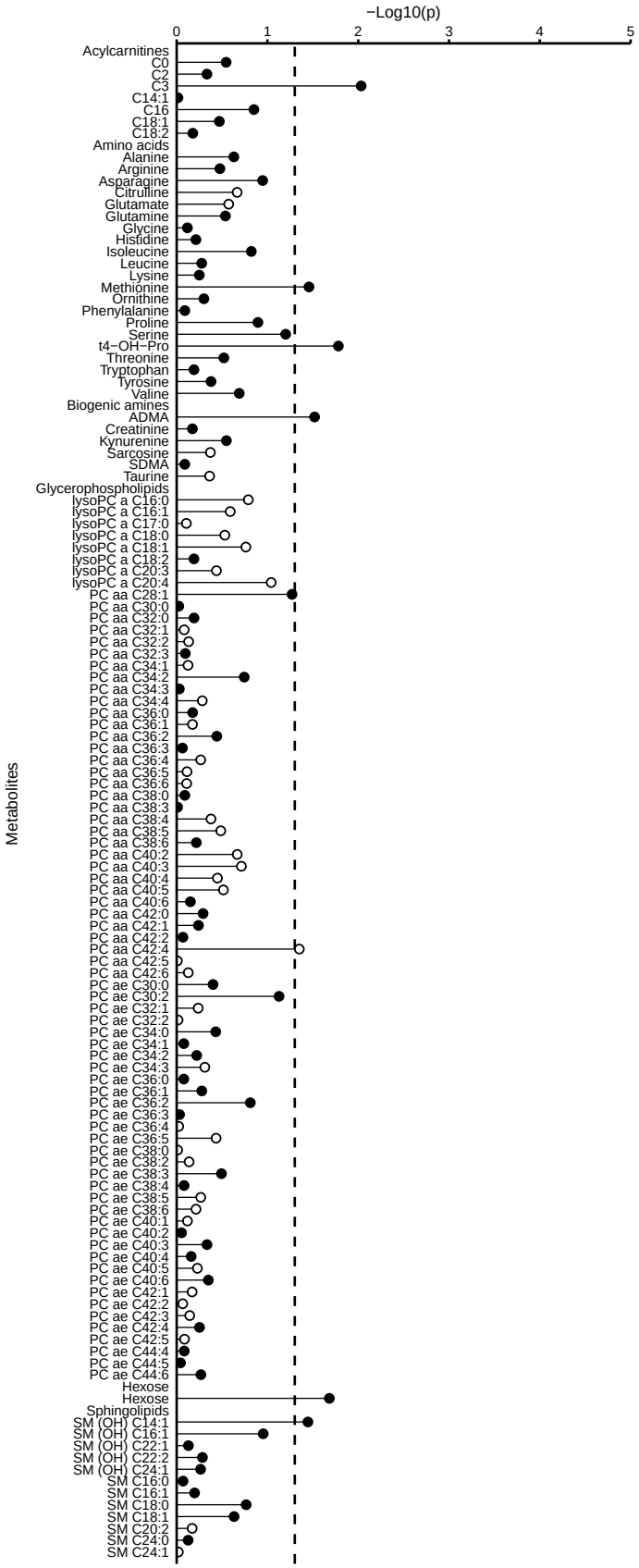


Figure 6.13: Statistical significance of associations between metabolite concentrations and risk of death from prostate cancer in EPIC

The analysis included 127 matched case-control sets. Statistical significance was plotted as $-$log10(p-values). The dashed line represents conventionally statistical significance at $\alpha = 0.05$; no associations were statistical significance after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). Filled circles represent positive associations and unfilled circles represent inverse associations. The p-values were derived from a conditional logistic regression using log-metabolite concentration as a continuous variable and adjusting for exact age, BMI, smoking, alcohol intake, educationand marital status.

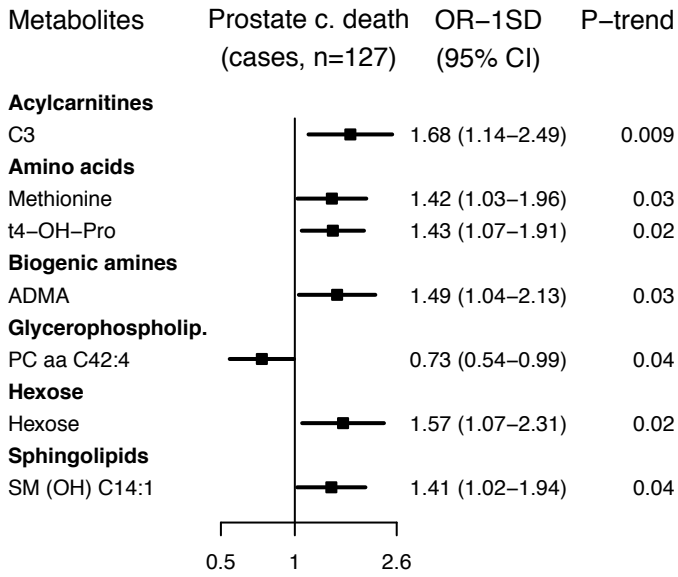


Figure 6.14: Odds ratios for death from prostate cancer by selected concentration of metabolites in EPIC

Metabolites with $p_{trend} < 0.05$ were included in the figure; no associations were statistically significant after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). Odds ratios for one standard deviation increase in metabolite concentrations, 95%CI and p_{trend} were derived from a conditional logistic regression using log-metabolite concentration divided by the standard deviation of log-metabolite concentration as a continuous variable and adjusting for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown).

7. Metabolite patterns and prostate cancer risk

7.1 Introduction

Metabolomics might help identify novel risk factors for prostate cancer, for which few risk factors are established. In Chapter 6, an analysis of individual metabolite concentrations and risk of prostate cancer was presented. Seven metabolites from several metabolite classes were found to be associated with risk of overall prostate cancer at a conventional level of significance but not after controlling for multiple testing. The results stratified by time to diagnosis suggested that lower concentrations of citrulline might be a marker of sub-clinical prostate cancer. Analyses by tumour sub-types indicated possible associations between higher concentrations of several glycerophospholipids and decreased risk of high-risk tumours, especially advanced stage prostate cancer. Suggestive associations were also observed between death from prostate cancer and seven metabolites from several metabolite classes.

Looking at metabolites individually will not capture patterns in the metabolite profile originating from the correlations between metabolites [140]; such patterns might represent metabolites which are involved in the same metabolic pathway [201] or originate from the same dietary sources. Moreover, associations of individual metabolites with risk might be weak and thus missed [140]. Finally, the risk of false positive findings increases with the number of tests performed and has to be controlled for when investigating many individual associations.

To overcome these limitations of univariate analyses, multivariate approaches, including dimension reduction methods, can be applied. Dimension reduction methods summarise the original data into fewer new variables (components), describing latent structures of the original data [140]. Two such methods are the principal component analysis (PCA) and treelet transform (TT). PCA is the most commonly used method and explains the maximum variation in the data by including all metabolites in the components. TT is a more recently developed method and patterns identified by TT may be more easily interpreted than those from PCA because only some metabolites contribute to each component; the trade off is that less of the total variance in metabolite concentrations is explained [160, 202].

7.1.1 Previous studies

Principal component analysis has been widely applied to describe dietary patterns in studies of diet and disease outcomes [203] and has also been used to characterise metabolite profile in nested and classic case-control studies of metabolites in relation to prostate cancer risk [11, 29]. Although TT is less frequently used, it has been used to describe patterns in serum metabolite

profile [201], adipose tissue fatty acids [204, 205] and dietary intake [160, 206, 207] in studies of these factors in relation to anthropometry and disease outcomes, including prostate cancer risk [204]. The one previous study which used TT to identify serum metabolite profile also used PCA and both methods identified four comparable patterns characterised by amino acids, amino acid derivatives, fatty acids, and sugar compounds and carboxylic acids, respectively [201]. The authors concluded that TT is a method well-suited for summarising high-dimensional data in epidemiological studies, and that patterns identified can readily be analysed in relation to other variables.

Little is known about the prospective association between metabolite concentrations and risk of prostate cancer. Apart from the paper published based on the work described in Chapter 6 [19], to our knowledge, only five publications from four small studies (n_{cases} = 74 to 380) have been published. Of these [28, 29, 82, 83, 84], only two investigated groups of metabolites defined using PCA or other methods, in addition to the univariate analyses [28, 29]. No associations between principal components and risk of overall or aggressive prostate cancer were reported [29]. To our knowledge, the current analysis is the first study to investigate metabolite patterns and risk of death from prostate cancer.

7.2 Aim

The aim of this analysis was to investigate the prospective associations of patterns in pre-diagnostic metabolite concentrations, using PCA and TT, with risk of prostate cancer overall, as well as by time to diagnosis and by tumour characteristics, and with risk of death from prostate cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). The studied metabolites were acylcarnitines, amino acids, biogenic amines, glycerophospholipids, hexose and sphingolipids.

7.3 Methods

7.3.1 Study population

The EPIC cohort has been described in detail in Chapter 3. In brief, the cohort includes 153,400 men recruited from 19 centres in eight European countries (Denmark, Germany, Greece, Italy, Netherlands, Spain, Sweden and United Kingdom) between 1992 and 2000, when they were aged mostly between 35 and 70 years. At recruitment, detailed information was collected on dietary intake, lifestyle, anthropometry and previous disease, and 139,600 men also gave a blood sample [99].

All participants gave written informed consent to participate in the EPIC cohort. Approval for the current study was obtained from the Internal Review Board of the International Agency for Research on Cancer (IARC), Lyon, France (Project No. 14-09) and participating centres as required.

For the current analysis, men were eligible if they had blood stored at the central biobank at IARC (15 centres in Germany, Greece, Italy, Netherlands, Spain and United Kingdom), the date of blood collection was known and if no cancer (except non-melanoma skin cancer) had been diagnosed at the time of blood collection.

7.3.1.1 Follow-up and case and control selection

Information on cancer incidence, tumour sub-types and vital status was obtained via record linkage to population-based cancer and mortality registries in most countries. However in Germany and Greece, a combination of methods was used, including health insurance records, local (municipality-, hospital- or physician-based) cancer and pathology registries, and active follow-up via participants or their next-of-kin; self-reported cancer incidence was verified via medical records.

Cases: Prostate cancer was defined as code C61 in the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) and cases were men diagnosed with prostate cancer after blood collection and prior to the end of follow-up. For these analyses, samples were available for cases from the first three phases of centralisation of follow-up data; with end of follow-up ranging between centres from 2007 to 2008.

Histological grade was known for 76.8% of cases, and 1,041 men were diagnosed with low-intermediate grade tumours (Gleason score < 8 or coded as well, moderately or poorly differentiated tumours), while 183 men were diagnosed with high grade disease (Gleason score ≥ 8 or coded as undifferentiated tumours). Information on tumour stage was available for 62.9% of cases; 702 men had localised tumours (tumour-node-metastasis (TNM) system score of $\leq$ T2 and N0/X and M0, or stage coded as localised), 300 men had advanced tumours (TNM score of $\mathrm{T3-4}$ and/or $\mathrm{N1-3}$ and/or M1, or coded as advanced), 181 men had aggressive tumours (a subset of advanced disease defined as TNM score of T4 and/or $\mathrm{N1-3}$ and/or M1), while 821 men had non-aggressive tumours.

Matched controls: Each case was matched to one control participant, selected randomly among male cohort participants who were alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the case. Matching criteria were study centre, length of follow-up, age at blood collection ($\pm$6 months), time of day ($\pm$1 hour) and fasting status (i.e.

time since last meal; $< 3, 3-6, > 6$ hours) at blood collection. An incidence density sampling procedure was used such that a control could become a case at a later date or be a control for more than one case.

7.3.2 Blood collection and laboratory analysis

A standardised protocol for blood collection and processing was followed (see Section 3.2.1 on page 23).

For this analysis, data from follow-up phase I and parts of III (together referred to as assay round I; i.e. the data which were used in the univariate analyses; see Section 6.3 on page 119), and data from follow-up phase II (assay round II) were included.

For both assay rounds, plasma samples (with citrate as an anticoagulant) were assayed at IARC, using the Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbrcuk, Austria) and following the procedure recommended by the vendor. A triple quadrupole mass spectrometer (Triple Quad 4500; AB Sciex, Framingham, Massachusetts, USA) was used to quantify a total of 142 and 144 metabolites for assay round I and II, respectively; amino acids ($n = 22$ for both rounds) and biogenic amines ($n = 11$ for both rounds) were separated by ultra-high performance liquid chromatography before injection into the mass spectrometer, whereas flow injection analysis was used for acylcarnitines ($n = 18$ and 19 for assay rounds I and II, respectively), glycerophospholipids ($n = 78$ for both rounds), hexose ($n = 1$ for both rounds) and sphingolipids ($n = 12$ and 13 for assay rounds I and II, respectively).

Samples from matched case-control sets were assayed in the same assay round and analytic batch, each of which included quality control samples of pooled plasma (6 to 11 per batch and a total of 255 for assay round I and 4 to 8 per batch and a total of 122 for assay round II). Laboratory personnel were blinded to sample category, i.e. case, study control or quality control.

7.3.3 Exclusions of metabolites and men

The following exclusion criteria were applied for each metabolite: more than 20% non-quantifiable assay results (missing, below the limit of detection (LOD), below limit of quantification (LOQ) or above the highest standard) and coefficient of variation (CV) higher than 20%. The exclusion criterion for non-quantifiable results was raised from 15% in the univariate analysis (described in Chapter 6) because for 17.2% of men ($n = 212$; all men in three batches) no data were provided by the laboratory for 11 to 13 glycerophospholipids due to high readings of internal quality control samples (Personal communication David Achaintre, IARC). Furthermore, men with missing information on any of the included metabolites were excluded. All exclusion criteria were applied separately to each assay round.

7.3.3.1 Assay round I

Of the 142 metabolites measured in assay round I, 17 metabolites had non-quantifiable measurements for more than 20% of men ($n > 433$ men) and were excluded (Table 6.1 on page 135 and Figure 7.1 on page 263). Only one metabolite (acylcarnitine C18) had non-quantifiable measurements for 15 to 20% (15.13%) of men; this metabolite was thus included in the present analysis as opposed to the univariate analysis. Two further metabolites were excluded as the CVs were greater than 20%. This left 123 metabolites. All 1,077 matched sets from the univariate analysis were included.

7.3.3.2 Assay round II

No data for an entire batch ($n = 70$ men) were provided by the laboratory due to technical issues. This batch was removed and was not considered any further (Figure 7.1 on page 263). 15 metabolites (eight acylcarnitines, one amino acid, five biogenic amines and one glycerophospholipid) had non-quantifiable measurements for more than 20% of men ($n > 250$ men) and two additional metabolites had CVs > 20% (Table 7.1 on page 230). These were excluded, leaving 127 of the 144 metabolites.

After exclusion of the batch with no data, data for 1,246 samples were available. Of these, one case had missing data on all amino acids and biogenic amines and was thus excluded along with his matched control. Furthermore, the 212 men from the three batches affected by analytical issues were excluded, leaving 1,032 men corresponding to 512 matched sets for analysis.

7.3.3.3 Combining the assay rounds

Two and six metabolites were measured only in assay round I and II, respectively, so for 121 metabolites data were available from both rounds and these were thus available for statistical analyses in this chapter; these were eight acylcarnitines, 21 amino acids, five biogenic amines, 74 glycerophospholipids, hexose and 12 sphingolipids.

From the two assay rounds, a total of 1,593 matched case-control sets were included in the analyses.

7.3.4 Statistical analysis

7.3.4.1 Data preparation

Any values below LOQ and higher than the highest standard were set to half the LOQ and to the highest standard, respectively (Table 7.1 on page 230). Values below LOD in assay round I were set to half the lowest measured value as values for the LODs were not available, while in assay round II values below LOD were set to half the mean LOD across batches.

All metabolite concentrations were logarithmically transformed to approximate the normal distribution.

7.3.4.2 Principal component analysis

The PCA was computed in control participants using ln-transformed metabolite concentrations and the covariance matrix (see Section 5.3.5.5 on page 60 for details of PCA). The number of components to retain was based on the scree plot and an *a priori* aim to explain approximately 60% of the total variation.

7.3.4.3 Treelet transform

Treelet transform was recently introduced into epidemiology [160]. Like PCA, TT describes patterns in multi-dimensional data. In contrast to PCA, TT produces more sparse components in which only highly correlated metabolites contribute; while this results in less variance in metabolite concentrations being explained than with PCA, the interpretation of patterns from TT is often easier [160, 202].

In TT, the two most closely correlated metabolites are joined together by local PCA and the two original variables are replaced with the first principal component [160, 202]. This procedure is repeated until all variables are joined into one. A cluster tree is used to visualise how metabolites are joined such that highly correlated metabolites are joined near the leaves of the tree (Figure 7.2 on page 264). In the current study, TT was carried out in control participants only and based on the covariance matrix using logarithmically transformed metabolite concentrations, as for PCA.

To obtain the components from TT, the cluster tree must be cut. The closer the cut is to the root of the cluster tree (where all original variables are joined), the more variance is explained by the components but less sparsity is obtained [160, 202]. For the retention of a given number of components, the optimal cut-level was determined by cross-validation. A cross-validation score for each cut-level is calculated as the mean variance explained by retained factors in 10 repeats. On a plot of cross validation score against cut-level, the optimal cut-level is found by the TT algorithm where the curve flattens (Figure 7.3 on page 265).

To inform the choice of number of components to retain and the corresponding cut-level, we ran TT for different numbers of retained components (two to 12), identified the optimal cut-level for each (ranging from 105 to 112) and inspected scree plots (see Figure 7.4 on page 265 for the final scree plot) [202]. The final TT was then computed specifying the number of components to retain and the corresponding optimal cut-level [160, 202]. A sensitivity analysis of the cut-level was performed by identifying treelet components at ± 3 cut-levels [160].

The stability of components was assessed by examining the sign patterns of component loadings in 100 repeat random sub-samples, comprising 80% of the data [202]. The sign pattern is determined by the sign of the loading (i.e. the contribution of the original metabolites to a given component [160]). For example the sign pattern for a component with loadings '$0.1, 0.0, 0.3, \ldots, -0, 2$' is '$+, 0, +, \ldots, -$'). The more frequent a sign pattern is within the 100 repeats, the more stable the corresponding component is (Tables 7.2 on page 233 and 7.3 on page 234).

Finally, for each component the participants are each assigned a score. The score is a linear combination of participant's concentration of the metabolites weighted by the loadings and reflects how similar a participant's metabolite profile is to the metabolite pattern described by the component [160, 202]. Unlike principal component scores which are uncorrelated, treelet components scores have some degree of inter-correlation and thus Pearson's correlation coefficients between treelet component scores were calculated.

7.3.4.4 Comparison of components from PCA and treelet transform

The components derived from PCA and TT were compared by calculating Pearson's correlation coefficients between the component scores.

7.3.4.5 Risk analysis

Both for PCA and TT, conditional logistic regression was used to estimate the risk of overall prostate cancer by fifths of each component score. Tests for linear trend were computed across the medians of the fifths by coding each fifth with its median value and modelling this as a continuous variable. This method was chosen over a linear model using all data points because not all components were normally distributed.

A minimally-adjusted model, conditioned on the matching variables and adjusted for exact age at blood collection (continuously), and a multivariable-adjusted model, additionally adjusting for body mass index (BMI; fourths of the distribution; unknown), smoking status (never; past; current; unknown), alcohol intake ($< 10; 10-19; 20-39; \geq 40$ g/d; unknown), educational attainment (primary or none; secondary; degree level; unknown) and marital status (not married or cohabiting; married or cohabiting; unknown), were fitted. The results of the minimally- and multivariable-adjusted models did not materially differ and thus only results from the multivariable-adjusted model are presented.

Conditional logistic regression models similar to the model for overall prostate cancer risk were fitted for sub-groups by time to diagnosis (≤ 5 and > 5 years) and by tumour subtypes (low-intermediate and high grade, localised and advanced stage, and non-aggressive and aggressive prostate cancer), after assigning to the controls the value of their matched case for

the outcome variables, e.g. grade. However, due to the lower number of cases in the sub-group analyses, thirds rather than fifths were used. Tests for heterogeneity in the association between metabolite concentrations and prostate cancer risk by sub-groups were done using the likelihood ratio χ^2 test, which compared models with and without an interaction term between the linear trend variables (across the medians) and the outcome variable of interest, e.g. grade.

We also assessed the risk of death from prostate cancer using the subset of cases who were known to have subsequently died of prostate cancer during follow-up (prostate cancer listed as the underlying cause of death on the death certificate; $n = 194$). The original matching was preserved, and matched sets, in which the control had died before the case ($n = 18$) or where information on vital status for the control was not available ($n = 4$), were excluded, leaving 172 matched sets in this analysis.

7.3.4.6 Statistical significance

All p-values were two-sided and p-values less than 0.05 were considered to be statistically significant.

7.3.4.7 Statistical software

All analyses were conducted in Stata Statistical Package version 14.1 (Stata Corp., Texas, USA). The 'tt' add-on for STATA used for TT is available from http://www.gorst.dk. The power calculation reported in the discussion was conducted in Power V3.0, version 3.0.0 (National Cancer Institute, Maryland, USA; see Section 7.5.4.3 on page 227).

7.4 Results

7.4.1 Participant characteristics

Participants were on average 60 years of age at blood collection (range 40 to 77 years; Table 7.4 on page 237) and the mean age of diagnosis was 66 years and 9 months (range 51 to 84 years). No material differences in lifestyle characteristics or anthropometry were observed between cases and controls.

7.4.2 Principal component analysis

7.4.2.1 The principal components

The first seven principal components (PC1 to PC7) explained 60.3% of the total variance in the metabolite data and a drop in the variance explained by each component was seen after seven components (Table 7.5 on page 238 and Figure 7.5 on page 266). These seven components were

thus retained for further analyses and each explained 21.4%, 9.6%, 7.1%, 6.6%, 5.8%, 5.2% and 4.5% of the total variation in the original data, respectively.

PC1 was characterised by positive loadings on many of the glycerophospholipids and sphingolipids (Figure 7.6 on page 267 and Table 7.6 on page 239). In **PC2**, strong positive loadings were seen for some acylcarnitines, especially C18:2, amino acids glutamate and ornithine, biogenic amine taurine, and sphingolipid SM C20:2. Furthermore, strong negative loadings were observed for arginine and hexose. **PC3** was characterised by both positive and negative loadings for some glycerophospholipids, especially PC aa C32:1 (positive), and a negative and positive loading for arginine and glutamate, respectively. On **PC4**, several acylcarnitines loaded positively, especially C14:1 and C3; glycerophospholipids loaded both positively and negatively and sphingolipid SM C20:2 loaded negatively. **PC5** was characterised by negative loadings for some glycerophospholipids, mainly diacyl-phosphatidylcholines, and positive loadings on some sphingolipids, especially SM C20:2. On **PC6**, most acylcarnitines loaded negatively, especially C14:1, but not C3. Moreover, the component was characterised by positive loadings on several amino acids and the biogenic amine sarcosine. Finally, **PC7** was characterised by a strong positive loading of acylcarnitine C3 and inverse loadings for lysophosphatidylcholines.

7.4.2.2 Associations between principal components and prostate cancer risk

Overall prostate cancer: No significant associations were observed between the retained principal components and risk of overall prostate cancer (Table 7.7 on page 242).

Time to diagnosis: Three components were significantly associated with prostate cancer diagnosed within the first five years of follow-up, while no associations were observed at longer follow-up (Table 7.8 on page 243). Men with higher scores for PC3 (positive and negative loadings on some glycerophospholipids and amino acids) had a lower risk of prostate cancer ($\text{OR}_{\text{top vs bottom third}} = 0.63$; 95%CI: $0.45 - 0.89$, $p_{trend} = 0.006$) and the test for heterogeneity in the association with prostate cancer by time to diagnosis was also significant ($p_{het} = 0.004$). For PC6 (negative loadings on acylcarnitines and positive loading on a few amino acids and sarcosine) and PC7 (positive loadings on acylcarnitine C3 and negative loadings on lysophosphatidylcholines), higher risk was observed for men with higher scores ($\text{OR}_{\text{top vs bottom third}} = 1.53$; 95%CI: $1.09 - 2.15$, $p_{trend} = 0.01$ and $\text{OR}_{\text{top vs bottom third}} = 1.45$; 95%CI: $1.01 - 2.10$, $p_{trend} = 0.049$, respectively).

High grade: A lower risk of high grade prostate cancer was observed for higher scores on PC1 (positive loadings on phospholipids; $\text{OR}_{\text{top vs bottom third}} = 0.52$; 95%CI: $0.29 - 0.93$; $p_{trend} =$

0.02; Table 7.9 on page 244).

Advanced stage: Men in the top third of PC1 (positive loadings on glycerophospholipids and sphingolipids) score had a 54% lower risk of advanced stage prostate cancer than men in the bottom third ($OR_{\text{top vs bottom third}} = 0.46$; 95%CI: $0.29 - 0.71$, $p_{trend} = 0.0005$; Table 7.10 on page 245). The test for heterogeneity in the association of this component with localised (no association) and advanced prostate cancer, respectively, was statistically significant ($p_{het} = 0.0001$).

Aggressive prostate cancer No associations were observed between principal components and risk of aggressive or non-aggressive disease (Table 7.11 on page 246). However, weak heterogeneity in the association of PC7 (positive loadings on acylcarnitine C3 and negative loadings on lysophosphatidylcholines) with risk of non-aggressive (inverse) and aggressive prostate cancer (positive), respectively, was observed ($p_{het} = 0.04$).

Death from prostate cancer: Men with higher scores on PC7 (positive loadings on acylcarnitine C3 and negative loadings on lysophosphatidylcholines) had a higher risk of dying from prostate cancer ($OR_{\text{top vs bottom third}} = 4.88$; 95%CI: $2.18 - 10.89$, $p_{trend} = 0.0001$; Table 7.12 on page 247).

7.4.3 Treelet transform

7.4.3.1 Treelet transform components

Seven treelet components (TC1 to TC7) were retained for further analysis based on the scree plot and together they explained 41.7% of the total variance (Figure 7.4 on page 265 and Table 7.5 on page 238). The cross-validation showed that the optimal cut-level for this number of components was 105 (Figure 7.3 on page 265). The cluster tree shows at which level metabolites are joined according to how correlated they are, and metabolites joined before cut-level 105 belong to the same component (Figure 7.2 on page 264).

TC1 explained 19.9% of the total variance and was characterised by positive loading on the vast majority of the glycerophospholipids (mainly non-lysoPCs) and sphingolipids (Figure 7.7 on page 269 and Table 7.13 on page 248). **TC2** explained 5.7% of the total variance and was characterised by positive loadings for acylcarnitines, amino acids glutamate and ornithine, and biogenic amine taurine. **TC3**, explaining 4.3% of the total variance, included a subset of the glycerophospholipids in TC1 but was characterised mainly by negative loadings with some positive loadings. Only acylcarnitine C2 and C14:1 loaded, positively, on **TC4**, which explained 3.8% of the total variance. **TC5** explained 3.1% of the total variance and all but

one lysophosphatidylcholine loaded positively on this component. Arginine and hexose were the only metabolites contributing, with positive loadings, to **TC6**, which explained 1.7% of the total variance. **TC7** comprised a subset of metabolites in TC3 and thereby TC1. Several diacyl-phosphatidylcholines loaded negatively on this component, while a few other diacyl-phosphatidylcholines, acyl-alkyl-phosphatidylcholines and sphingolipids loaded positively. This component comprised the phosphatidylcholines with shorter fatty acids; the total fatty acid length in these ranged from 30 to 38 carbon atoms. TC7 explained 3.2% of the total variance.

Correlations between treelet components scores: Correlation coefficients (r) between treelet component scores in general were low (Table 7.14 on page 250). However, there were moderate correlations between TC1 (positive loadings on phospholipids) and TC5 (positive loadings on lysophosphatidylcholines; $r = 0.31$), and between TC2 (positive loadings on acyl-carnitines, amino acids glutamate and ornithine, and biogenic amine taurine) and TC6 (arginine and hexose; $r = -0.67$).

Stability of treelet components: In Table 7.2 on page 233, the results of the stability assessment of the treelet components are shown and in Table 7.3 on page 234 the corresponding sign patterns are displayed. In general, components derived from the sub-samples were very similar to the overall components, indicating high stability of the components.

TC1, TC4 and TC5 fully overlapped with sign patterns 1, 6 and 7, which all had a frequency of 1, indicating extremely highly stable components. TC2 fully matched sign pattern 3, and with the exception of one metabolite (acylcarnitine C18), also matched sign pattern 2, which occurred in 80% and 20% of the repeats respectively. Similarly, TC3 overlapped perfectly with sign pattern 4 and nearly with sign pattern 5 (there was a difference in the sign of PC aa C38:5); the frequencies were 0.57 and 0.23, respectively. TC6 and TC7 completely corresponded to sign patterns 8 and 10, which had a frequency of 0.88 and 0.32, respectively. Sign pattern 9 did not resemble any of the seven retained treelet components.

In summary, TC1, TC4 and TC5 were highly stable, closely followed by TC2 and TC6, while the stability was only slightly lower for TC3. In contrast, TC7 was not very stable.

Sensitivity analysis of cut-level for treelet transform: TC1 to TC5 and TC7 derived by cutting the cluster tree both at cut-level 102 and 108 (Tables 7.15 on page 251 and 7.16 on page 254, respectively) were identical or very similar to TC1 to TC2 and TC6 in the main analysis using a cut-level of 105. However, TC7 in the main analysis was not replicated at cut-levels 102 or at 108, reinforcing the limited stability of this component.

Metabolites not represented in retained treelet components: Due to the sparse nature of treelet components, several metabolites included in the study did not contribute to any of the seven retained components (Table 7.13 on page 248 and Figure 7.2 on page 264), i.e. two acylcarnitines (C0 and C3), 18 amino acids (all but arginine, glutamate and ornithine), four biogenic amines (all but taurine) and a sphingolipid (SM C20:2).

7.4.3.2 Comparison of components from PCA and TT

The first two components in both PCA and TT were very similar and scores were highly positively correlated ($r = 0.99$ for PC1/TC1 and $r = 0.85$ for PC2/TC2; Table 7.17 on page 256). Similarly, but less strongly, PC4 was positively correlated with TC4 ($r = 0.52$) but also TC5 ($r = 0.44$). Scores for PC3, PC5, PC6 and PC7 were inversely correlated with TC7, TC3, TC4 and TC5, respectively ($-0.76 \leq r \leq -0.47$), and PC3 was further inversely correlated with TC3 ($r = -0.53$).

7.4.3.3 Association between treelet components and prostate cancer risk

Overall prostate cancer: None of the seven treelet components were significantly associated with risk of overall prostate cancer (Table 7.18 on page 257).

Time to diagnosis: Analyses of risk by time to diagnosis showed that TC3 (positive and negative loadings on certain phospholipids) was associated with a higher risk of prostate cancer within five years of follow-up ($\text{OR}_{\text{top vs bottom third}} = 1.54$; 95%CI: $1.08 - 2.20$, $p_{trend} = 0.02$; Table 7.19 on page 258). At longer follow-up, no association was observed ($\text{OR}_{\text{top vs bottom third}} = 0.89$; 95%CI: $0.70 - 1.14$, $p_{trend} = 0.4$) and the test for heterogeneity in the association by time to diagnosis was significant ($p_{het} = 0.01$).

High grade: Men with higher scores on TC1 (positive loadings on phospholipids) and TC6 (positive loadings on arginine and hexose) had lower and higher risk of high grade prostate cancer, respectively (TC1: $\text{OR}_{\text{top vs bottom third}} = 0.47$; 95%CI: $0.26 - 0.86$, $p_{trend} = 0.01$; TC6: $\text{OR}_{\text{top vs bottom third}} = 2.33$; 95%CI: $1.22 - 4.43$, $p_{trend} = 0.02$; Table 7.20 on page 259). In addition, the test for heterogeneity of the association with prostate cancer by grade was significant for the latter component ($p_{het} = 0.01$).

Advanced stage: TC1 (positive loadings on phospholipids), TC2 (positive loadings on some acylcarnitines, glutamate, ornithine and taurine) and TC4 (positive loadings on acylcarnitines C2 and C14:1) were all inversely associated with risk of advanced disease; the risk in the top third of the component score was approximately 50% lower than in the bottom third (TC1:

$OR_{\text{top vs bottom third}} = 0.48$; 95%CI: $0.31 - 0.74$, $p_{trend} = 0.0009$; TC2: $OR_{\text{top vs bottom third}} = 0.52$; 95%CI: $0.30 - 0.93$, $p_{trend} = 0.04$; and TC4: $OR_{\text{top vs bottom third}} = 0.54$; 95%CI: $0.34 - 0.85$; $p_{trend} = 0.008$; Table 7.21 on page 260). Moreover, for TC1 there was significant heterogeneity in the associations with localised and advanced tumour stage ($p_{het} = 0.0002$).

Aggressive prostate cancer: No associations were observed between the retained treelet components and risk of aggressive disease, while men with higher scores on TC4 (positive loadings on acylcarnitines C2 and C14:1) had a lower risk of being diagnosed with non-aggressive disease ($OR_{\text{top vs bottom third}} = 0.74$; 95%CI: $0.56 - 0.97$; $p_{trend} = 0.03$; $p_{het} = 0.6$; Table 7.22 on page 261).

Death from prostate cancer: No significant associations were observed between treelet components and risk of death from prostate cancer (Table 7.23 on page 262) .

7.5 Discussion

7.5.1 Main findings

In this study, patterns in metabolite profiles based on the correlation structure of 121 pre-diagnostic metabolite concentrations were identified using PCA and TT. The two methods produced somewhat comparable patterns but the majority of patterns identified by TT were easier to interpret than the principal components, because of their sparsity. Furthermore, the retained treelet components except one were very stable, confirming that TT is a useful method for summarising metabolomics data [201].

The primary pattern identified by both methods, which was characterised by positive loadings on glycerophospholipids and sphingolipid (PC1 and TC1), was inversely associated with risk of high grade and advanced prostate cancer, suggesting that men with higher concentrations of these metabolites might have a lower risk of high grade and advanced prostate cancer. Furthermore, two patterns identified by TT were both inversely related to risk of advanced prostate cancer; one pattern was characterised by positive loadings on four acylcarnitines (C16, C18, C18:1 and C18:2), amino acids glutamate and ornithine, and biogenic amine taurine (TC2), and the other by the two acylcarnitines C2 and C14:1 (TC4). The latter pattern was also associated with non-aggressive disease, possibly reflecting a link to stage T3 tumours, which are included both in advanced and non-aggressive, because in the current study aggressive is the subset of advanced prostate cancer with T4 (rather than T3 and T4).

In contrast, two further patterns were related to increased risk of prostate cancer death and high grade disease, respectively. Higher risk of death from prostate cancer was observed

in men with higher scores on a pattern characterised by positive loadings on acylcarnitine C3 and negative loadings on lysophosphatidylcholines (PC7). Likewise, a higher risk of high grade disease was found in men with metabolite profiles characterised by higher concentrations of amino acid arginine and hexose (TC6).

7.5.1.1 Comparison with results from univariate analysis

The results described in the current chapter were in general compatible with findings from the analysis of individual metabolites (Chapter 6).

Firstly, there was some evidence from both the univariate and the multivariate analyses that some associations might be due to existing prostate cancer. Both citrulline alone, and several patterns identified in the present analysis (including patterns mainly characterised by acylcarnitines, amino acids, glycerophospholipids and sphingolipids; PC3, PC6, PC7 and TC3) were associated with risk only within five years of follow-up and might thus be markers of sub-clinical disease rather than being relevant in the aetiology of prostate cancer.

Secondly, the lower risks of high grade and advanced prostate cancer with higher concentrations of glycerophospholipids and sphingolipids were consistent across the univariate and both PCA and TT.

Thirdly, there was consistency in findings of an inverse relationship between acylcarnitines other than C3 and advanced prostate cancer. Higher concentrations of four individual acylcarnitines (C0, C14:1, C18:1 and C18:2) were associated with lower risk of advanced stage disease at the conventional significance level ($p < 0.05$), and so were two patterns characterised by higher concentrations of four acylcarnitines (C16, C18, C18:1 and C18:2) plus a few additional metabolites (TT2), and by two acylcarnitines only (C2 and C14:1; TT4), respectively.

Finally, findings in the current multivariate analysis, that a pattern characterised by high concentrations of C3 and low concentrations of lysophosphatidylcholine (PC7) was related to higher risk of death from prostate cancer, also correspond to the results of the univariate analysis. In the univariate analysis, higher concentration of C3 was associated with higher risk of aggressive disease and death from prostate cancer, while lower concentrations of three lysophosphatidylcholines were associated with higher risk of aggressive disease, all at a conventional significance level but not after multiple testing.

In contrast to the findings of the univariate analysis, at a conventional statistical significance level, no associations were found with overall risk of prostate cancer in the current analysis.

In summary, findings from the univariate and multivariate analyses were mostly consistent but the multivariate analyses drew attention to groups of metabolites, such as the acylcarnitines, lysophosphatidylcholines and arginine, the results for which were more pronounced when they were part of a pattern rather than when they were analysed as individual metabolites.

7.5.2 Other studies

Comparing results from metabolomics studies is not straightforward due to the use of different laboratory assays, the use of different biological media (e.g. serum or plasma [166]), and the use of different statistical methods. Furthermore, features general to epidemiological studies might also complicate comparison of results, e.g. different population characteristics and for prostate cancer in particular, heterogeneity in tumour sub-types (case mix) and in prostate cancer detection methods (based on PSA test or clinical symptoms). However, to get a fuller understanding of the relationship between metabolites and risk, replication of results is essential.

Few other prospective studies have been published on metabolite concentrations and risk of prostate cancer [28, 29, 82, 83, 84], and only two of these have grouped metabolites by chemical class, pathway and/or correlation structure [28, 29]. For details of the designs in these two studies please see Section 6.5.2 on page 129.

A study from the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) investigated the association between chemical classes of metabolites (e.g. energy metabolites, lipids, amino acids, nucleotides, and co-factors and vitamins) and risk of prostate cancer [28]. Two of these classes, namely energy metabolites (involved in the Krebs cycle) and lipid metabolites (including glycerophospholipids) were both inversely associated with risk of aggressive prostate cancer (defined by combining stage and grade information), while no association with any class of metabolites was observed for overall and non-aggressive prostate cancer. The finding for the lipid class is in line with the inverse associations reported here between metabolite patters characterised by high concentrations of glycerophospholipids and sphingolipids (PC1 and TC1) and risk of high grade and advanced stage tumours. In accordance with the ATBC analysis, the present analysis did not find associations with these patterns and risk of overall, low-intermediate grade or localised prostate caner. Metabolites involved in the Krebs cycle were not measured in the current study. Finally, our finding of a positive association between a pattern characterised by arginine and hexose, and risk of high grade prostate cancer is consistent with the positive association between arginine and risk of aggressive prostate cancer (defined by combining information on stage and grade; in univariate analyses) reported from ATBC [28].

Analyses in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) investigating metabolites grouped by biological pathway found inverse associations of metabolites in the branched-chain fatty acid, medium-chain fatty acid and tocopherol pathways with risk of overall prostate cancer [29]. Moreover, metabolites in the branched-chain fatty acid pathway were positively associated with risk of aggressive disease, while metabolites involved in the urea cycle (including arginine, citrulline and ornithine) and metabolism of arginine, proline and tryptophan were inversely associated with aggressive disease. No associations with risk of

prostate cancer were observed in the analysis grouping metabolites by chemical class or with the first ten principal components, the structure of which were not described in the paper.

Results for lipids, and arginine and related metabolites, in relation to risk of aggressive disease were in opposite directions in PLCO (positively and inversely, respectively) [29] compared to ATBC (inversely and positively, respectively) [28]. Our results were mostly in agreement with the findings in ATBC, although comparison is complex due to the differences in measured lipid species and the differences in definitions of tumour sub-types. These differences in results between PLCO on one side, and ATBC and EPIC on the other may be due to screening of the entire PLCO population and thus difference in case-mix [29], as screening was not common in ATBC [28] or EPIC.

7.5.3 Potential mechanisms

Little is known about mechanisms potentially underlying the associations between metabolites and prostate cancer risk reported here. It is however likely that some of the associations observed reflect alterations in metabolite profiles driven by existing tumours while other associations may be due to effects of metabolites and related pathways on the development of the tumour. Evidence from basic science might help develop our understanding of such mechanisms.

7.5.3.1 Glycerophospholipids and sphingolipids

Altered circulating lipid concentrations in men who are subsequently diagnosed with cancer may be due to altered lipid uptake and metabolism by rapidly proliferating cancer cells in sub-clinical tumours; these metabolites are required for membrane synthesis and lipid-based cell signalling [178, 192].

Sphingomyelins are synthesised from phosphatidylcholines [178] (the types of sphingolipids and glycerophospholipids, respectively, investigated here), which might explain why these two metabolite classes cluster together in patterns. For details on phosphatidylcholines and sphingomyelins in relation to cancer, please refer to Section 6.5.3.4 on page 132.

Like other phosphatidylcholines, lysophosphatidylcholines are involved in both pro-cancerous processes and apoptosis [177], and in line with our results, decreased concentrations of lysophosphatidylcholines have been observed in patients with lung and liver cancer [191].

7.5.3.2 Acylcarnitines

Acylcarnitines are composed of carnitine (acylcarnitine C0) and fatty acids, and the main function of these compounds is to facilitate transport of fatty acids into the mitochondria for breakdown [208].

Evidence from cell and animal studies of colon cancer supports a possible role of acylcarnitines in cancer development. Carnitine (C0) has been shown to have apoptotic and anti-proliferative effects on colon cancer cells [209]. Furthermore, supplementation of mice with carnitine (C0) and other acylcarnitines has been reported to limit tumour formation and progression from pre-cancerous lesions to macroscopic tumours [210, 211].

Case-control studies in humans have found lower concentrations of some acylcarnitines in patients with lung [212] and bladder [213] cancer compared to controls. This is in line with the inverse association of acylcarnitines other than C3 with risk, reported here.

The reason for the apparent opposite directions of the association of acylcarnitine C3 (propionylcarnitine; positive) and other acylcarnitines (inverse) with prostate cancer observed here requires further investigation.

7.5.3.3 Arginine

In contrast to the current results, two studies have reported lower concentrations of arginine in plasma, serum and urine of prostate cancer patients than healthy controls [214, 215], and an inverse association between arginine and insulin-like growth factor-I (IGF-I) has been observed in women [78]. However, in Chapter 9, a positive association between arginine and IGF-I concentration is reported, which might partly explain the positive association between arginine and prostate cancer risk observed in the current analysis. Furthermore, arginine might be involved in maintaining malignancies, including prostate cancer, and arginine deprivation has been used in treatment of prostate cancer [216, 217].

7.5.3.4 Hexose

A positive association for hexose is not in line with findings from other prospective studies in which higher fructose intake [218], higher blood glucose concentrations [219, 220] and diabetes [54] were all related to a lower risk of developing prostate cancer. Profound changes in glucose metabolism of cancer cells, however, are well known [221]. Normal prostate cells depend on glycolysis to survive and perform their main function, namely to excrete citrate, as the Krebs cycle is inactive (in contrast to most other normal cells in which citrate is a substrate in the Krebs cycle). In contrast to normal prostate cells, in prostate cancer cells the Krebs cycle is active and citrate is consumed along with glucose and lactate. A further adaption of glucose metabolism is seen in metastatic prostate cancer cells in which the Krebs cycle is again inactivated, the glycolytic rate is increased and lactate is accumulated; this is known as the Warburg effect.

7.5.4 Strengths and limitations

7.5.4.1 PCA and TT

Strengths of PCA and TT include the ability to summarise high-dimensional data like metabolomics data and thereby reduce issues of multiple testing and bring out patterns in data which may be overlooked in univariate analyses.

On the other hand, there are limitations to these methods. Both methods produce mathematically derived patterns, which might be assumed to reflect metabolic pathways due to the tight correlations within patterns [201], but the biological relevance of such patterns is not clear [160]. However, the grouping of glycerophospholipids and sphingolipids in the first PC and TT, for example, might reflect a biological pathway as sphingomyelins are synthesised from phosphatidylcholines [178]. Furthermore, both methods construct patterns which explain variance in the metabolomics data without taking into account the outcome of interest, e.g. prostate cancer. Such patterns might not necessarily be the most relevant for the outcome.

So-called supervised methods, such as the partial least squares method which maximises the covariance between the metabolites and the outcome [158], or pathway analyses might help identify patterns potentially more relevant to risk [11]. Components explaining less variance than the retained components could potentially also be related to risk of prostate cancer, but for both methods a limited number of components have to be selected for dimension reduction purposes.

Nonetheless, we did find associations with risk of prostate cancer among the first seven patterns, which were derived mathematically without regards to case-control status. This suggests that PCA and TT are suitable for investigations of metabolites and risk of prostate cancer.

Strengths and limitations of using covariance matrix of ln-transformed metabolite concentrations are described in Section 5.3.5.5 on page 60.

7.5.4.2 PCA versus TT

PCA has been widely used to describe dietary patterns in studies of diet and disease outcomes [203] and more recently, it has also been used to identify metabolite patterns in prostate cancer cases and controls [11]. Although not as widely used as PCA, TT has been useful in identifying patterns in serum metabolite and lipid profiles [201], in adipose tissue fatty acids [204, 205], in dietary intake [160, 206, 207] and in co-morbidity of traumatic brain injuries [222]. Furthermore, associations of these patterns with various outcomes, including anthropometric measures [201], and risk of myocardial infarction [160], breast cancer [206] and prostate cancer [204] have been reported.

The main strength of TT over PCA is the easier interpretation of patterns owing to the

sparse components generated by the TT algorithm. The treelet components which were related to risk of prostate cancer in our analysis were clear and interpretable; for example, a clear pattern including specific acylcarnitines (TC4) was observed. In contrast, all metabolites contribute to each of the principal components and subjective judgement therefore has to guide which loadings to emphasise. Rotation of principal components may be used to improve interpretability of PCA results [156]. However, rotation is based on several subjective criteria, which can severely affect the resulting components. Thus, rotation was not used in the present analysis and TT provided a more objective way of improving interpretation.

The trade off for having more sparse components in TT than PCA is that TT explained less of the total variance in the original metabolomics data than PCA, which explains the maximum variation possible [156]. In the current analysis, 42% and 60% of the total variance was explained by the seven components for TT and PCA, respectively).

The main limitation of TT is that a cut-level has to be chosen to generate components, and this might affect component composition, while no such choice has to be made in PCA. However, the choice of cut-level for this TT analysis was data-driven using cross-validation, and components generated at ±3 cut-levels produced similar components to those of the main analysis, except for TC7 (characterised by both positive and negative loadings on phospholipids).

7.5.4.3 Other strength and limitations

Strengths and limitations related to parts of the study other than the statistical analysis are the same as for the analysis of individual metabolite concentrations and are thus only mentioned briefly here (see Section 6.5.4 on page 132 for more details).

The main strengths are the large study sample, facilitating more robust analyses than the previous published studies, and the well-characterised cohort, allowing multivariable-adjustment and thus minimising confounding. With a total of 1,593 cases, we had more than 80% power to detect an OR of 1.33 or 0.75 in the top fifth compared to the bottom fifth. In the sub-group analyses, we had more that 80% power to detect an OR of 1.77 or 0.56 in the top versus bottom third for advanced stage prostate cancer ($n_{cases} = 300$), while the corresponding numbers were 2.12 or 0.47 for high grade and aggressive disease, and death from prostate cancer ($n_{cases} = 172$ to 183).

The choice of a targeted metabolomics assay limits the identification of metabolite patterns and thereby the extent of our search for potential risk factors for prostate cancer to the covered metabolites. A further limitation is the fact that only one blood sample per participant was available. While this might lead to some attenuation of risk estimates, studies of long-term reproducibility of metabolite concentrations have suggested that a single measurement may be adequate for most metabolites [126, 146, 193]. Moreover, pre-analytical factors, such as use

of citrate as anticoagulant, of non-fasting blood samples, of some samples that had not been immediately cooled (EPIC-Oxford samples only) and of samples that had been thawed and refrozen, might potentially have attenuated our risk estimates. However, to minimise this, citrate was used in all samples, cases and controls were matched on fasting status and on study centre, and samples from matched sets had gone through the same number of freeze-thaw cycles.

7.5.5 Further work

The patterns in metabolite profile and their relation to prostate cancer risk, described here, require replication in other populations. In deriving metabolite patterns, the association with prostate cancer could be taken into account using supervised methods, such as the partial least squares method [158]. Moreover, analyses focusing on amino acids might further advance the field; amino acids were not well-represented by the retained components in the current analysis but there is an *a priori* hypothesis about their relationship with prostate cancer risk (see Section 2.2.6 on page 15).

Future studies should aim to develop our understanding of the temporality and thereby the causality, of the associations between metabolites and prostate cancer risk. Unlike previous smaller studies [28, 29], we have not investigated the association between metabolites and risk of tumour sub-types by time to diagnosis; we assessed that the number of cases within each sub-group were still inadequate for such an analysis. However, the robustness of such an analysis will increase when metabolite data from more cases and controls become available. Data from cases diagnosed during the latest follow-up phase of EPIC are expected in 2018. Moreover, research utilising genetic instruments for metabolite concentrations [223] may provide insight into possible causal relationships between specific metabolites and prostate cancer risk (see Section 10.3.3 on page 349).

Further experimental studies in animals and *in vitro* could also help develop our understanding of mechanisms underlying the potential associations described here.

7.6 Conclusion

In conclusion, we found evidence that patterns in pre-diagnostic plasma metabolite profile, identified using PCA and TT, are associated with risk of more aggressive sub-types of prostate cancer (high grade, advanced and aggressive disease) and death from prostate cancer. In contrast, there was little evidence of associations with risk overall or with risk of less aggressive prostate tumours (low-intermediate grade, localised stage and non-aggressive disease). Our results suggest that men with metabolite profiles characterised by higher concentrations of glycerophospholipids and sphingomyelins, and by acylcarnitines, certain amino acids and a biogenic amine might have a

lower risk of high grade and/or advanced stage prostate cancer. Furthermore, the results indicate that men with higher concentration of acylcarnitine C3 in combination with lower concentrations of lysophosphatidylcholines might have a higher risk of prostate cancer death, and that men with higher concentrations of the amino acid arginine and hexose might have a higher risk of high grade disease.

Although this study is the largest to date, that we are aware of, replication in larger populations will help provide more robust results and help us understand if these patterns might reflect sub-clinical markers of high-risk prostate cancer or whether they might play a role in aetiology.

7.7 Tables

Table 7.1: Completeness and coefficients of variation for the metabolomics assay and corresponding exclusions of metabolites for assay round II[1]

Metabolites	Non-quantifiable results[2]				CV[3], %
	Missing, %	Outside measurable range			
		Samples, %	Type (conc, µmol/L)	Total, %	
Acylcarnitines					
C0	0	0		0.00	5.43
C10	0	83.07	<LOD (0.411)	**83.07**	-
C10:1	0	27.77	<LOD (0.103)	**27.77**	6.98
C12	0	69.18	<LOD (0.129)	**69.18**	-
C12:1	0	5.54	<LOD (0.103)	5.54	13.02
C14	0	78.57	<LOD (0.038)	**78.57**	-
C14:1	0	0.40	<LOD (0.016)	0.40	12.14
C14:2	0	50.32	<LOD (0.022)	**50.32**	3.01
C16	0	0		0.00	8.60
C16:1	0	85.39	<LOD (0.040)	**85.39**	-
C18	0	1.36	<LOD (0.020)	1.36	9.87
C18:1	0	0		0.00	7.95
C18:2	0	0.16	<LOD (0.015)	0.16	10.38
C2	0	0		0.00	5.19
C3	0	0		0.00	7.78
C4	0	0.08	<LOD (0.079)	0.08	8.24
C3-DC (C4-OH)	0	75.60	<LOD (0.049)	**75.60**	13.86
C4:1	0	85.47	<LOD (0.056)	**85.47**	13.15
C5	0	3.77	<LOD (0.097)	3.77	6.56
Amino acids					
Alanine	0.08	0		0.08	7.41
Arginine	0.08	0.08	<LOQ (5)	0.16	8.28
Asparagine	0.08	0		0.08	7.05
Aspartate	0.08	32.50	<LOQ (5)	**32.58**	18.69
Citrulline	0.08	0		0.08	13.65
Glutamine	0.08	0		0.08	9.58
Glutamate	0.08	0		0.08	10.23
Glycine	0.08	0		0.08	12.42
Histidine	0.08	0		0.08	7.00
Isoleucine	0.08	0		0.08	8.99
Leucine	0.08	0		0.08	7.99
Lysine	0.08	0		0.08	13.90
Methionine	0.08	0		0.08	12.38
Ornithine	0.08	0		0.08	14.45
Phenylalanine	0.08	0.08	>highest std. (400)	0.16	8.33
Proline	0.08	0		0.08	8.68
Serine	0.08	0		0.08	8.69
t4-OH-Pro	0.08	0		0.08	11.43
Threonine	0.08	0		0.08	7.67
Tryptophan	0.08	0		0.08	8.01
Tyrosine	0.08	0		0.08	9.82
Valine	0.08	0		0.08	11.41
Biogenic amines					
alpha-AAA	0.08	67.82	<LOQ (1)	**67.90**	11.63
Creatinine	0.08	0		0.08	4.80
Kynurenine	0.08	0.08	<LOQ (1)	0.16	13.03

Continues...

Table 7.1 continued

Metabolites	Non-quantifiable results[2]				CV[3], %
	Missing, %	Outside measurable range			
		Samples, %	Type (conc, µmol/L)	Total, %	
Putrescine	0.08	67.42	<LOQ (0.1)	**67.50**	-
Sarcosine	0.08	18.06	<LOQ (1)	18.14	11.92
Serotonin	0.08	86.04	<LOQ (0.1)	**86.12**	-
Spermidine	0.08	92.38	<LOQ (0.25)	**92.46**	-
Spermine	0.08	95.99	<LOQ (0.25)	**96.07**	-
Taurine	0.08	0		0.08	6.80
ADMA	0.08	3.85	<LOQ (0.25)	3.93	15.76
SDMA	0.08	0		0.08	**20.52**
Glycerophospholipids					
lysoPC a C16:0	0	0		0	6.03
lysoPC a C16:1	0	0		0	6.59
lysoPC a C17:0	0	0		0	6.76
lysoPC a C18:0	0	0		0	6.24
lysoPC a C18:1	0	0		0	6.27
lysoPC a C18:2	0	0		0	6.43
lysoPC a C20:3	0	0		0	10.67
lysoPC a C20:4	0	0		0	6.56
lysoPC a C28:1	17.17	0		17.17	11.36
PC aa C28:1	0	0		0	10.05
PC aa C30:0	0	0		0	7.78
PC aa C32:0	0	0		0	7.20
PC aa C32:1	0	0		0	10.62
PC aa C32:2	0	0		0	15.71
PC aa C32:3	0	0		0	11.29
PC aa C34:1	0	0		0	6.97
PC aa C34:2	0	0		0	7.07
PC aa C34:3	0	0		0	6.42
PC aa C34:4	0	0		0	8.22
PC aa C36:0	0	0		0	18.56
PC aa C36:1	0	0		0	8.19
PC aa C36:2	0	0		0	7.41
PC aa C36:3	0	0		0	6.74
PC aa C36:4	0	0		0	5.52
PC aa C36:5	0	0		0	6.28
PC aa C36:6	0	0		0	16.59
PC aa C38:0	0	0		0	11.12
PC aa C38:3	0	0		0	6.40
PC aa C38:4	0	0		0	6.06
PC aa C38:5	0	0		0	6.16
PC aa C38:6	0	0		0	6.16
PC aa C40:1	17.17	63.00	<LOD (0.395)	**80.18**	5.28
PC aa C40:2	17.17	0		17.17	6.49
PC aa C40:3	17.17	0		17.17	7.10
PC aa C40:4	0	0		0.00	6.89
PC aa C40:5	0	0		0.00	6.48
PC aa C40:6	0	0		0.00	5.95
PC aa C42:0	0	0		0.00	11.11
PC aa C42:1	0	0		0.00	11.30
PC aa C42:2	0	0.40	<LOD (0.058)	0.40	10.49
PC aa C42:4	17.17	0		17.17	7.45
PC aa C42:5	17.17	0		17.17	9.85
PC aa C42:6	17.17	1.44	<LOD (0.212)	18.62	8.42
PC ae C30:0	17.17	0.40	<LOD (0.113)	17.58	9.15

Continues...

Table 7.1 continued

Metabolites	Non-quantifiable results[2]				CV[3], %
	Missing, %	Outside measurable range			
		Samples, %	Type (conc, µmol/L)	Total, %	
PC ae C30:2	17.17	0		17.17	8.83
PC ae C32:1	0	0		0.00	8.14
PC ae C32:2	0	0		0.00	9.48
PC ae C34:0	0	0		0.00	10.18
PC ae C34:1	0	0		0.00	7.87
PC ae C34:2	0	0		0.00	6.67
PC ae C34:3	0	0		0.00	6.34
PC ae C36:2	0	0		0.00	6.06
PC ae C36:3	0	0		0.00	6.69
PC ae C36:4	0	0		0.00	6.36
PC ae C36:5	0	0		0.00	6.13
PC ae C38:0	0	0		0.00	**20.23**
PC ae C38:2	0	0		0.00	8.78
PC ae C38:3	0	0		0.00	6.40
PC ae C38:4	0	0		0.00	6.42
PC ae C38:5	0	0		0.00	6.94
PC ae C38:6	0	0		0.00	6.06
PC ae C40:1	0	0		0.00	10.24
PC ae C40:2	0	0		0.00	8.94
PC ae C40:3	17.17	0		17.17	5.15
PC ae C40:4	0	0		0.00	6.17
PC ae C40:5	0	0		0.00	5.84
PC ae C40:6	0	0		0.00	6.30
PC ae C42:1	11.40	0.08	<LOD (0.159)	11.48	8.10
PC ae C42:2	0	0		0.00	8.39
PC ae C42:3	0	0		0.00	7.39
PC ae C42:4	0	0		0.00	6.91
PC ae C42:5	0	0		0.00	5.55
PC ae C44:3	11.40	0.88	<LOD (0.025)	12.28	13.09
PC ae C44:4	0	0		0.00	12.50
PC ae C44:5	0	0		0.00	7.85
PC ae C44:6	0	0		0.00	7.37
Hexose		0		0.00	0.00
Hexose	0	2.49	<LOQ (1500)	2.49	4.86
Sphingomyelins		0		0.00	0.00
SM (OH) C14:1	0	0		0.00	6.38
SM (OH) C22:2	0	0		0.00	9.32
SM (OH) C24:1	0	0		0.00	18.87
SM C16:0	0	0		0.00	6.45
SM C16:1	0	0		0.00	6.12
SM C18:0	0	0		0.00	6.40
SM C18:1	0	0		0.00	7.55
SM C20:2	0	0		0.00	11.56
SM C24:0	0	0		0.00	8.02
SM C24:1	0	0		0.00	7.87
SM C26:1	0	0		0.00	17.90

[1] Metabolites marked in grey were excluded from the statistical analyses because $>$ 20% of samples had non-quantifiable assay results or the coefficient of variation (CV) was $>$ 20%. The reason for exclusion is marked in bold

[2] Limit of detection (LOD) is applicable to acylcarnitines, glycerophospholipids, sphingolipids; limit of quantification (LOQ) is applicable to amino acids, biogenic amines and hexose

[3] CV could not be calculated for all metabolites due to too many quality control samples having assay results outside the measurable range

Table 7.2: Stability assessment for treelet transform[1]

Sign pattern	Frequency	Avg. rank	Avg. variance
1	1.00	1.00	2.26
2	0.20	2.00	0.60
3	0.80	2.00	0.65
4	0.57	3.00	0.49
5	0.23	3.00	0.51
6	1.00	4.01	0.43
7	1.00	5.05	0.40
8	0.88	6.19	0.38
9	0.41	6.71	0.37
10	0.32	6.78	0.37

[1] Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites retaining seven components at cut-level 105. Analysis was repeated 100 times in random sub-samples in 80% of the data

Table 7.3: Sign patterns of treelet components for stability assessment[1]

Metabolites	Sign patterns									
	1	2	3	4	5	6	7	8	9	10
Acylcarnitines										
C0	0	0	0	0	0	0	0	0	0	0
C14:1	0	0	0	0	0	+	0	0	0	0
C16	0	+	+	0	0	0	0	0	0	0
C18	0	0	+	0	0	0	0	0	0	0
C18:1	0	+	+	0	0	0	0	0	0	0
C18:2	0	+	+	0	0	0	0	0	0	0
C2	0	0	0	0	0	+	0	0	0	0
C3	0	0	0	0	0	0	0	0	+	0
Amino acids										
Alanine	0	0	0	0	0	0	0	0	0	0
Arginine	0	0	0	0	0	0	0	+	0	0
Asparagine	0	0	0	0	0	0	0	0	0	0
Citrulline	0	0	0	0	0	0	0	0	0	0
Glutamine	0	0	0	0	0	0	0	0	0	0
Glutamate	0	+	+	0	0	0	0	0	0	0
Glycine	0	0	0	0	0	0	0	0	0	0
Histidine	0	0	0	0	0	0	0	0	0	0
Isoleucine	0	0	0	0	0	0	0	0	0	0
Leucine	0	0	0	0	0	0	0	0	0	0
Lysine	0	0	0	0	0	0	0	0	0	0
Methionine	0	0	0	0	0	0	0	0	0	0
Ornithine	0	+	+	0	0	0	0	0	0	0
Phenylalanine	0	0	0	0	0	0	0	0	0	0
Proline	0	0	0	0	0	0	0	0	0	0
Serine	0	0	0	0	0	0	0	0	0	0
t4-OH-Pro	0	0	0	0	0	0	0	0	0	0
Threonine	0	0	0	0	0	0	0	0	0	0
Tryptophan	0	0	0	0	0	0	0	0	0	0
Tyrosine	0	0	0	0	0	0	0	0	0	0
Valine	0	0	0	0	0	0	0	0	0	0
Biogenic amines										
Creatinine	0	0	0	0	0	0	0	0	0	0
Kynurenine	0	0	0	0	0	0	0	0	0	0
Sarcosine	0	0	0	0	0	0	0	0	0	0
Taurine	0	+	+	0	0	0	0	0	0	0
ADMA	0	0	0	0	0	0	0	0	0	0
Glycerophospholipids										
lysoPC a C16:0	0	0	0	0	0	0	+	0	0	0
lysoPC a C16:1	+	0	0	-	-	0	0	0	0	-
lysoPC a C17:0	0	0	0	0	0	0	+	0	0	0
lysoPC a C18:0	0	0	0	0	0	0	+	0	0	0
lysoPC a C18:1	0	0	0	0	0	0	+	0	0	0
lysoPC a C18:2	0	0	0	0	0	0	+	0	0	0
lysoPC a C20:3	0	0	0	0	0	0	+	0	0	0
lysoPC a C20:4	0	0	0	0	0	0	+	0	0	0
PC aa C28:1	+	0	0	-	-	0	0	0	0	+
PC aa C30:0	+	0	0	-	-	0	0	0	0	-

Continues...

Table 7.3 continued

Metabolites	Sign patterns									
	1	2	3	4	5	6	7	8	9	10
PC aa C32:0	+	0	0	-	-	0	0	0	0	-
PC aa C32:1	+	0	0	-	-	0	0	0	0	-
PC aa C32:2	+	0	0	-	-	0	0	0	0	-
PC aa C32:3	+	0	0	-	-	0	0	0	0	+
PC aa C34:1	+	0	0	-	-	0	0	0	0	-
PC aa C34:2	+	0	0	-	-	0	0	0	0	+
PC aa C34:3	+	0	0	-	-	0	0	0	0	-
PC aa C34:4	+	0	0	-	-	0	0	0	0	-
PC aa C36:0	+	0	0	+	+	0	0	0	0	0
PC aa C36:1	+	0	0	-	-	0	0	0	0	-
PC aa C36:2	+	0	0	-	-	0	0	0	0	+
PC aa C36:3	+	0	0	-	-	0	0	0	0	-
PC aa C36:4	+	0	0	0	0	0	0	0	0	0
PC aa C36:5	+	0	0	+	+	0	0	0	0	0
PC aa C36:6	+	0	0	+	+	0	0	0	0	0
PC aa C38:0	+	0	0	+	+	0	0	0	0	0
PC aa C38:3	+	0	0	-	-	0	0	0	0	-
PC aa C38:4	+	0	0	0	0	0	0	0	0	0
PC aa C38:5	+	0	0	-	+	0	0	0	0	-
PC aa C38:6	+	0	0	+	+	0	0	0	0	0
PC aa C40:2	+	0	0	+	+	0	0	0	0	0
PC aa C40:3	+	0	0	+	+	0	0	0	0	0
PC aa C40:4	+	0	0	0	0	0	0	0	0	0
PC aa C40:5	+	0	0	-	-	0	0	0	0	-
PC aa C40:6	+	0	0	+	+	0	0	0	0	0
PC aa C42:0	+	0	0	0	0	0	0	0	0	0
PC aa C42:1	+	0	0	0	0	0	0	0	0	0
PC aa C42:2	+	0	0	+	+	0	0	0	0	0
PC aa C42:4	+	0	0	0	0	0	0	0	0	0
PC aa C42:5	+	0	0	+	+	0	0	0	0	0
PC aa C42:6	+	0	0	+	+	0	0	0	0	0
PC ae C30:0	+	0	0	-	-	0	0	0	0	+
PC ae C30:2	+	0	0	-	-	0	0	0	0	+
PC ae C32:1	+	0	0	-	-	0	0	0	0	+
PC ae C32:2	+	0	0	-	-	0	0	0	0	+
PC ae C34:0	+	0	0	-	-	0	0	0	0	+
PC ae C34:1	+	0	0	-	-	0	0	0	0	+
PC ae C34:2	+	0	0	-	-	0	0	0	0	+
PC ae C34:3	+	0	0	-	-	0	0	0	0	+
PC ae C36:0	+	0	0	+	+	0	0	0	0	0
PC ae C36:1	+	0	0	-	-	0	0	0	0	+
PC ae C36:2	+	0	0	-	-	0	0	0	0	+
PC ae C36:3	+	0	0	-	-	0	0	0	0	+
PC ae C36:4	+	0	0	0	0	0	0	0	0	0
PC ae C36:5	+	0	0	0	0	0	0	0	0	0
PC ae C38:2	+	0	0	-	-	0	0	0	0	+
PC ae C38:3	+	0	0	-	-	0	0	0	0	+
PC ae C38:4	+	0	0	0	0	0	0	0	0	0
PC ae C38:5	+	0	0	0	0	0	0	0	0	0
PC ae C38:6	+	0	0	+	+	0	0	0	0	0

Continues...

Table 7.3 continued

Metabolites	Sign patterns									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
PC ae C40:1	+	0	0	+	+	0	0	0	0	0
PC ae C40:2	+	0	0	-	-	0	0	0	0	+
PC ae C40:3	+	0	0	-	-	0	0	0	0	+
PC ae C40:4	+	0	0	0	0	0	0	0	0	0
PC ae C40:5	+	0	0	0	0	0	0	0	0	0
PC ae C40:6	+	0	0	+	+	0	0	0	0	0
PC ae C42:1	+	0	0	0	0	0	0	0	0	0
PC ae C42:2	+	0	0	+	+	0	0	0	0	0
PC ae C42:3	+	0	0	+	+	0	0	0	0	0
PC ae C42:4	+	0	0	0	0	0	0	0	0	0
PC ae C42:5	+	0	0	0	0	0	0	0	0	0
PC ae C44:4	+	0	0	0	0	0	0	0	0	0
PC ae C44:5	+	0	0	0	0	0	0	0	0	0
PC ae C44:6	+	0	0	0	0	0	0	0	0	0
Hexose										
Hexose	0	0	0	0	0	0	0	+	0	0
Sphingolipids										
SM (OH) C14:1	+	0	0	-	-	0	0	0	0	+
SM (OH) C16:1	+	0	0	-	-	0	0	0	0	+
SM (OH) C22:1	+	0	0	0	0	0	0	0	0	0
SM (OH) C22:2	+	0	0	-	-	0	0	0	0	+
SM (OH) C24:1	+	0	0	0	0	0	0	0	0	0
SM C16:0	+	0	0	0	0	0	0	0	0	0
SM C16:1	+	0	0	0	0	0	0	0	0	0
SM C18:0	+	0	0	0	0	0	0	0	0	0
SM C18:1	+	0	0	0	0	0	0	0	0	0
SM C20:2	0	0	0	0	0	0	0	0	0	0
SM C24:0	+	0	0	0	0	0	0	0	0	0
SM C24:1	+	0	0	0	0	0	0	0	0	0

[1] Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 105. Analysis was repeated 100 times in random sub-samples in 80% of the data

Table 7.4: Characteristics of 1,593 prostate cancer cases and 1,593 controls in EPIC

Characteristic	Cases, n=1593	Controls, n=1593
Age at blood collection, years (SD)	60.0 (7.1)	59.9 (7.1)
Height, cm (SD)[1]	172.0 (7.0)	172.3 (7.1)
BMI, kg/m² (SD)[1]	26.8 (3.4)	26.9 (3.6)
Smoking, n (%)[1]		
Never	509 (32.5)	463 (29.6)
Former	697 (44.5)	724 (46.3)
Current	359 (22.9)	377 (24.1)
Alcohol consumption, n (%)[1]		
<10 g/day	701 (44.2)	699 (43.9)
10-19 g/day	305 (19.2)	303 (19.0)
20-40 g/day	300 (18.9)	339 (21.3)
≥40 g/day	280 (17.7)	250 (15.7)
Physical activity, n (%)[1]		
Inactive	411 (26.5)	416 (26.8)
Moderately inactive	512 (33.0)	502 (32.3)
Moderately active	354 (22.8)	348 (22.4)
Active	275 (17.7)	289 (18.6)
Marital status, n (%)[1]		
Married or cohabiting	1100 (89.7)	1110 (90.2)
Not married or cohabiting	126 (10.3)	120 (9.8)
Educational attainment, n (%)[1]		
Primary or equivalent	611 (40.9)	631 (42.0)
Secondary	537 (35.9)	558 (37.2)
Degree	347 (23.2)	312 (20.8)
Cases only		
Age at diagnosis, years (SD)	66.8 (7.01)	-
Time to diagnosis, n (%)[2]		
<2 years	145 (9.1)	-
2-<4 years	207 (13.0)	-
4-<6 years	305 (19.2)	-
6-<8 years	300 (18.8)	-
8-<10 years	278 (17.5)	-
≥10 years	358 (22.5)	-
Year of diagnosis, median (range)	2002 (1994-2008)	-
Grade, n (%)[1, 3]		
Low-intermediate grade	1041 (85.1)	-
High grade	183 (14.9)	-
Stage, n (%)[1, 4]		
Localised	702 (70.1)	-
Advanced	300 (29.9)	-
Non-aggressive	821 (81.9)	-
Aggressive	181 (18.1)	-
Death from prostate cancer, n (%)[1, 5]	172 (11.4)	-

[1] Unknown values for some participants; the calculations of percentages exclude missing values

[2] Time between blood collection and diagnosis

[3] Gleason score <8 or coded as well, moderately or poorly differentiated for low-intermediate grade and Gleason score ≥ 8 or coded as undifferentiated for high grade

[4] Categorised using the TNM system; localised: $\leq$ T2 and N0/X and M0, or coded as localised; advanced: T3 $-$ 4 and/or N1 $-$ 3 and/or M1, or coded as advanced; and aggressive: T4 and/or N1 $-$ 3 and/or M1. Categories are not mutually exclusive so numbers do not add up; percentages were calculated separately for localised and advanced, and for non-aggressive and aggressive

[5] Death from prostate cancer (prostate cancer listed as the underlying cause of death on the death certificate) during follow-up; 194 died of prostate cancer but 22 were excluded from further analysis as their matched control had died before them (n $=$ 18) or vital status was unknown for the control (n $=$ 4)

Table 7.5: Variance in metabolite concentrations explained by the first seven components from principal component analysis and treelet transform[1]

Components	Variance explained, %	
	PCA	TT[2]
Component 1	21.44	19.90
Component 2	9.61	5.73
Component 3	7.13	4.33
Component 4	6.61	3.75
Component 5	5.80	3.08
Component 6	5.20	1.73
Component 7	4.48	3.18
Total	60.27	41.70

[1] Principal component analysis and treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites

[2] Adjusted for correlation between treelet component scores

Table 7.6: Loadings for the first seven principal components[1]

Metabolites	Principal component loadings						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Acylcarnitines							
C0	0.0192	-0.0027	0.0434	0.0710	-0.0191	-0.0017	0.0565
C14:1	0.0304	-0.1125	0.0346	0.3762	0.0138	-0.5397	0.1770
C16	0.0758	0.0953	0.1525	0.0436	0.0306	-0.1530	-0.0197
C18	0.0843	0.1470	0.1327	0.1406	0.1396	-0.1585	0.1384
C18:1	0.0321	0.2161	0.1102	0.1571	-0.0391	-0.1853	-0.0742
C18:2	-0.0060	0.4089	0.0693	0.1646	-0.0292	-0.0763	-0.0850
C2	0.0288	-0.0504	0.0630	0.1228	0.0141	-0.1903	-0.0060
C3	-0.0131	0.0257	0.1038	0.2982	0.0028	0.0680	0.5867
Amino acids							
Alanine	0.0023	0.0727	0.0645	-0.0131	-0.0410	0.1084	0.0529
Arginine	-0.0013	-0.3638	-0.1916	0.0345	-0.0036	0.1652	0.1558
Asparagine	0.0003	0.0770	-0.0096	0.0400	0.0018	0.1062	0.0284
Citrulline	0.0067	0.0543	-0.0158	0.0086	0.0152	0.0371	0.0347
Glutamine	0.0110	-0.0097	-0.0313	0.0191	0.0329	0.0582	0.0028
Glutamate	0.0019	0.3187	0.2189	-0.0211	-0.1252	0.0420	0.1050
Glycine	0.0092	0.0931	-0.0281	0.0253	0.0391	0.0752	-0.0065
Histidine	0.0123	0.0638	0.0256	0.0486	-0.0142	0.0761	0.0278
Isoleucine	-0.0003	0.0279	0.0602	0.0696	-0.0692	0.1638	0.1199
Leucine	0.0032	0.0414	0.0552	0.0586	-0.0586	0.1341	0.1090
Lysine	0.0044	0.0404	0.0177	0.0718	-0.0606	0.0996	0.0910
Methionine	-0.0078	-0.0135	0.0027	0.0806	-0.0490	0.1731	0.0973
Ornithine	0.0132	0.2373	0.0814	0.0554	-0.0168	0.1065	0.0359
Phenylalanine	0.0048	0.0593	0.0326	0.0387	-0.0229	0.0864	0.0805
Proline	0.0199	0.0698	0.0652	-0.0054	-0.0002	0.1852	0.0832
Serine	-0.0099	0.0895	-0.0383	0.0732	0.0191	0.0775	0.0139
t4-OH-Pro	0.0149	-0.0544	0.0116	0.1274	-0.0338	0.2023	0.0743
Threonine	-0.0081	0.0513	-0.0075	0.0594	-0.0240	0.1252	0.0362
Tryptophan	0.0154	-0.0058	0.0336	0.0535	-0.0301	0.0911	0.0601
Tyrosine	0.0050	0.0289	0.0715	0.0477	-0.0619	0.1269	0.1041
Valine	0.0046	0.0113	0.0467	0.0508	-0.0572	0.1068	0.0945
Biogenic amines							
Creatinine	0.0135	0.0290	-0.0037	0.0005	-0.0010	0.0160	0.0409
Kynurenine	0.0105	0.0256	0.0365	0.0234	0.0200	0.0159	0.1138
Sarcosine	0.0107	0.0009	0.0777	0.1211	-0.1074	0.2935	0.0586
Taurine	0.0251	0.2198	0.0238	0.0068	-0.0523	0.0187	0.0350
ADMA	0.0019	0.0704	0.0329	0.0299	0.0029	0.0466	0.0509
Glycerophospholipids							
lysoPC a C16:0	0.0581	0.0734	0.0829	0.0870	-0.0081	0.0586	-0.1198
lysoPC a C16:1	0.1113	0.0626	0.1841	-0.0114	0.0161	0.0170	-0.1023
lysoPC a C17:0	0.0826	0.1275	-0.0506	0.0445	0.1208	0.0810	-0.0554
lysoPC a C18:0	0.0475	0.0746	0.0121	0.1281	0.0154	0.0740	-0.1412
lysoPC a C18:1	0.0654	0.0112	0.0384	0.1540	0.0367	0.0375	-0.1862
lysoPC a C18:2	0.0442	0.0378	-0.0425	0.0825	0.1237	0.1618	-0.1475
lysoPC a C20:3	0.0687	-0.0764	0.1640	0.1336	0.0916	0.0605	-0.1713
lysoPC a C20:4	0.0430	-0.0561	0.0628	0.1982	0.0341	0.0409	-0.2491
PC aa C28:1	0.1264	-0.0074	0.0133	-0.0670	0.1005	-0.0065	0.0685
PC aa C30:0	0.1674	-0.0745	0.1433	-0.1034	0.0246	0.0317	0.0469

Continues...

Table 7.6 continued

Metabolites	Principal component loadings						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
PC aa C32:0	0.1077	-0.0465	0.0606	0.0213	-0.0096	-0.0034	-0.0117
PC aa C32:1	0.2031	-0.1708	0.3635	-0.1400	-0.0387	-0.0467	0.0206
PC aa C32:2	0.1444	-0.0385	0.1648	-0.1365	0.0425	0.1013	0.0518
PC aa C32:3	0.1301	0.0496	-0.0280	-0.1131	0.0706	0.0112	0.0546
PC aa C34:1	0.1180	-0.1039	0.1408	0.0376	-0.0373	-0.0411	-0.0419
PC aa C34:2	0.0663	-0.0124	0.0339	-0.0095	0.0298	0.0451	-0.0242
PC aa C34:3	0.1478	0.0098	0.1254	-0.1519	0.0706	0.0212	0.0171
PC aa C34:4	0.1553	-0.0876	0.1644	-0.0497	-0.0246	0.0727	-0.0347
PC aa C36:0	0.1001	-0.0110	-0.1416	0.0538	-0.1542	0.0263	-0.0071
PC aa C36:1	0.1157	-0.0831	0.0926	0.0659	-0.0299	-0.0112	0.0023
PC aa C36:2	0.0640	-0.0117	0.0081	0.0315	0.0272	0.0613	-0.0208
PC aa C36:3	0.0818	-0.0140	0.0902	0.0213	0.0483	0.0326	-0.0160
PC aa C36:4	0.0810	-0.0864	0.1014	0.0899	-0.0159	0.0189	-0.0934
PC aa C36:5	0.2174	-0.0528	0.0111	-0.1281	-0.3556	-0.1041	0.0437
PC aa C36:6	0.1932	-0.0179	-0.0161	-0.1049	-0.2530	-0.0142	0.0694
PC aa C38:0	0.0850	0.0139	-0.1926	0.0994	-0.1763	-0.0068	-0.0150
PC aa C38:3	0.0842	-0.0539	0.1214	0.0572	-0.0200	0.0180	-0.0099
PC aa C38:4	0.0657	-0.0743	0.0757	0.1247	-0.0216	0.0092	-0.0832
PC aa C38:5	0.1357	-0.0250	0.0507	-0.0152	-0.1001	-0.0485	-0.0332
PC aa C38:6	0.1133	-0.0116	-0.0700	0.0282	-0.2491	-0.0649	0.0261
PC aa C40:2	0.1087	0.1059	-0.1265	-0.0901	-0.0520	-0.0075	0.0139
PC aa C40:3	0.1291	0.0835	-0.0693	-0.1041	-0.0294	-0.0241	-0.0102
PC aa C40:4	0.0912	-0.0460	0.1335	0.0474	0.0370	0.0113	-0.0606
PC aa C40:5	0.1384	-0.0151	0.1286	-0.0598	-0.0316	-0.0214	-0.0274
PC aa C40:6	0.1097	0.0133	-0.0661	0.0171	-0.2660	-0.0578	0.0533
PC aa C42:0	0.0864	0.0938	-0.1891	0.0316	-0.0787	0.0008	-0.0191
PC aa C42:1	0.0745	0.0799	-0.1823	0.0548	-0.1014	0.0260	-0.0342
PC aa C42:2	0.0826	0.0596	-0.1469	0.0469	-0.1492	0.0215	-0.0275
PC aa C42:4	0.0817	0.0222	-0.0163	0.0181	0.0025	0.0176	-0.0580
PC aa C42:5	0.1389	0.0686	-0.0157	-0.1348	-0.0816	-0.0533	0.0232
PC aa C42:6	0.1360	0.0620	-0.0310	-0.0772	-0.1194	-0.0103	0.0218
PC ae C30:0	0.1544	-0.0068	0.0168	-0.1360	0.1182	0.0205	0.0604
PC ae C30:2	0.1225	0.0378	-0.0565	-0.0924	0.0950	-0.0032	0.0348
PC ae C32:1	0.1108	-0.0386	-0.0124	0.0180	0.0425	0.0099	-0.0157
PC ae C32:2	0.1160	-0.0177	-0.0763	0.0141	0.0000	0.0046	-0.0131
PC ae C34:0	0.1548	-0.0339	0.0026	-0.0392	0.0527	0.0176	0.0638
PC ae C34:1	0.1153	-0.0515	0.0282	0.0161	0.0761	-0.0112	0.0085
PC ae C34:2	0.0909	-0.0192	-0.0576	0.0328	0.1110	0.0776	-0.0070
PC ae C34:3	0.0959	-0.0105	-0.0741	0.0558	0.0949	0.0828	-0.0611
PC ae C36:0	0.1232	-0.0218	-0.0328	0.0272	-0.0779	-0.0506	0.0017
PC ae C36:1	0.1280	-0.0284	0.0040	-0.0078	0.0874	-0.0097	0.0458
PC ae C36:2	0.0991	0.0314	-0.0768	-0.0202	0.1313	0.0663	0.0306
PC ae C36:3	0.0905	-0.0134	-0.0443	0.0439	0.0965	0.0684	-0.0364
PC ae C36:4	0.0715	-0.0974	0.0117	0.1401	0.0492	0.0530	-0.0655
PC ae C36:5	0.0912	-0.0980	-0.0275	0.1387	-0.0427	0.0373	-0.0776
PC ae C38:2	0.1010	0.0523	-0.0570	-0.0103	0.0770	0.0488	0.0363
PC ae C38:3	0.1098	0.0000	0.0126	-0.0090	0.1260	0.0295	0.0428
PC ae C38:4	0.0818	-0.0589	-0.0005	0.0915	0.1144	0.0329	-0.0442
PC ae C38:5	0.0691	-0.0567	-0.0371	0.1457	-0.0095	0.0206	-0.0689
PC ae C38:6	0.1034	-0.0471	-0.1141	0.1034	-0.1402	0.0051	-0.0179

Continues...

Table 7.6 continued

Metabolites	Principal component loadings						
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
PC ae C40:1	0.1190	0.0447	-0.0606	0.0086	-0.0872	0.0348	-0.0700
PC ae C40:2	0.1260	0.0256	-0.0768	-0.0495	0.0418	-0.0304	0.0580
PC ae C40:3	0.1030	0.0522	-0.0545	-0.0320	0.0946	0.0129	0.0182
PC ae C40:4	0.0861	0.0018	-0.0365	0.0443	0.1021	0.0193	-0.0524
PC ae C40:5	0.1042	0.0109	-0.0753	0.0390	0.0115	0.0071	-0.0471
PC ae C40:6	0.1088	0.0348	-0.1552	0.0220	-0.0733	-0.0076	0.0259
PC ae C42:1	0.0882	0.0339	-0.0145	0.0223	-0.0147	0.0285	-0.0870
PC ae C42:2	0.1287	0.0429	-0.0487	-0.0288	-0.0308	-0.0252	0.0052
PC ae C42:3	0.1106	0.0734	-0.1055	0.0079	-0.0444	0.0068	-0.0268
PC ae C42:4	0.0903	0.0463	-0.0688	0.0304	0.1046	0.0455	-0.0588
PC ae C42:5	0.0749	0.0302	-0.0788	0.0508	0.0393	0.0010	-0.0554
PC ae C44:4	0.0791	0.0365	-0.0329	0.0438	0.0724	0.0386	-0.0684
PC ae C44:5	0.0715	-0.0159	-0.0526	0.1372	0.0404	0.0017	-0.0967
PC ae C44:6	0.0616	0.0646	-0.1314	0.0825	0.0029	0.0334	-0.0705
Hexose							
Hexose	-0.0141	-0.2617	-0.0445	0.0241	-0.0200	0.0505	0.1163
Sphingolipids							
SM (OH) C14:1	0.1087	0.0267	-0.0707	-0.0333	0.1686	0.0032	0.1044
SM (OH) C16:1	0.0858	-0.0013	-0.1001	0.0327	0.1454	-0.0086	0.1146
SM (OH) C22:1	0.0689	-0.0123	-0.0552	0.0841	0.0779	0.0080	0.0750
SM (OH) C22:2	0.0934	0.0139	-0.0870	0.0184	0.1109	-0.0169	0.0597
SM (OH) C24:1	0.0548	-0.0312	-0.0691	0.1164	0.0722	0.0230	0.0623
SM C16:0	0.0638	-0.0019	-0.0753	0.0776	0.0661	-0.0190	0.0318
SM C16:1	0.0671	-0.0097	-0.0447	0.0580	0.0675	-0.0288	0.0203
SM C18:0	0.0729	-0.0504	-0.0227	0.0734	0.0520	-0.0457	0.0663
SM C18:1	0.0698	-0.0153	-0.0432	0.0521	0.0742	-0.0528	0.0534
SM C20:2	0.1120	0.1559	-0.1380	-0.2007	0.2957	-0.1535	0.1957
SM C24:0	0.0588	-0.0214	-0.0206	0.1041	0.0150	0.0113	0.0228
SM C24:1	0.0576	-0.0220	-0.0831	0.1011	-0.0016	-0.0625	0.0254

[1] Principal component analysis using the covariance matrix of ln-transformed concentrations of 121 metabolites

Table 7.7: Principal component scores and risk of overall prostate cancer in 1,593 matched case-control sets in EPIC[1]

Components	Fifths of principal component score									
	1	2		3		4		5		p_{trend}[2]
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	
PC1	Ref	1.30	(1.04 - 1.63)	1.31	(1.04 - 1.65)	1.11	(0.88 - 1.40)	1.04	(0.81 - 1.33)	0.9
PC2	Ref	1.27	(1.00 - 1.61)	1.04	(0.79 - 1.36)	1.34	(1.01 - 1.79)	1.13	(0.80 - 1.58)	0.5
PC3	Ref	0.85	(0.68 - 1.07)	0.89	(0.71 - 1.12)	0.91	(0.72 - 1.15)	0.90	(0.70 - 1.17)	0.6
PC4	Ref	1.02	(0.81 - 1.28)	0.82	(0.64 - 1.06)	0.96	(0.74 - 1.25)	0.85	(0.64 - 1.14)	0.3
PC5	Ref	0.95	(0.76 - 1.20)	1.08	(0.86 - 1.36)	0.84	(0.66 - 1.07)	1.02	(0.79 - 1.32)	0.8
PC6	Ref	1.21	(0.96 - 1.52)	1.11	(0.87 - 1.40)	1.10	(0.86 - 1.40)	1.29	(1.01 - 1.65)	0.1
PC7	Ref	0.96	(0.75 - 1.23)	1.00	(0.77 - 1.29)	1.15	(0.89 - 1.49)	1.22	(0.93 - 1.59)	0.1

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the fifths of component scores

Table 7.8: Principal component scores and risk of prostate cancer by time to diagnosis in 1,593 matched case-control sets in EPIC[1]

≤5 years to diagnosis, n = 501 cases

Components	Thirds of principal component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
PC1	Ref	1.18	(0.85 - 1.63)	1.11	(0.80 - 1.55)	0.6
PC2	Ref	0.94	(0.66 - 1.34)	1.09	(0.69 - 1.73)	0.7
PC3	Ref	0.95	(0.68 - 1.33)	0.63	(0.45 - 0.89)	0.006
PC4	Ref	1.08	(0.78 - 1.49)	0.96	(0.64 - 1.42)	0.9
PC5	Ref	1.05	(0.75 - 1.45)	0.83	(0.59 - 1.17)	0.3
PC6	Ref	1.16	(0.82 - 1.63)	1.53	(1.09 - 2.15)	0.01
PC7	Ref	1.09	(0.77 - 1.55)	1.45	(1.01 - 2.10)	0.05

>5 years to diagnosis, n = 1,092 cases

Components	Thirds of principal component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
PC1	Ref	1.12	(0.90 - 1.38)	0.93	(0.75 - 1.17)	0.6	0.4
PC2	Ref	1.08	(0.84 - 1.38)	1.06	(0.78 - 1.46)	0.7	0.9
PC3	Ref	0.95	(0.76 - 1.17)	1.19	(0.94 - 1.52)	0.1	0.004
PC4	Ref	0.88	(0.70 - 1.12)	0.91	(0.68 - 1.21)	0.5	0.7
PC5	Ref	1.11	(0.89 - 1.39)	1.10	(0.86 - 1.40)	0.5	0.2
PC6	Ref	1.14	(0.92 - 1.41)	1.04	(0.83 - 1.32)	0.7	0.1
PC7	Ref	1.11	(0.88 - 1.39)	1.14	(0.89 - 1.47)	0.3	0.5

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component scores

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and time to diagnosis ($\leq$5 versus $>$5 years)

Table 7.9: Principal component scores and risk of prostate cancer by grade in 1,224 matched case-control sets in EPIC[1]

Low-intermediate grade, n = 1,041 cases

Components	Thirds of principal component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
PC1	Ref	1.21	(0.97 - 1.50)	0.97	(0.77 - 1.23)	0.9
PC2	Ref	1.03	(0.81 - 1.31)	1.02	(0.74 - 1.40)	0.9
PC3	Ref	1.01	(0.80 - 1.26)	0.98	(0.77 - 1.25)	0.8
PC4	Ref	0.95	(0.75 - 1.21)	0.92	(0.69 - 1.22)	0.6
PC5	Ref	1.11	(0.89 - 1.39)	1.04	(0.81 - 1.34)	0.7
PC6	Ref	1.16	(0.93 - 1.44)	1.14	(0.90 - 1.45)	0.3
PC7	Ref	1.04	(0.82 - 1.31)	1.04	(0.80 - 1.34)	0.8

High grade, n = 183 cases

Components	Thirds of principal component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
PC1	Ref	0.98	(0.54 - 1.80)	0.52	(0.29 - 0.93)	0.02	0.1
PC2	Ref	1.04	(0.53 - 2.03)	0.75	(0.31 - 1.80)	0.5	0.3
PC3	Ref	1.04	(0.57 - 1.89)	0.99	(0.53 - 1.85)	>0.9	0.3
PC4	Ref	0.67	(0.37 - 1.23)	0.47	(0.21 - 1.05)	0.06	0.1
PC5	Ref	0.84	(0.46 - 1.57)	1.07	(0.57 - 2.01)	0.9	0.6
PC6	Ref	1.12	(0.59 - 2.12)	1.30	(0.68 - 2.47)	0.4	0.8
PC7	Ref	1.06	(0.55 - 2.03)	1.63	(0.83 - 3.21)	0.2	0.2

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component score.

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and grade (low-intermediate versus high grade)

Table 7.10: Principal component scores and risk of prostate cancer by stage (localised and advanced) in 1,002 matched case-control sets in EPIC[1]

Localised, n = 702 cases

Components	Thirds of principal component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
PC1	Ref	1.74	(1.31 - 2.32)	1.25	(0.95 - 1.66)	0.1
PC2	Ref	0.97	(0.72 - 1.31)	1.10	(0.75 - 1.61)	0.6
PC3	Ref	0.94	(0.72 - 1.24)	0.89	(0.66 - 1.20)	0.4
PC4	Ref	1.01	(0.77 - 1.34)	0.92	(0.64 - 1.33)	0.7
PC5	Ref	0.89	(0.68 - 1.17)	0.90	(0.66 - 1.22)	0.5
PC6	Ref	1.12	(0.85 - 1.48)	1.06	(0.79 - 1.41)	0.7
PC7	Ref	0.78	(0.57 - 1.07)	0.87	(0.63 - 1.20)	0.5

Advanced, n = 300 cases

Components	Thirds of principal component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
PC1	Ref	0.68	(0.44 - 1.06)	0.46	(0.29 - 0.71)	0.0005	0.0001
PC2	Ref	1.11	(0.70 - 1.76)	0.68	(0.36 - 1.29)	0.3	0.4
PC3	Ref	1.02	(0.66 - 1.57)	1.02	(0.63 - 1.64)	0.9	0.6
PC4	Ref	0.84	(0.56 - 1.27)	0.81	(0.45 - 1.43)	0.4	0.7
PC5	Ref	1.09	(0.69 - 1.73)	1.35	(0.85 - 2.14)	0.2	0.3
PC6	Ref	0.96	(0.60 - 1.51)	1.32	(0.81 - 2.17)	0.3	0.5
PC7	Ref	1.19	(0.75 - 1.89)	1.06	(0.65 - 1.72)	0.8	0.3

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component scores

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and stage (localised versus advanced stage)

Table 7.11: Principal component scores and risk of prostate cancer by stage (non-aggressive and aggressive) in 1,002 matched case-control sets in EPIC[1]

Non-aggressive, n = 821 cases

Components	Thirds of principal component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
PC1	Ref	1.44	(1.12 - 1.86)	1.02	(0.79 - 1.31)	0.9
PC2	Ref	0.97	(0.74 - 1.27)	1.07	(0.76 - 1.52)	0.7
PC3	Ref	0.97	(0.76 - 1.25)	0.98	(0.74 - 1.29)	0.9
PC4	Ref	1.00	(0.77 - 1.30)	0.85	(0.61 - 1.18)	0.4
PC5	Ref	0.97	(0.75 - 1.25)	0.98	(0.74 - 1.30)	0.9
PC6	Ref	1.08	(0.83 - 1.39)	1.14	(0.87 - 1.50)	0.3
PC7	Ref	0.78	(0.59 - 1.04)	0.84	(0.63 - 1.13)	0.3

Aggressive, n = 181 cases

Components	Thirds of principal component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
PC1	Ref	0.72	(0.39 - 1.33)	0.60	(0.34 - 1.06)	0.08	0.08
PC2	Ref	1.19	(0.58 - 2.41)	0.65	(0.26 - 1.61)	0.2	0.3
PC3	Ref	0.92	(0.54 - 1.59)	0.80	(0.42 - 1.51)	0.5	0.7
PC4	Ref	0.93	(0.57 - 1.51)	1.40	(0.63 - 3.07)	0.6	0.6
PC5	Ref	0.89	(0.49 - 1.63)	1.12	(0.62 - 2.00)	0.7	>0.9
PC6	Ref	1.03	(0.57 - 1.87)	0.91	(0.48 - 1.72)	0.8	0.5
PC7	Ref	2.05	(1.02 - 4.11)	1.70	(0.89 - 3.27)	0.1	0.04

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component scores

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and stage (non-aggressive versus aggressive)

Table 7.12: Principal component scores and risk of death from prostate cancer in 172 matched case-control sets in EPIC[1,2]

Components	Thirds of principal component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
PC1	Ref	1.10	(0.60 - 2.00)	0.83	(0.45 - 1.52)	0.5
PC2	Ref	0.94	(0.45 - 1.97)	0.63	(0.27 - 1.49)	0.2
PC3	Ref	0.65	(0.36 - 1.18)	0.90	(0.44 - 1.83)	0.8
PC4	Ref	1.25	(0.69 - 2.27)	1.27	(0.58 - 2.80)	0.5
PC5	Ref	1.31	(0.72 - 2.38)	1.12	(0.60 - 2.09)	0.7
PC6	Ref	0.96	(0.54 - 1.70)	1.25	(0.65 - 2.40)	0.5
PC7	Ref	3.36	(1.62 - 7.01)	4.88	(2.18 - 10.89)	0.0001

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Death from prostate cancer: prostate cancer listed as the underlying cause of death on the death certificate during follow-up; 22 were excluded from further analysis as their matched control had died before them (n $= 18$) or vital status was unknown for the control (n $= 4$)

[3] Computed across the medians of the thirds of component scores

Table 7.13: Loadings for the first seven treelet components[1]

Metabolites	Treelet component loadings						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
Acylcarnitines							
C0							
C14:1				0.9536			
C16		0.2388					
C18		0.3004					
C18:1		0.3212					
C18:2		0.6025					
C2				0.3012			
C3							
Amino acids							
Alanine							
Arginine						0.8521	
Asparagine							
Citrulline							
Glutamine							
Glutamate		0.5013					
Glycine							
Histidine							
Isoleucine							
Leucine							
Lysine							
Methionine							
Ornithine		0.2615					
Phenylalanine							
Proline							
Serine							
t4-OH-Pro							
Threonine							
Tryptophan							
Tyrosine							
Valine							
Biogenic amines							
Creatinine							
Kynurenine							
Sarcosine							
Taurine		0.2588					
ADMA							
Glycerophospholipids							
lysoPC a C16:0					0.2844		
lysoPC a C16:1	0.1196		-0.0987				-0.1556
lysoPC a C17:0					0.3224		
lysoPC a C18:0					0.3177		
lysoPC a C18:1					0.3978		
lysoPC a C18:2					0.3785		
lysoPC a C20:3					0.4586		
lysoPC a C20:4					0.4498		
PC aa C28:1	0.1205		-0.0994				0.1716
PC aa C30:0	0.1658		-0.1368				-0.2157

Continues...

Table 7.13 continued

Metabolites	Treelet component loadings						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
PC aa C32:0	0.0899		-0.0741				-0.1169
PC aa C32:1	0.2743		-0.2263				-0.3568
PC aa C32:2	0.1820		-0.1502				-0.2368
PC aa C32:3	0.1141		-0.0941				0.1625
PC aa C34:1	0.1247		-0.1029				-0.1622
PC aa C34:2	0.0539		-0.0445				0.0768
PC aa C34:3	0.1480		-0.1221				-0.1925
PC aa C34:4	0.1678		-0.1384				-0.2183
PC aa C36:0	0.1270		0.1804				
PC aa C36:1	0.1192		-0.0983				-0.1550
PC aa C36:2	0.0558		-0.0460				0.0794
PC aa C36:3	0.0746		-0.0616				-0.0971
PC aa C36:4	0.0819						
PC aa C36:5	0.2479		0.3522				
PC aa C36:6	0.1891		0.2687				
PC aa C38:0	0.1334		0.1895				
PC aa C38:3	0.1000		-0.0825				-0.1301
PC aa C38:4	0.0861						
PC aa C38:5	0.1079		-0.0890				-0.1404
PC aa C38:6	0.1356		0.1926				
PC aa C40:2	0.1307		0.1857				
PC aa C40:3	0.1343		0.1907				
PC aa C40:4	0.0820						
PC aa C40:5	0.1297		-0.1070				-0.1687
PC aa C40:6	0.1437		0.2042				
PC aa C42:0	0.0941						
PC aa C42:1	0.0913						
PC aa C42:2	0.1100		0.1563				
PC aa C42:4	0.0813						
PC aa C42:5	0.1356		0.1926				
PC aa C42:6	0.1255		0.1783				
PC ae C30:0	0.1629		-0.1344				0.2319
PC ae C30:2	0.1236		-0.1020				0.1760
PC ae C32:1	0.0879		-0.0725				0.1251
PC ae C32:2	0.0935		-0.0771				0.1331
PC ae C34:0	0.1385		-0.1143				0.1973
PC ae C34:1	0.1114		-0.0919				0.1587
PC ae C34:2	0.0870		-0.0718				0.1239
PC ae C34:3	0.1009		-0.0833				0.1438
PC ae C36:0	0.1073		0.1525				
PC ae C36:1	0.1203		-0.0993				0.1714
PC ae C36:2	0.1082		-0.0892				0.1541
PC ae C36:3	0.0840		-0.0693				0.1195
PC ae C36:4	0.0853						
PC ae C36:5	0.0898						
PC ae C38:2	0.0914		-0.0754				0.1301
PC ae C38:3	0.1032		-0.0851				0.1469
PC ae C38:4	0.0690						

Continues...

Table 7.13 continued

Metabolites	Treelet component loadings						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
PC ae C38:5	0.0700						
PC ae C38:6	0.1152		0.1637				
PC ae C40:1	0.0926		0.1316				
PC ae C40:2	0.1176		-0.0970				0.1675
PC ae C40:3	0.0939		-0.0775				0.1338
PC ae C40:4	0.0597						
PC ae C40:5	0.0635						
PC ae C40:6	0.1037		0.1473				
PC ae C42:1	0.0782						
PC ae C42:2	0.0940		0.1335				
PC ae C42:3	0.0947		0.1346				
PC ae C42:4	0.0773						
PC ae C42:5	0.0633						
PC ae C44:4	0.0789						
PC ae C44:5	0.0824						
PC ae C44:6	0.0818						
Hexose							
Hexose						0.5234	
Sphingolipids							
SM (OH) C14:1	0.1281		-0.1057				0.1824
SM (OH) C16:1	0.1188		-0.0980				0.1692
SM (OH) C22:1	0.0681						
SM (OH) C22:2	0.1054		-0.0869				0.1501
SM (OH) C24:1	0.0857						
SM C16:0	0.0503						
SM C16:1	0.0500						
SM C18:0	0.0683						
SM C18:1	0.0647						
SM C20:2							
SM C24:0	0.0608						
SM C24:1	0.0645						

[1] Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 105

Table 7.14: Pearson's correlation coefficients between treelet component scores

	TC1	TC2	TC3	TC4	TC5	TC6	TC7
TC1	1						
TC2	0.06	1					
TC3	0.00	0.07	1				
TC4	0.07	0.06	0.02	1			
TC5	0.31	0.13	-0.15	0.04	1		
TC6	0.02	-0.67	-0.03	0.12	-0.12	1	
TC7	0.08	0.00	0.10	-0.06	0.02	-0.01	1

[1] Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 105

Table 7.15: Loadings for the first seven treelet components at cut-level 102[1]

Metabolites	Treelet component loadings at cut-level 102						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
Acylcarnitines							
C0							
C14:1				0.9536			
C16		0.2503					
C18							
C18:1		0.3367					
C18:2		0.6316					
C2				0.3012			
C3							
Amino acids							
Alanine							
Arginine							0.8521
Asparagine							
Citrulline							
Glutamine							
Glutamate		0.5255					
Glycine							
Histidine							
Isoleucine							
Leucine							
Lysine							
Methionine							
Ornithine		0.2742					
Phenylalanine							
Proline							
Serine							
t4-OH-Pro							
Threonine							
Tryptophan							
Tyrosine							
Valine							
Biogenic amines							
Creatinine							
Kynurenine							
Sarcosine							
Taurine		0.2714					
ADMA							
Glycerophospholipids							
lysoPC a C16:0					0.2844		
lysoPC a C16:1	0.1250		-0.0987				
lysoPC a C17:0					0.3224		
lysoPC a C18:0					0.3177		
lysoPC a C18:1					0.3978		
lysoPC a C18:2					0.3785		
lysoPC a C20:3					0.4586		
lysoPC a C20:4					0.4498		
PC aa C28:1	0.1260		-0.0994				
PC aa C30:0	0.1733		-0.1368				

Continues...

Table 7.15 continued

Metabolites	Treelet component loadings at cut-level 102						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
PC aa C32:0	0.0939		-0.0741				
PC aa C32:1	0.2867		-0.2263				
PC aa C32:2	0.1902		-0.1502				
PC aa C32:3	0.1192		-0.0941				
PC aa C34:1	0.1303		-0.1029				
PC aa C34:2	0.0564		-0.0445				
PC aa C34:3	0.1547		-0.1221				
PC aa C34:4	0.1754		-0.1384				
PC aa C36:0	0.1327		0.1804				
PC aa C36:1	0.1246		-0.0983				
PC aa C36:2	0.0583		-0.046				
PC aa C36:3	0.0780		-0.0616				
PC aa C36:4	0.0856						
PC aa C36:5	0.2591		0.3522				
PC aa C36:6	0.1976		0.2687				
PC aa C38:0	0.1394		0.1895				
PC aa C38:3	0.1046		-0.0825				
PC aa C38:4	0.0900						
PC aa C38:5	0.1128		-0.0890				
PC aa C38:6	0.1417		0.1926				
PC aa C40:2	0.1366		0.1857				
PC aa C40:3	0.1403		0.1907				
PC aa C40:4	0.0857						
PC aa C40:5	0.1355		-0.1070				
PC aa C40:6	0.1502		0.2042				
PC aa C42:0						0.4173	
PC aa C42:1						0.4045	
PC aa C42:2	0.1150		0.1563				
PC aa C42:4	0.0850						
PC aa C42:5	0.1417		0.1926				
PC aa C42:6	0.1312		0.1783				
PC ae C30:0	0.1702		-0.1344				
PC ae C30:2	0.1292		-0.1020				
PC ae C32:1	0.0918		-0.0725				
PC ae C32:2	0.0977		-0.0771				
PC ae C34:0	0.1448		-0.1143				
PC ae C34:1	0.1165		-0.0919				
PC ae C34:2	0.0909		-0.0718				
PC ae C34:3	0.1055		-0.0833				
PC ae C36:0	0.1122		0.1525				
PC ae C36:1	0.1258		-0.0993				
PC ae C36:2	0.1131		-0.0892				
PC ae C36:3	0.0877		-0.0693				
PC ae C36:4	0.0892						
PC ae C36:5	0.0939						
PC ae C38:2	0.0955		-0.0754				
PC ae C38:3	0.1078		-0.0851				
PC ae C38:4	0.0722						
PC ae C38:5	0.0732						
PC ae C38:6	0.1204		0.1637				

Continues...

Table 7.15 continued

Metabolites	Treelet component loadings at cut-level 102						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
PC ae C40:1	0.0968		0.1316				
PC ae C40:2	0.1229		-0.0970				
PC ae C40:3	0.0982		-0.0775				
PC ae C40:4	0.0624						
PC ae C40:5						0.2815	
PC ae C40:6	0.1084		0.1473				
PC ae C42:1	0.0817						
PC ae C42:2	0.0982		0.1335				
PC ae C42:3	0.0990		0.1346				
PC ae C42:4						0.3427	
PC ae C42:5						0.2805	
PC ae C44:4						0.3495	
PC ae C44:5						0.3651	
PC ae C44:6						0.3626	
Hexose							
Hexose							0.5234
Sphingolipids							
SM (OH) C14:1	0.1339		-0.1057				
SM (OH) C16:1	0.1242		-0.0980				
SM (OH) C22:1							
SM (OH) C22:2	0.1101		-0.0869				
SM (OH) C24:1							
SM C16:0							
SM C16:1							
SM C18:0							
SM C18:1							
SM C20:2							
SM C24:0							
SM C24:1							

[1] Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 102, i.e. -3 cut-levels of the main treelet transform

Table 7.16: Loadings for the first seven treelet components at cut-level 108[1]

Metabolites	Treelet component loadings at cut-level 108						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
Acylcarnitines							
C0							
C14:1				0.9536			
C16		0.2388					
C18		0.3004					
C18:1		0.3212					
C18:2		0.6025					
C2				0.3012			
C3							
Amino acids							
Alanine						0.1813	
Arginine							0.8521
Asparagine						0.1713	
Citrulline							
Glutamine							
Glutamate		0.5013					
Glycine						0.1829	
Histidine						0.1116	
Isoleucine						0.3243	
Leucine						0.2827	
Lysine						0.2181	
Methionine						0.2570	
Ornithine		0.2615					
Phenylalanine						0.1731	
Proline						0.2592	
Serine						0.1805	
t4-OH-Pro							
Threonine						0.1985	
Tryptophan						0.1646	
Tyrosine						0.2378	
Valine						0.2339	
Biogenic amines							
Creatinine							
Kynurenine							
Sarcosine						0.5346	
Taurine		0.2588					
ADMA							
Glycerophospholipids							
lysoPC a C16:0					0.2844		
lysoPC a C16:1	0.1196		-0.0987				
lysoPC a C17:0					0.3224		
lysoPC a C18:0					0.3177		
lysoPC a C18:1					0.3978		
lysoPC a C18:2					0.3785		
lysoPC a C20:3					0.4586		
lysoPC a C20:4					0.4498		
PC aa C28:1	0.1205		-0.0994				
PC aa C30:0	0.1658		-0.1368				

Continues...

Table 7.16 continued

Metabolites	Treelet component loadings at cut-level 108						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
PC aa C32:0	0.0899		-0.0741				
PC aa C32:1	0.2743		-0.2263				
PC aa C32:2	0.1820		-0.1502				
PC aa C32:3	0.1141		-0.0941				
PC aa C34:1	0.1247		-0.1029				
PC aa C34:2	0.0539		-0.0445				
PC aa C34:3	0.1480		-0.1221				
PC aa C34:4	0.1678		-0.1384				
PC aa C36:0	0.1270		0.1804				
PC aa C36:1	0.1192		-0.0983				
PC aa C36:2	0.0558		-0.0460				
PC aa C36:3	0.0746		-0.0616				
PC aa C36:4	0.0819						
PC aa C36:5	0.2479		0.3522				
PC aa C36:6	0.1891		0.2687				
PC aa C38:0	0.1334		0.1895				
PC aa C38:3	0.1000		-0.0825				
PC aa C38:4	0.0861						
PC aa C38:5	0.1079		-0.0890				
PC aa C38:6	0.1356		0.1926				
PC aa C40:2	0.1307		0.1857				
PC aa C40:3	0.1343		0.1907				
PC aa C40:4	0.0820						
PC aa C40:5	0.1297		-0.1070				
PC aa C40:6	0.1437		0.2042				
PC aa C42:0	0.0941						
PC aa C42:1	0.0913						
PC aa C42:2	0.1100		0.1563				
PC aa C42:4	0.0813						
PC aa C42:5	0.1356		0.1926				
PC aa C42:6	0.1255		0.1783				
PC ae C30:0	0.1629		-0.1344				
PC ae C30:2	0.1236		-0.1020				
PC ae C32:1	0.0879		-0.0725				
PC ae C32:2	0.0935		-0.0771				
PC ae C34:0	0.1385		-0.1143				
PC ae C34:1	0.1114		-0.0919				
PC ae C34:2	0.0870		-0.0718				
PC ae C34:3	0.1009		-0.0833				
PC ae C36:0	0.1073		0.1525				
PC ae C36:1	0.1203		-0.0993				
PC ae C36:2	0.1082		-0.0892				
PC ae C36:3	0.0840		-0.0693				
PC ae C36:4	0.0853						
PC ae C36:5	0.0898						
PC ae C38:2	0.0914		-0.0754				
PC ae C38:3	0.1032		-0.0851				
PC ae C38:4	0.0690						
PC ae C38:5	0.0700						
PC ae C38:6	0.1152		0.1637				

Continues...

Table 7.16 continued

Metabolites	Treelet component loadings at cut-level 108						
	TC1	TC2	TC3	TC4	TC5	TC6	TC7
PC ae C40:1	0.0926		0.1316				
PC ae C40:2	0.1176		-0.0970				
PC ae C40:3	0.0939		-0.0775				
PC ae C40:4	0.0597						
PC ae C40:5	0.0635						
PC ae C40:6	0.1037		0.1473				
PC ae C42:1	0.0782						
PC ae C42:2	0.0940		0.1335				
PC ae C42:3	0.0947		0.1346				
PC ae C42:4	0.0773						
PC ae C42:5	0.0633						
PC ae C44:4	0.0789						
PC ae C44:5	0.0824						
PC ae C44:6	0.0818						
Hexose							
Hexose							0.5234
Sphingolipids							
SM (OH) C14:1	0.1281		-0.1057				
SM (OH) C16:1	0.1188		-0.0980				
SM (OH) C22:1	0.0681						
SM (OH) C22:2	0.1054		-0.0869				
SM (OH) C24:1	0.0857						
SM C16:0	0.0503						
SM C16:1	0.0500						
SM C18:0	0.0683						
SM C18:1	0.0647						
SM C20:2							
SM C24:0	0.0608						
SM C24:1	0.0645						

[1] Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 108, i.e. $+3$ cut-levels of the main treelet transform

Table 7.17: Pearson's correlation coefficients between principal and treelet component scores[1]

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
TC1	0.99	-0.06	-0.01	-0.04	-0.06	0.01	0.02
TC2	0.12	0.85	0.32	0.22	-0.06	-0.13	0.03
TC3	-0.03	0.22	-0.53	0.06	-0.76	-0.14	-0.02
TC4	0.09	-0.20	0.07	0.52	0.02	-0.67	0.18
TC5	0.37	0.07	0.16	0.44	0.20	0.23	-0.47
TC6	-0.02	-0.76	-0.27	0.06	-0.02	0.21	0.22
TC7	0.10	0.26	-0.74	0.08	0.42	0.05	0.15

[1] Principal component analysis and treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components from each analysis

Table 7.18: Treelet component scores and risk of overall prostate cancer in 1,593 matched case-control sets in EPIC[1]

Components	Fifths of treelet component score								
	1	2		3		4		5		p_{trend}[2]
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	
TC1	Ref	1.26	(1.01 - 1.58)	1.39	(1.11 - 1.75)	1.11	(0.87 - 1.40)	1.08	(0.85 - 1.39)	0.8
TC2	Ref	0.90	(0.72 - 1.13)	0.74	(0.58 - 0.95)	0.78	(0.60 - 1.00)	0.86	(0.65 - 1.15)	0.2
TC3	Ref	0.87	(0.69 - 1.09)	0.75	(0.59 - 0.94)	0.95	(0.75 - 1.20)	0.95	(0.74 - 1.21)	>0.9
TC4	Ref	1.01	(0.80 - 1.27)	0.85	(0.67 - 1.08)	0.81	(0.63 - 1.04)	0.84	(0.66 - 1.07)	0.1
TC5	Ref	0.94	(0.75 - 1.18)	0.80	(0.63 - 1.02)	0.86	(0.68 - 1.09)	0.86	(0.67 - 1.12)	0.2
TC6	Ref	1.08	(0.84 - 1.39)	1.02	(0.78 - 1.32)	1.12	(0.86 - 1.46)	1.09	(0.83 - 1.44)	0.5
TC7	Ref	1.40	(1.11 - 1.77)	1.28	(1.01 - 1.63)	1.21	(0.94 - 1.54)	1.35	(1.05 - 1.74)	0.1

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the fifths of component scores

Table 7.19: Treelet component scores and risk of prostate cancer by time to diagnosis in 1,593 matched case-control sets in EPIC[1]

≤5 years to diagnosis, n = 501 cases

Components	Thirds of treelet component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
TC1	Ref	1.18	(0.85 - 1.65)	1.06	(0.76 - 1.48)	0.8
TC2	Ref	0.84	(0.60 - 1.19)	0.93	(0.63 - 1.36)	0.7
TC3	Ref	1.03	(0.75 - 1.41)	1.54	(1.08 - 2.20)	0.02
TC4	Ref	0.88	(0.61 - 1.28)	0.77	(0.56 - 1.07)	0.1
TC5	Ref	0.73	(0.53 - 1.00)	0.73	(0.51 - 1.06)	0.08
TC6	Ref	1.03	(0.73 - 1.46)	1.15	(0.79 - 1.67)	0.5
TC7	Ref	1.28	(0.93 - 1.78)	1.32	(0.94 - 1.85)	0.1

>5 years to diagnosis, n = 1,092

Components	Thirds of treelet component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
TC1	Ref	1.13	(0.92 - 1.40)	0.97	(0.77 - 1.21)	0.8	0.7
TC2	Ref	0.82	(0.66 - 1.04)	0.87	(0.67 - 1.13)	0.3	0.8
TC3	Ref	0.76	(0.60 - 0.95)	0.89	(0.70 - 1.14)	0.4	0.01
TC4	Ref	0.81	(0.65 - 1.01)	0.84	(0.66 - 1.07)	0.1	0.9
TC5	Ref	0.90	(0.72 - 1.12)	0.99	(0.77 - 1.26)	0.9	0.4
TC6	Ref	0.99	(0.78 - 1.24)	1.04	(0.80 - 1.34)	0.8	0.8
TC7	Ref	0.90	(0.72 - 1.11)	0.99	(0.79 - 1.25)	0.9	0.2

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component scores

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and time to diagnosis ($\leq$5 versus $>$5 years)

Table 7.20: Treelet component scores and risk of prostate cancer by grade in 1,224 matched case-control sets in EPIC[1]

Low-intermediate grade, n = 1,041 cases

Components	Thirds of treelet component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
TC1	Ref	1.25	(1.00 - 1.55)	0.99	(0.78 - 1.25)	>0.9
TC2	Ref	0.77	(0.61 - 0.98)	0.82	(0.63 - 1.07)	0.1
TC3	Ref	0.83	(0.66 - 1.04)	1.05	(0.83 - 1.35)	0.7
TC4	Ref	0.76	(0.61 - 0.96)	0.83	(0.66 - 1.06)	0.1
TC5	Ref	0.95	(0.75 - 1.18)	1.01	(0.78 - 1.30)	1.0
TC6	Ref	0.96	(0.76 - 1.23)	0.85	(0.65 - 1.11)	0.2
TC7	Ref	1.00	(0.80 - 1.25)	1.02	(0.81 - 1.28)	0.9

High grade, n = 183 cases

Components	Thirds of treelet component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
TC1	Ref	1.03	(0.57 - 1.86)	0.47	(0.26 - 0.86)	0.01	0.08
TC2	Ref	0.48	(0.24 - 0.95)	0.46	(0.22 - 0.96)	0.07	0.6
TC3	Ref	0.86	(0.46 - 1.60)	0.90	(0.46 - 1.76)	0.8	0.6
TC4	Ref	0.97	(0.52 - 1.82)	0.63	(0.34 - 1.16)	0.1	0.6
TC5	Ref	0.67	(0.37 - 1.23)	0.60	(0.31 - 1.15)	0.1	0.1
TC6	Ref	1.04	(0.58 - 1.85)	2.33	(1.22 - 4.43)	0.02	0.01
TC7	Ref	1.17	(0.63 - 2.20)	1.06	(0.57 - 1.99)	0.8	0.4

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component scores

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and grade (low-intermediate versus high grade)

Table 7.21: Treelet component scores and risk of prostate cancer by stage (localised and advanced) in 1,002 matched case-control sets in EPIC[1]

Localised, n = 702 cases

Components	Thirds of treelet component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
TC1	Ref	1.64	(1.24 - 2.17)	1.27	(0.95 - 1.70)	0.1
TC2	Ref	0.86	(0.65 - 1.14)	0.92	(0.67 - 1.26)	0.6
TC3	Ref	0.86	(0.65 - 1.15)	1.20	(0.88 - 1.63)	0.2
TC4	Ref	0.68	(0.51 - 0.91)	0.89	(0.66 - 1.19)	0.4
TC5	Ref	1.03	(0.79 - 1.36)	1.05	(0.77 - 1.44)	0.7
TC6	Ref	0.94	(0.71 - 1.25)	0.89	(0.65 - 1.22)	0.5
TC7	Ref	1.10	(0.83 - 1.45)	1.06	(0.79 - 1.43)	0.7

Advanced, n = 300 cases

Components	Thirds of treelet component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
TC1	Ref	0.73	(0.47 - 1.12)	0.48	(0.31 - 0.74)	0.0009	0.0002
TC2	Ref	0.52	(0.32 - 0.86)	0.52	(0.30 - 0.93)	0.04	0.3
TC3	Ref	0.75	(0.50 - 1.13)	0.79	(0.50 - 1.26)	0.3	0.2
TC4	Ref	0.68	(0.42 - 1.10)	0.54	(0.34 - 0.85)	0.008	0.06
TC5	Ref	0.72	(0.47 - 1.11)	0.85	(0.51 - 1.43)	0.4	0.2
TC6	Ref	1.07	(0.67 - 1.70)	1.02	(0.61 - 1.69)	0.9	0.6
TC7	Ref	0.96	(0.61 - 1.53)	1.28	(0.81 - 2.03)	0.3	0.6

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; 10 $-$ 19; 20 $-$ 39; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component scores

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and stage (localised versus advanced stage)

Table 7.22: Treelet component scores and risk of prostate cancer by stage (non-aggressive and aggressive) in 1,002 matched case-control sets in EPIC[1]

Non-aggressive, n = 821 cases

Components	Thirds of treelet component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
TC1	Ref	1.38	(1.08 - 1.78)	1.02	(0.79 - 1.33)	0.8
TC2	Ref	0.79	(0.61 - 1.03)	0.85	(0.63 - 1.14)	0.3
TC3	Ref	0.85	(0.66 - 1.10)	1.10	(0.83 - 1.46)	0.4
TC4	Ref	0.59	(0.45 - 0.78)	0.74	(0.56 - 0.97)	0.03
TC5	Ref	0.94	(0.73 - 1.21)	1.01	(0.76 - 1.34)	>0.9
TC6	Ref	0.99	(0.76 - 1.29)	0.91	(0.68 - 1.22)	0.5
TC7	Ref	1.14	(0.89 - 1.47)	1.05	(0.81 - 1.37)	0.7

Aggressive, n = 181 cases

Components	Thirds of treelet component score						
	1	2		3		p_{trend}[2]	p_{het}[3]
		OR	95% CI	OR	95% CI		
TC1	Ref	0.78	(0.43 - 1.40)	0.64	(0.36 - 1.14)	0.1	0.1
TC2	Ref	0.59	(0.30 - 1.14)	0.67	(0.32 - 1.38)	0.4	>0.9
TC3	Ref	0.79	(0.45 - 1.39)	0.84	(0.46 - 1.53)	0.6	0.5
TC4	Ref	1.25	(0.68 - 2.30)	1.01	(0.57 - 1.81)	>0.9	0.6
TC5	Ref	0.83	(0.49 - 1.40)	0.71	(0.36 - 1.38)	0.3	0.2
TC6	Ref	0.91	(0.52 - 1.61)	1.09	(0.56 - 2.10)	0.9	0.9
TC7	Ref	0.75	(0.40 - 1.42)	1.59	(0.83 - 3.05)	0.2	0.6

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake ($<$ 10; $10-19$; $20-39$; $\geq$ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Computed across the medians of the thirds of component scores

[3] Likelihood ratio χ^2 test comparing models with and without an interaction term between the linear trend variable and stage (non-aggressive versus aggressive)

Table 7.23: Treelet component scores and risk of death from prostate cancer in 172 matched case-control sets in EPIC[1,2]

Components	Thirds of treelet component score					
	1	2		3		p_{trend}[2]
		OR	95% CI	OR	95% CI	
TC1	Ref	1.07	(0.60 - 1.92)	0.83	(0.45 - 1.54)	0.5
TC2	Ref	0.87	(0.44 - 1.75)	1.09	(0.54 - 2.20)	0.7
TC3	Ref	0.70	(0.36 - 1.34)	0.67	(0.32 - 1.39)	0.3
TC4	Ref	0.99	(0.54 - 1.82)	0.83	(0.45 - 1.54)	0.6
TC5	Ref	0.69	(0.38 - 1.27)	0.50	(0.24 - 1.06)	0.07
TC6	Ref	1.23	(0.70 - 2.17)	1.91	(0.99 - 3.69)	0.07
TC7	Ref	0.88	(0.46 - 1.71)	1.20	(0.64 - 2.26)	0.6

[1] Multivariable-adjusted model: adjusted for exact age (continuously), BMI (fourths; unknown), smoking (never; past; current; unknown), alcohol intake (< 10; $10 - 19$; $20 - 39$; ≥ 40 g/d; unknown), education (primary or none; secondary; degree level; unknown) and marital status (married or cohabiting; not married or cohabiting; unknown)

[2] Death from prostate cancer: prostate cancer listed as the underlying cause of death on the death certificate during follow-up; 22 were excluded from further analysis as their matched control had died before them (n $= 18$) or vital status was unknown for the control (n $= 4$)

[3] Computed across the medians of the thirds of component scores

7.8 FIGURES

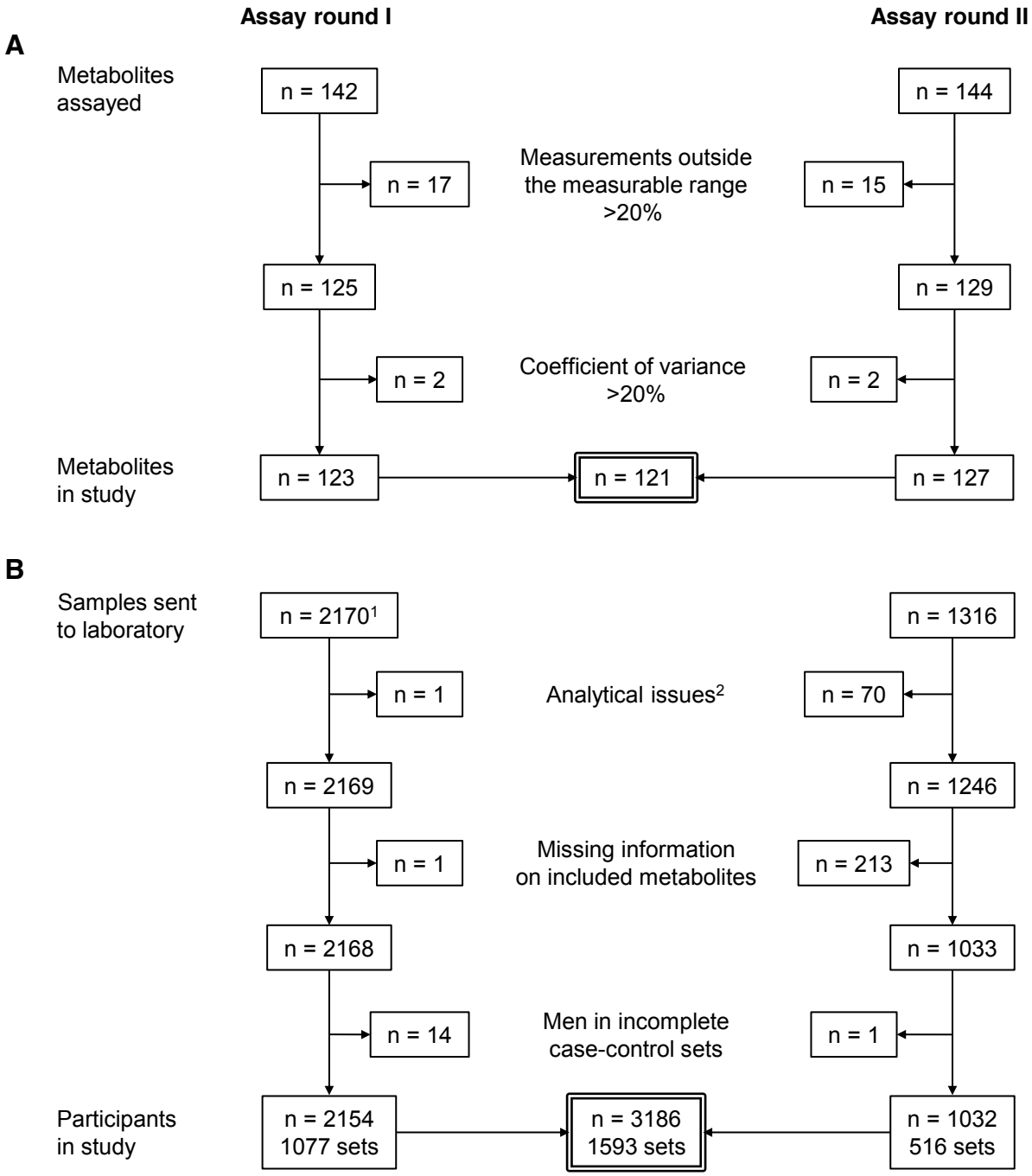


Figure 7.1: Flowchart of exclusions of metabolites and participants

A. Participants. B. Metabolites.

[1] There were 2,162 samples but eight samples contributed to the dataset twice because three controls became cases and five controls were controls for two cases each.

[2] Assay round I: too little blood volume; assay round II: data were missing for an entire batch.

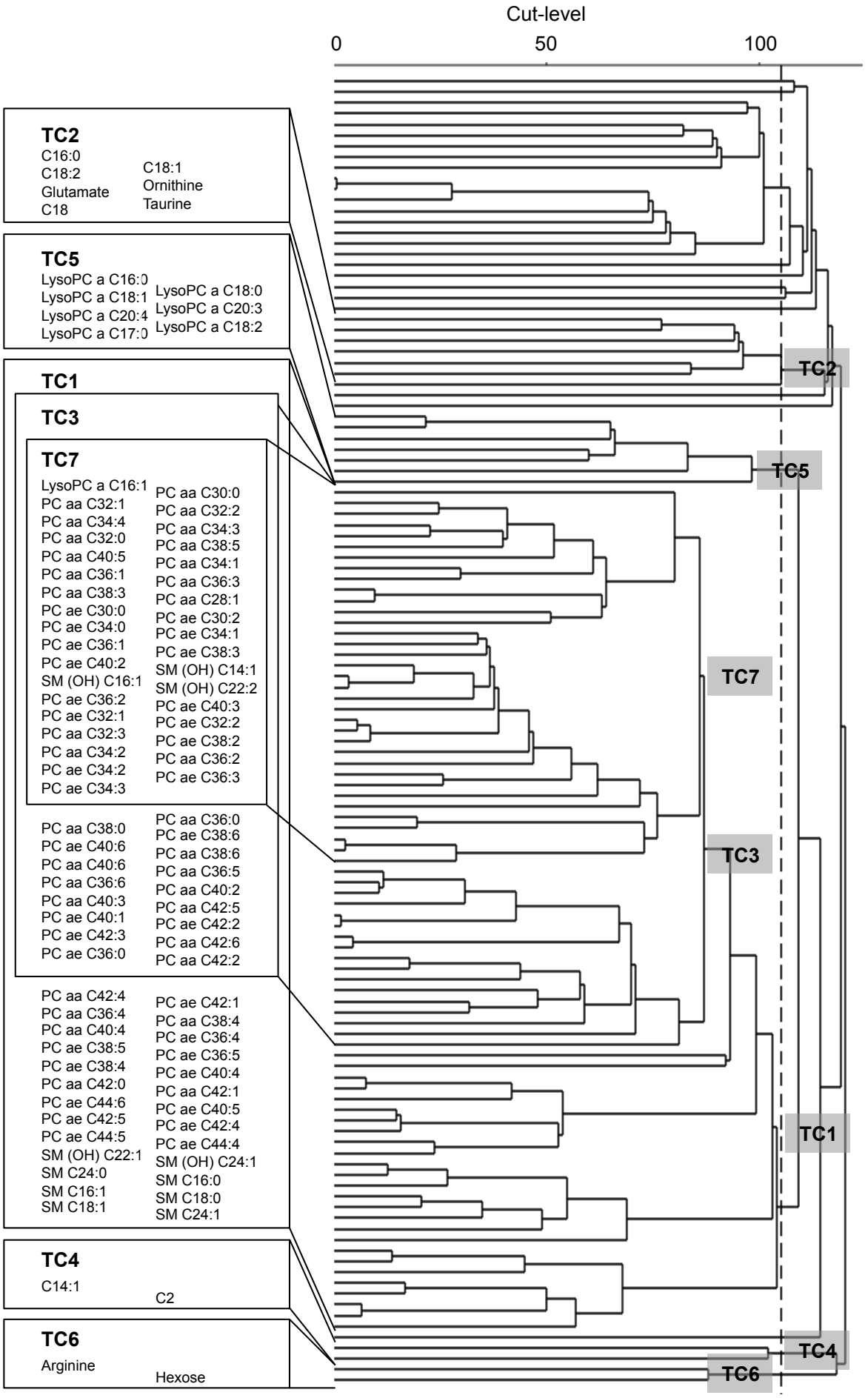


Figure 7.2: Cluster tree for treelet transform

Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 105.

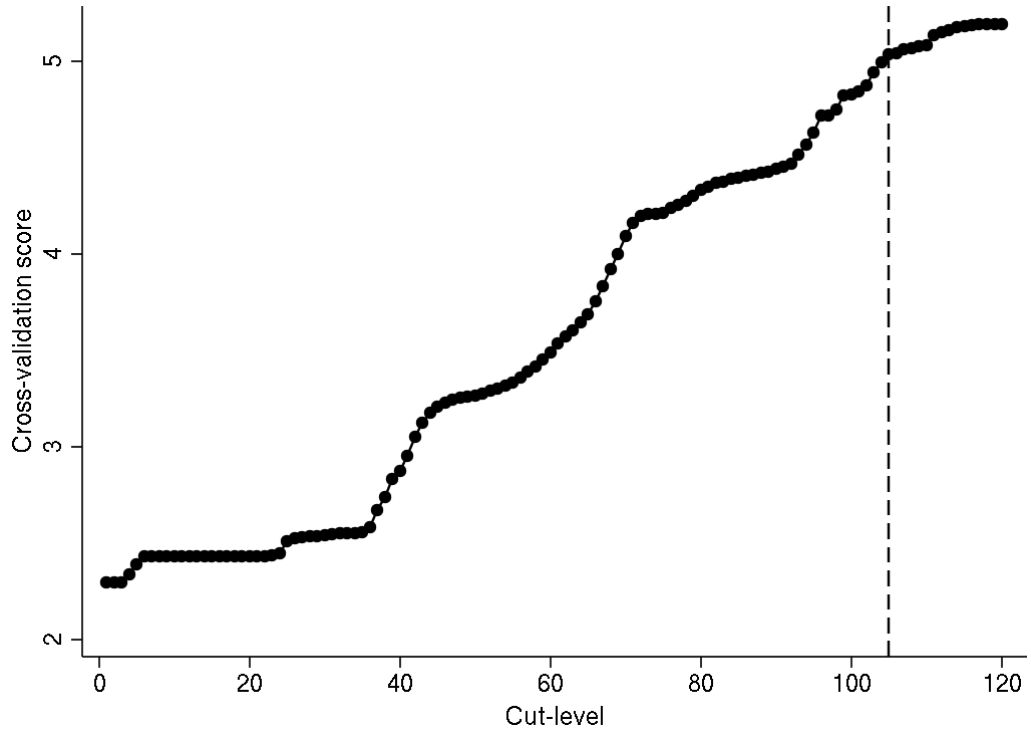


Figure 7.3: Cross-validation of cut-level for treelet transform

Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites. The dashed line indicates the optimal cut-level for the seven retained treelet components.

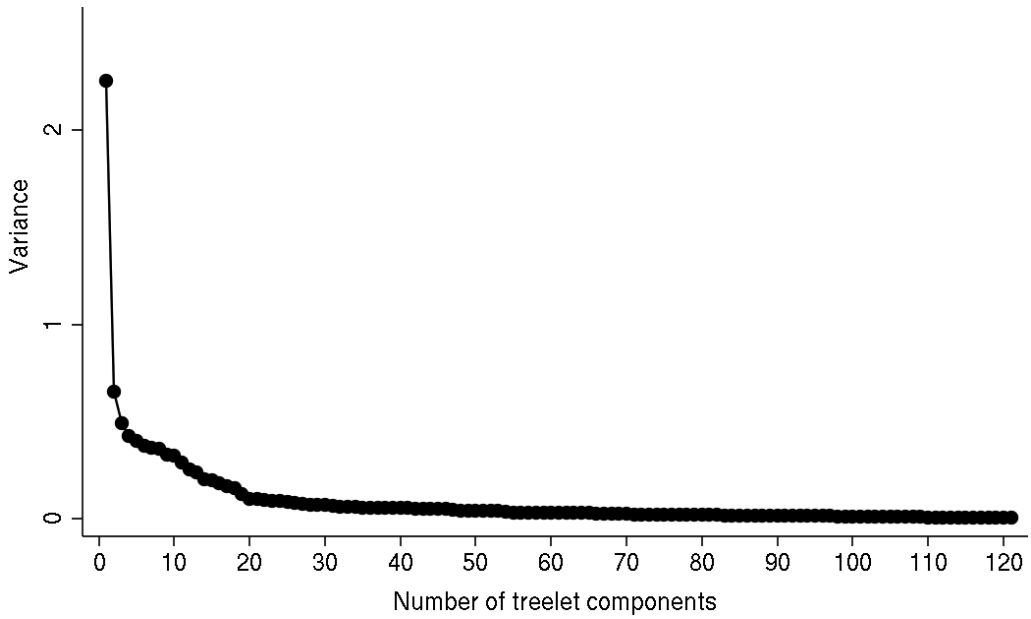


Figure 7.4: Scree plot for treelet transform

Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 105.

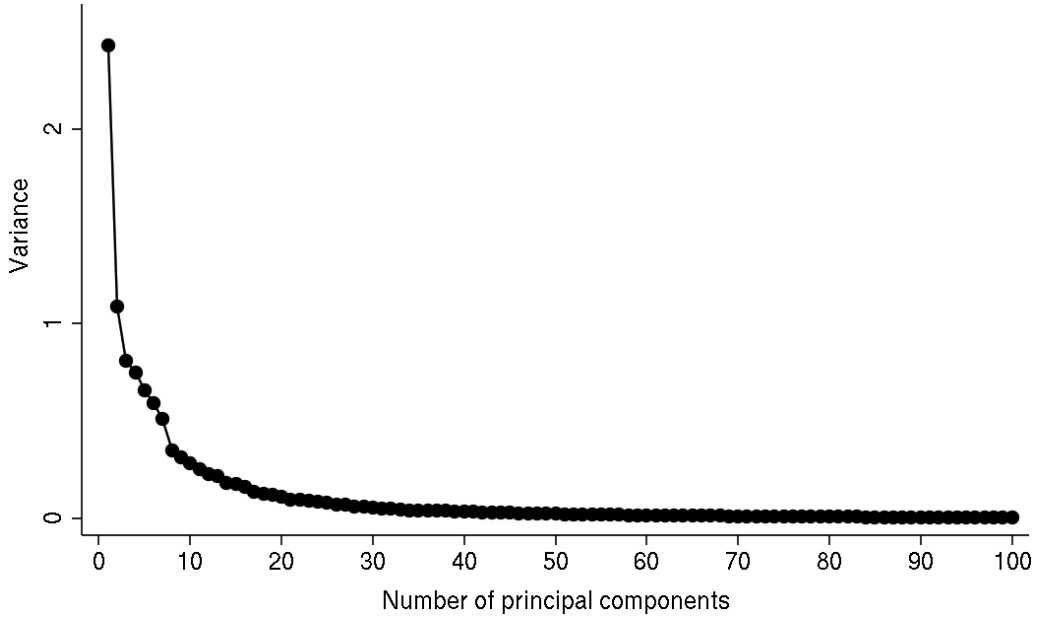


Figure 7.5: Scree plot for principal component analysis

Principal component analysis using the covariance matrix of ln-transformed concentrations of 121 metabolites.

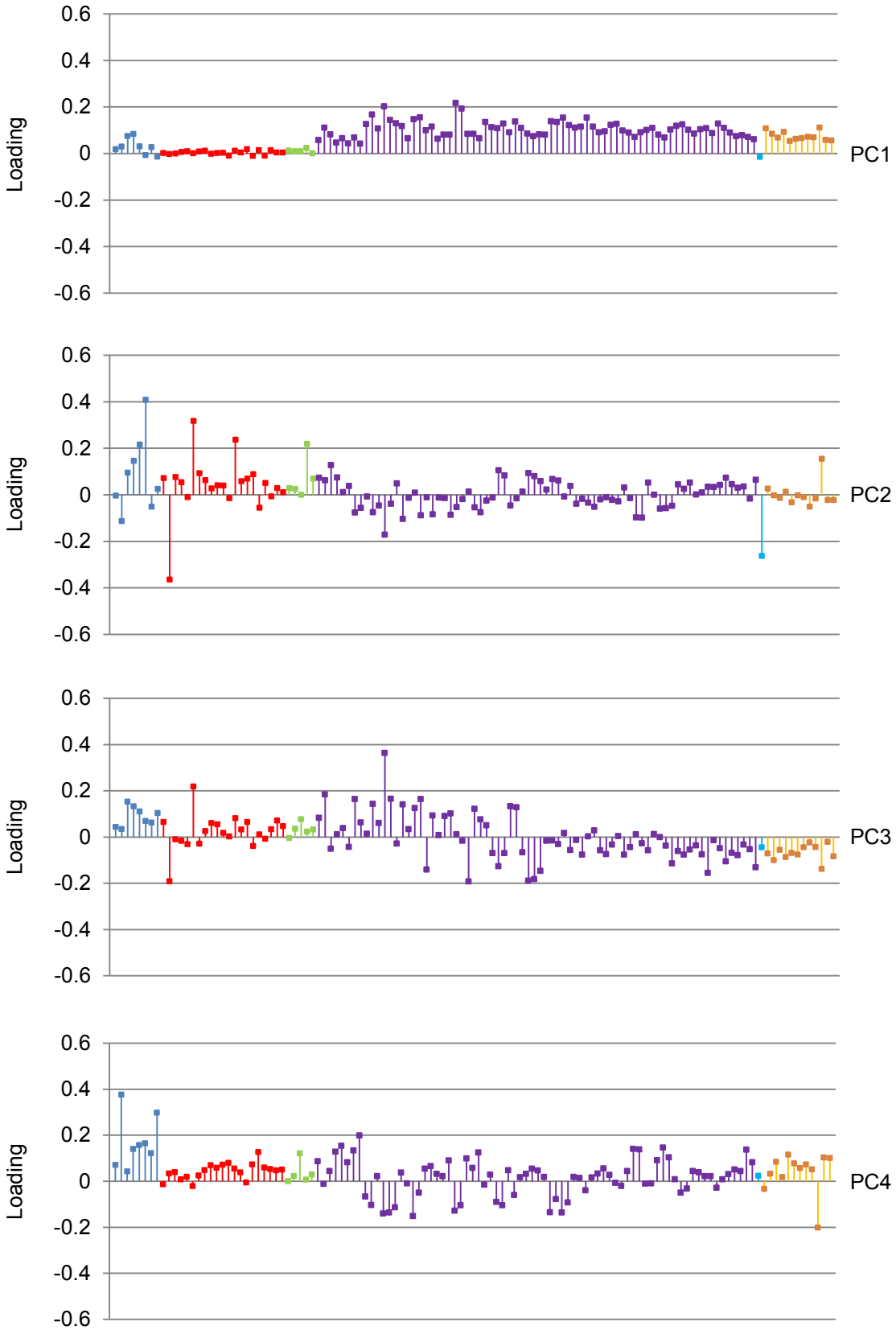


Figure 7.6: Loadings for the first seven principal components

Continues...

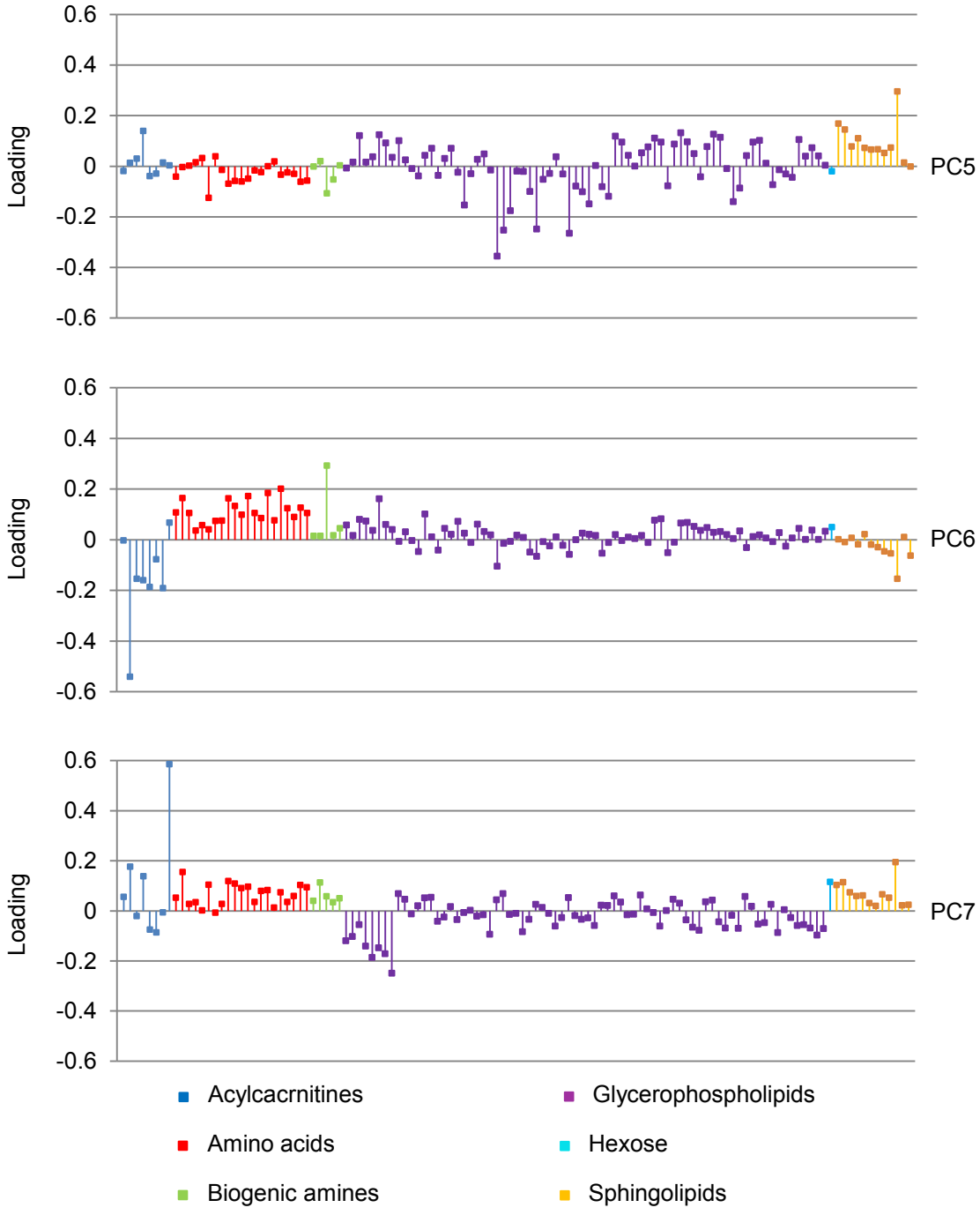


Figure 7.6 on the preceding page continued

Principal component analysis using the covariance matrix of ln-transformed concentrations of 121 metabolites.

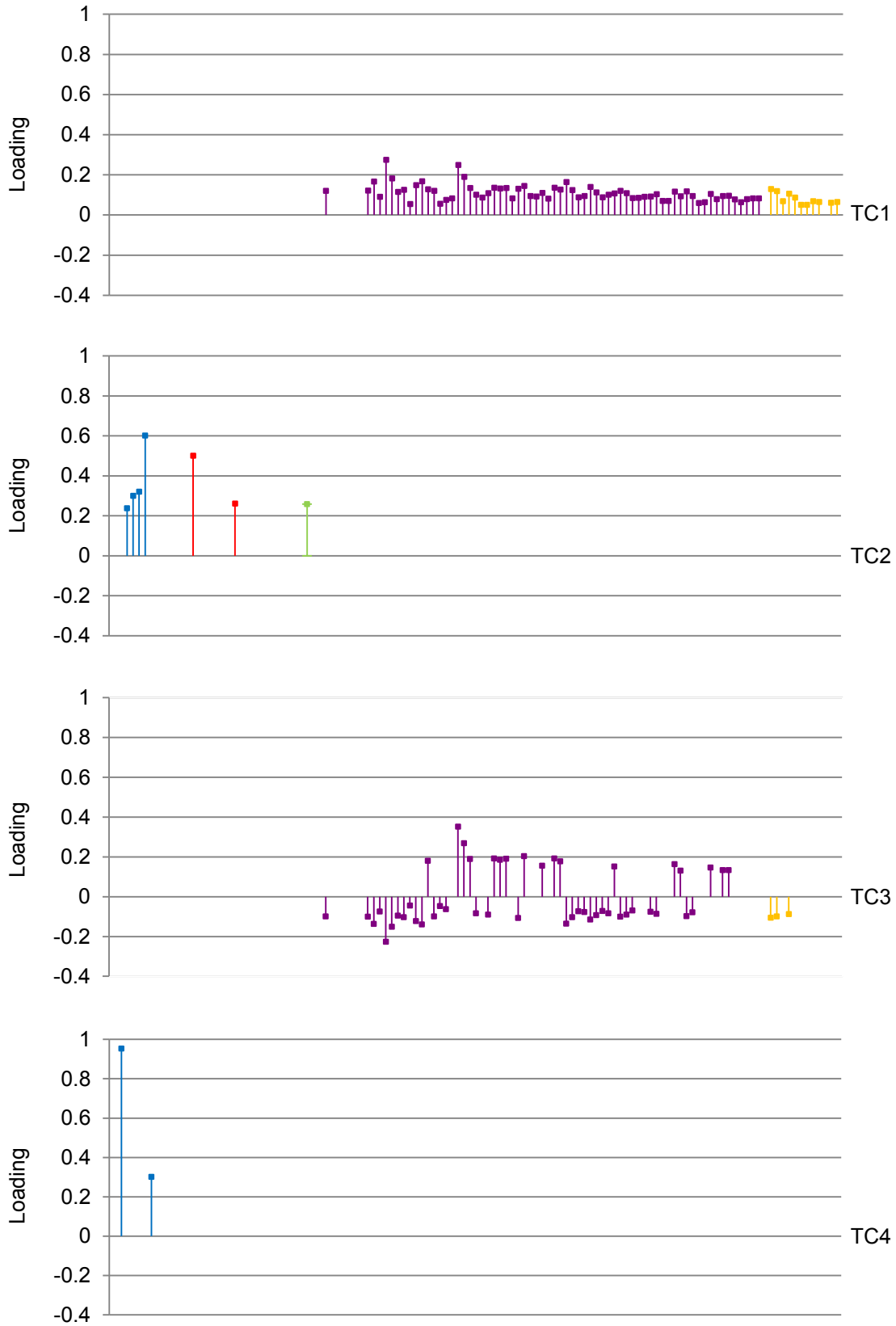


Figure 7.7: Loadings for the first seven treelet components

Continues...

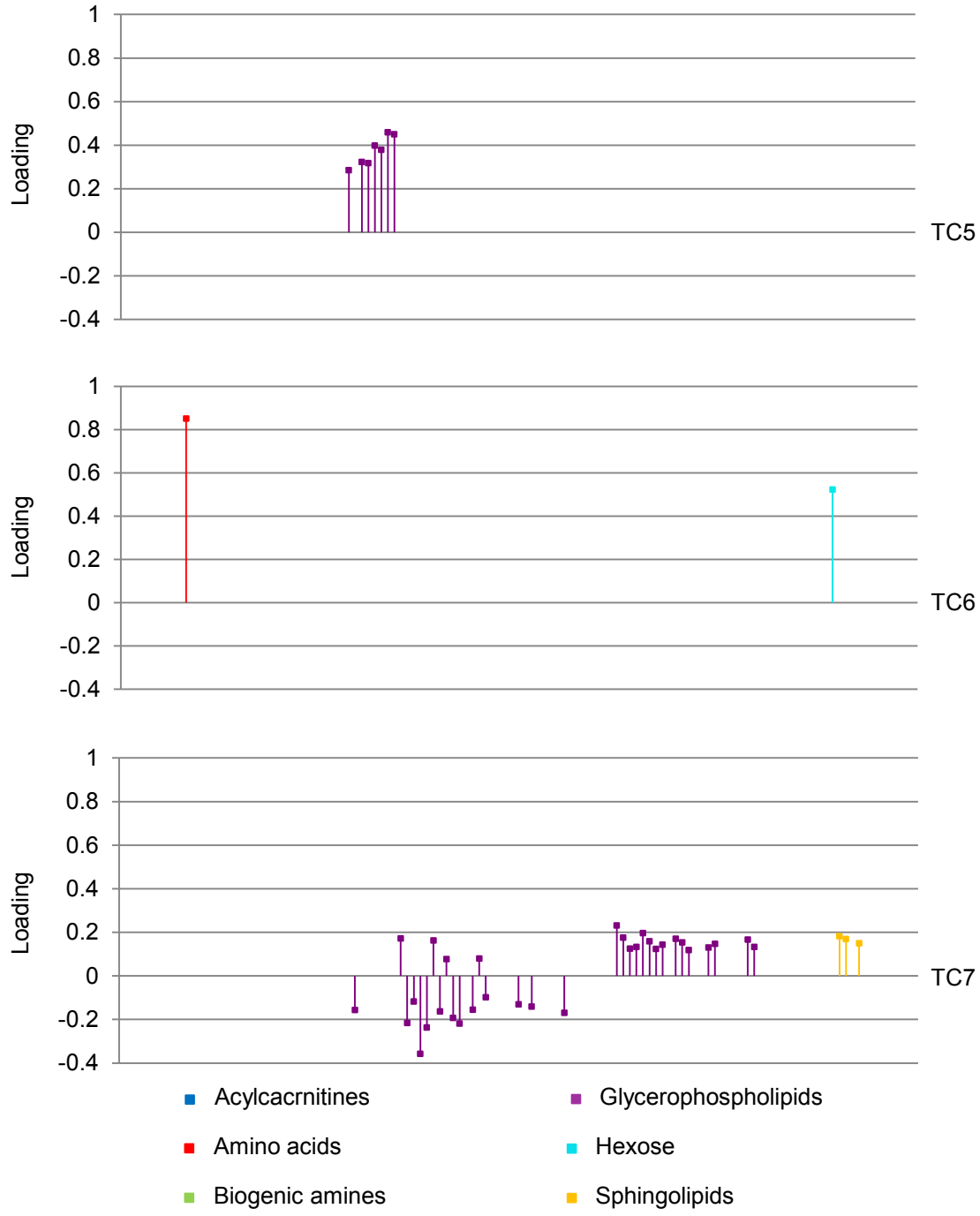


Figure 7.7 on the previous page continued

Treelet transform using the covariance matrix of ln-transformed concentrations of 121 metabolites and retaining seven components at cut-level 105.

8. Intake of protein and amino acid concentrations

8.1 Introduction

Protein from various dietary sources, for example dairy products and plants, might be associated differently with circulating concentration of insulin-like growth factor-I (IGF-I), which is an established risk factor for prostate cancer [2]. Published evidence suggests that intake of total protein [6, 7, 75, 76], animal protein [6, 7, 75, 224] and dairy protein [6, 7, 75, 76] but not other sub-types of animal protein [6] might be positively associated with circulating IGF-I concentration. For intake of plant protein, no associations [6, 7] or positive associations have been reported [75, 224]. While it is not clear if these associations are mediated by differences in circulating concentrations of amino acids, experimental research has found that higher intake of essential amino acids increases circulating IGF-I concentration [86, 87, 88].

In the current chapter and Chapter 9, the associations between protein intake from specific sources, and circulating concentrations of amino acids and IGF-I will be investigated; the current chapter focuses on the former association. Taken together, this work might help to inform the development of strategies for prevention of IGF-I-related cancers, including prostate cancer.

8.1.1 Previous studies

Limited observational data are available on the association between protein intake and circulating concentration of amino acids. To my knowledge, only one small study ($n = 32$) [77] and the work presented in Chapter 4 (Section 4.4.5.2 on page 41; [17]) have been published. Both of these analyses were conducted in samples including men from different diet groups defined by the intake of animal protein, e.g. meat-eaters, vegetarians and vegans. Anderson and Blendis [77] reported positive correlations of total habitual protein intake with plasma concentrations of isoleucine, leucine, valine, methionine, phenylalanine and tyrosine, and inverse correlations with the ratio of tryptophan to the sum of tryptophan, tyrosine, phenylalanine and branched-chain amino acids (leucine, isoleucine and valine), and with the ratio of tryptophan to branched-chain amino acids. After correction for multiple testing, we found inverse associations between intakes of non-soya plant protein and concentrations of lysine and tyrosine, and between soya protein intake and concentrations of valine, lysine, methionine and tyrosine, while a positive correlation was found between soya protein intake and glycine concentration (Table 4.7 on page 52).

Indirect evidence of an association between intake of protein and amino acid concentrations is available from analyses comparing plasma amino acid concentrations in individuals by diet groups, e.g. meat-eaters and vegans. Together, such studies have found lower concentrations

of most essential amino acids, i.e. isoleucine, leucine, valine, histidine, lysine, phenylalanine, threonine and tryptophan, and of other amino acids, such as arginine, tyrosine, citrulline and trans-4-hydroxyproline (t4-OH-Pro), in vegans than in meat-eaters [77, 134, 152]. In line with this, we found significant differences across the four EPIC-Oxford diet groups in mean concentrations of lysine, methionine, tryptophan and tyrosine, for which vegans had the lowest concentrations (Section 4.4.4.1 on page 40). We also found significant differences for alanine and glycine, for which vegans had higher concentrations than meat-eaters.

8.2 Aim

The aim of the current chapter was to investigate associations of intake of dietary protein from different sources, i.e. total protein, animal protein, dairy protein and plant protein, with plasma amino acid concentrations, sums and ratios in men from six countries who are participants in the European Prospective Investigation into Cancer and Nutrition (EPIC).

8.3 Methods

8.3.1 Study population

The EPIC study is a prospective cohort that was established with the aim of investigating associations between diet, lifestyle and cancer risk in men and women from ten European countries; details of the cohort have been described previously (Section 3.1 on page 21).

All participants gave written informed consent to participate in the EPIC cohort. Approval for the current study was obtained from the Internal Review Board of the International Agency for Research on Cancer (IARC), Lyon, France (Project No. 14-09) and participating centres as required.

Participants were eligible for the current cross-sectional analysis if they were control participants in the nested case-control study of metabolite patterns and risk of prostate cancer (Chapter 7). These were 1,699 men from 15 centres in Germany, Greece, Italy, Netherlands, Spain and United Kingdom, with no cancer (except non-melanoma skin cancer) at blood collection and at diagnosis of their index case. Men with no data on protein intake ($n = 2$) were excluded, leaving 1,697 men for further analysis.

8.3.2 Dietary data

Details of the collection of dietary data are provided in Section 3.2.2.1 on page 23. In brief, information about dietary intake over the past year was collected at recruitment using validated,

country-specific questionnaires designed to capture local dietary habits [99, 101, 102, 103, 104, 105, 106, 107].

In the current analysis, animal protein refers to protein from meat and meat products, fish and shellfish, egg and egg products, and dairy products [6]. Dairy protein refers to protein from milk, milk beverages, yogurt and cheese. Plant protein refers to total protein minus animal protein.

8.3.3 Blood collection and laboratory analysis

A standardised protocol for blood collection and processing was followed (see Section 3.2.1 on page 23 for details)

All plasma samples (with citrate as an anticoagulant) were assayed at IARC, using the Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbrcuk, Austria) and following the procedure recommended by the vendor. As described in Section 7.3.2 on page 212, blood samples were assayed in two rounds. For both rounds, a triple quadrupole mass spectrometer (Triple Quad 4500; AB Sciex, Framingham, Massachusetts, USA) was used to quantify 22 amino acids; these were separated by ultra-high performance liquid chromatography before injection into the mass spectrometer.

Each assay batch included quality control samples from one plasma pool and laboratory personnel were blinded to the sample category, i.e. study samples or quality control samples.

Information of completeness and coefficients of variation (CV) for the metabolomics assay are presented in Tables 6.1 on page 135 and 7.1 on page 230 for each of the two assay rounds, respectively. Across the two assay rounds, CVs for amino acids ranged from 6.0% to 18.7%, and all amino acids, except aspartate, had none or very few non-quantifiable measurements (i.e. missings or below the limit of quantification (LOQ); $< 0.7\%$). For aspartate, 22.4% and 32.6% of samples had concentrations below the LOQ in assay round I and II, respectively. The procedure for dealing with measurements below the LOQ is described in Section 7.3.4.1 on page 213. All 22 measured amino acid were included in the analysis.

8.3.4 Statistical analysis

8.3.4.1 Data preparation

Seven sum variables and five ratios of amino acids were calculated. The sum variables were: branched-chain amino acids (sum of isoleucine, leucine and valine), essential amino acids that cannot be synthesised endogenously [128] (sum of branched-chain amino acids, histidine, lysine, methionine, phenylalanine, threonine and tryptophan), non-essential amino acids (sum of alanine, arginine, asparagine, aspartate, glutamate, glutamine, glycine, proline, serine and tyrosine),

non-standard amino acids that are synthesised from other amino acids after incorporation into proteins [137] (sum of citrulline, ornithine and t4-OH-Pro), glucogenic amino acids that are broken down to form glucose (sum of isoleucine, valine, histidine, methionine, phenylalanine, threonine, tryptophan, alanine, arginine, asparagine, aspartate, glutamate, glutamine, glycine, proline, serine and tyrosine) [225], ketogenic amino acids that are broken down to form ketone bodies (sum of isoleucine, leucine, lysine, phenylalanine, tryptophan, threonine and tyrosine) [225], and ureagenic amino acids that are involved in the urea cycle (sum of arginine, citrulline and ornithine) [137].

Four ratios of single amino acids were calculated and these partly reflect the activity of the enzyme responsible for the conversion of the amino acid in the denominator to the amino acid in the numerator. These ratios were: citrulline to arginine ratio (enzyme: nitric oxide synthase (NOS) [226]; nitric oxide can promote carcinogenesis and variants in the NOS3 gene coding for NOS might be related to prostate cancer risk [227]), citrulline to ornithine ratio (enzyme: ornithine carbamoyltransferase [226]), ornithine to arginine ratio (enzyme: arginase [228], which may protect prostate tumour cells from immune response [229]), and tyrosine to phenylalanine ratio (phenylalanine hydrolase). The Fischer ratio, an indicator of liver damage [230], was also calculated (the ratio of branched-chain amino acids to the sum of phenylalanine, tryptophan and tyrosine).

All amino acid variables were logarithmically-transformed to approximate the normal distribution and log-transformed variables were used for all analyses. Intakes of protein, expressed as percentage of energy, were not transformed.

8.3.4.2 Participant characteristics

Participant characteristic, selected nutrient intakes and factors related to blood collection were presented as means or percentages for continuous and categorical variables, respectively, adjusted for study centre by thirds of total protein intake (% of energy).

8.3.4.3 Potential confounding factors

For all the following analyses, a minimally-adjusted model, including energy intake (fourths of the distribution, i.e. $< 7.87; 7.87 - 9.59; 9.60 - 11.40; > 11.40$ MJ), study centre and metabolite batch, was fitted to account for variation in participants' volume of food intake and to account for any variation in measurement of amino acid concentration which might arise from differences in blood handling between centres and any batch effects.

Furthermore, two multivariable-adjusted models were fitted; the first further adjusting for age group ($< 55; 55 - 59; 60 - 64; 65 - 69; \geq 70$ years) and body mass index (BMI; fourths of the distribution, i.e. $< 24.6; 24.6 - 26.5; 26.6 - 28.8; > 28.8$ kg/m^2; unknown), and the second

adjusting yet further for fasting status (no; yes ($>$ 6 hours since last meal); unknown) and time of day (6:00 − 8:59; 9:00 − 11:59; 12:00 − 14:59; 15:00 − 20:00; unknown) at blood collection. Results of the two multivariable-adjusted models did not materially differ, so only results from the former, including energy intake, study centre, metabolite batch, age and BMI, are shown.

8.3.4.4 Correlation analyses

Partial correlations between intakes of protein (total, animal, dairy and plant protein), and concentrations and ratios of amino acid and were computed.

8.3.4.5 Amino acid concentrations and ratios by fifths of protein intake

Adjusted geometric mean amino acid concentrations and ratios were calculated by fifths of protein intakes.

Multiple linear regression models of amino acid concentrations and ratios, and protein intakes were fitted to calculate p-values for linear trend and the percentage difference in amino acid concentrations and ratios for an 80 percentile increment in protein intake. In this regression, the fifths were re-scored as 0.00, 0.25, 0.50, 0.75 and 1.00 and treated as a continuous variable; a unit increase in this variable represents an 80 percentile increase in protein intake because the mid-points of the bottom and top fifths are the 10th and 90th percentiles of the protein intake, respectively [2]. The percentage increase in amino acid concentrations and ratios for an 80 percentile increment in protein intake, was calculated as $(e^{\beta} - 1) \times 100$, in which β is the regression coefficient [231].

8.3.4.6 Statistical significance

All tests for statistical significance were two-sided. The Benjamini–Hochberg false discovery rate (FDR)-controlling procedure was used to account for multiple testing [154] with $\alpha = 0.05$. This allows 5% of statistical significant findings to be false, on average. For further details on this method, please refer to Section 6.3.4.5 on page 124.

8.3.4.7 Statistical power

With a sample size of 1,697 men, we had more than 80% power to detect a difference in amino acid concentration for an 80 percentile increase in protein intake of approximately 6 to 9%.

8.3.4.8 Statistical software

All statistical analyses were conducted in Stata Statistical Software Package, version 14 (Stata Corporation, College Station, Texas, USA). Figures were drawn using R: A language and envir-

onment for statistical computing, version 3.3.3 (Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/). The power calculation was done in Quanto, version 1.2.4.

8.4 Results

8.4.1 Participant characteristics

On average, participants were 60 years of age at blood collection. Men with higher total protein intakes (% of energy) intake were heavier (4.0 kg difference) and had higher BMI (1.3 kg/m^2 difference) than men with lower protein intakes (Table 8.1 on page 285). Men consuming higher percentages of energy from total protein also consumed higher proportions of protein from each source, especially animal protein (5.9 percentage points of energy difference), and this was coupled with lower total energy intakes (−1.7 MJ difference). Furthermore, men in the top third of total protein intake drank less alcohol, were less physically active, were less educated and were more likely to have fasted prior to blood collection.

8.4.2 Intake of protein types and amino acid concentrations

For the associations between intake of protein types and amino acid concentrations, results from the multivariable-adjusted model (including energy intake, study centre, metabolite batch, age and BMI) were in general similar or slightly weaker than those from the minimally-adjusted model (not including age and BMI). The correlation coefficients were attenuated by no more than 0.03. Results of the linear regressions were somewhat weakened for glutamate ($\leq$ 4.7 percentage points for the percentage difference in glutamate concentration per an 80 percentile increment in protein intakes). Results for other amino acids were attenuated to a lesser extent ($\leq$ 2.3, $\leq$ 2.4, $\leq$ 0.5, $\leq$ 0.7 percentage points for total, animal, dairy and plant protein intakes, respectively). Results for the minimally-adjusted model are therefore presented in table format only (Tables 8.2 on page 286 to 8.10 on page 299), while results from the multivariable-adjusted model are described below.

Associations between protein intakes and selected individual amino acids are shown in Figures 8.1 on page 314 to 8.4 on page 317.

8.4.2.1 Total protein intake

Correlations: Positive correlations of total protein intake (% of energy) with concentrations of leucine, valine, the sum of branched-chain amino acids, lysine, the sum of essential amino acids, t4-OH-Pro and the Fischer ratio were observed, after taking multiple testing into account ($0.07 \leq r \leq 0.12$; $1 \times 10^{-6} \leq p \leq 0.009$; Table 8.11 on page 300). In contrast, inverse correlations

were found for total protein intake with glycine, citrulline, and its ratios with arginine and ornithine ($-0.09 \leq r \leq -0.06$; $0.0002 \leq p \leq 0.01$).

Additionally, correlations between total protein intake and isoleucine (positive), non-essential amino acids (inverse) and ketogenic amino acids (positive) were found, at a conventional significance level ($p < 0.05$).

Amino acid concentrations and ratios by fifths of total protein intake: The mean amino acid concentrations and ratios by fifths of total protein intake (% of energy) are shown in Table 8.12 on page 302, and the percentage increase in amino acids per an 80 percentile increase in total protein intake is shown in Table 8.13 on page 304. Concentration of glycine and t4-OH-Pro varied significantly by fifths of total protein intake ($p = 0.0009$ and 0.0001, respectively), with an inverse and positive linear trend, respectively, after correcting for multiple testing (Figures 8.3 on page 316 and 8.4 on page 317, A panels). For an 80 percentile increment in total protein intake, glycine decreased by 5.6% (95% confidence interval (CI): $2.5 - 8.6\%$, $p_{trend} = 0.0005$) and t4-OH-Pro increased by 15.9% (95%CI: $8.2 - 24.1\%$, $p_{trend} = 2 \times 10^{-5}$).

At a conventional significance level, leucine, valine (Figure 8.1 on page 314, panel A) and sum of branched-chain amino acids varied by fifths of total protein intake with evidence of positive linear associations. Moreover, the sum of non-essential amino acids and the Fischer ratio varied by fifths of total protein intake, while lysine, the sum of essential amino acids, the sum of ketogenic amino acids (all positively), citrulline, and its ratios with arginine and ornithine (all inversely) were linearly associated with total protein intake, all at a conventional significance level.

8.4.2.2 Animal protein intake

Results for intake of animal protein were in general in line with those for total protein intake.

Correlations: Concentrations of leucine, valine, the sum of branched-chain amino acids, lysine, the sum of essential amino acids, t4-OH-Pro and the Fischer ratio were positively correlated with intake of protein from animal sources (% of energy; $0.06 \leq r \leq 0.15$; $7 \times 10^{-10} \leq p \leq 0.02$; Table 8.11 on page 300), while alanine, asparagine, glutamine, glycine, the sum of non-essential amino acids, citrulline and the citrulline to arginine ratio were inversely correlated with animal protein intake ($-0.12 \leq r \leq -0.06$; $2 \times 10^{-6} \leq p \leq 0.02$), after controlling the FDR.

Moreover, correlations of animal protein intake with isoleucine (positively), the glucogenic (inversely) and ketogenic (positively) amino acid concentrations, and the citrulline to ornithine ratio (inversely), were observed at a conventional significance level.

Amino acid concentrations and ratios by fifths of animal protein intake: As for total protein, concentrations of glycine and t4-OH-Pro varied by fifths of animal protein intake (% of energy; $p_{diff} = 4 \times 10^{-5}$ and 2×10^{-6}, respectively; Table 8.14 on page 305), with an inverse and positive linear trend, respectively, after multiple testing was taken into account (Table 8.15 on page 307; Figures 8.3 on page 316 and 8.4 on page 317, B panels). There was also evidence of linear associations of animal protein intake and leucine, valine, the sum of branched-chain amino acids (all positive; Figure 8.1 on page 314, panel B), the sum of non-essential amino acids, citrulline (Figure 8.4 on page 317, panel B) and its ratio with arginine (all inverse), after correction for multiple testing. For these, the differences across fifths of animal protein intake were however, significant only at a conventional significance level. The strongest linear association was with t4-OH-Pro, which differed by 21.6% (95%CI: 13.5−30.2%; $p_{trend} = 3 \times 10^{-8}$) per an 80 percentile increase in intake of animal protein.

Additionally, at a conventional significance level, isoleucine, the Fischer ratio (both positively) and glutamine (inversely) differed by fifths of animal protein intake with linear relationships. Finally, linear associations were observed for lysine, the sum of essential amino acids and the sum of ketogenic amino acids (all positive), alanine and the citrulline to ornithine ratio (both inverse), at a conventional significance level.

8.4.2.3 Dairy protein intake

Correlations: Dairy protein intake (% of energy) was positively correlated with concentrations of leucine, valine, the sum of branched-chain amino acids, phenylalanine, tryptophan, the sum of essential amino acids, tyrosine and the sum of ketogenic amino acids ($0.08 \leq r \leq 0.11$; $1 \times 10^{-5} \leq p \leq 0.001$; Table 8.11 on page 300), while inversely correlated with t4-OH-Pro ($r = -0.11$; $p = 5 \times 10^{-6}$), after accounting for the number of tests performed. Furthermore, intake of dairy protein was positively correlated with isoleucine, lysine, proline and at a conventionally significance level.

Amino acid concentrations and ratios by fifths of dairy protein intake: Concentrations of leucine, valine, the sum of branched-chain amino acids, phenylalanine, tryptophan, the sum of essential amino acids, tyrosine, t4-OH-Pro and the sum of ketogenic amino acids differed by fifths of dairy protein intake (% of energy), after correction for multiple testing ($2 \times 10^{-5} \leq p_{diff} \leq 0.001$; Table 8.16 on page 308). All of these (except t4-OH-Pro) and isoleucine, lysine and proline were positively, linearly associated with dairy protein, after controlling the FDR (Table 8.17 on page 310; Figures 8.1 on page 314, 8.2 on page 315 and 8.3 on page 316, C panels). The difference in amino acid concentrations per an 80 percentile increase in dairy protein intake ranged from 3.9% (95%CI: $0.9 - 7.0$%; $p_{trend} = 0.01$ for lysine) to 7.4% (95%CI:

$4.1 - 10.8\%$; $p_{trend} = 8 \times 10^{-6}$ for tyrosine). In contrast, an inverse linear association was observed with t4-OH-Pro concentration, which was 11.8% (95%CI: $6.1 - 17.0\%$; $p_{trend} = 0.0001$) lower in men with the highest versus the lowest consumption of dairy protein (Figure 8.4 on page 317, panel C).

Furthermore, glutamate and proline concentrations varied across fifths of dairy protein intake, at a conventional significance level.

8.4.2.4 Plant protein intake

Compared to intakes of total, animal and dairy protein, the associations with amino acids were weaker for plant protein intake, and the associations were often with other amino acids or in the opposite direction than those observed for the other protein types.

Correlations: Intake of plant protein (% of energy) was positively correlated with histidine concentration, after taking multiple testing into account ($r = 0.08$; $p = 0.001$; Table 8.11 on page 300). Positive correlations with asparagine, glutamine, glycine, the sum of non-essential amino acids, ornithine, the sum of glucogenic and the sum of ureagenic amino acids were also observed, at a conventional significance level.

Amino acid concentrations and ratios by fifths of plant protein intake: Histidine concentration was 4.1% (95%CI: $1.9 - 6.4\%$; $p_{trend} = 0.0003$) higher in men with the highest intake of plant protein (% of energy) compared to men with the lowest intake, and this linear association was significant after correcting for multiple testing (Table 8.18 on page 311; Figure 8.2 on page 315, panel D). The difference across fifths of plant protein intake however, did not remain statistically singificant after controlling the FDR (Table 8.19 on page 312).

Moreover, the sum of non-essential amino acids, ornithine and the sum of non-standard amino acids varied by fifths of plant protein intake with positive linear trends, at a conventional significance level. There was also evidence of positive linear associations with asparagine, citrulline, the sum of glucogenic amino acids and the sum of ureagenic amino acids, at a conventional significance level.

8.5 Discussion

8.5.1 Main findings

In this first large study of intake of protein from various sources and circulating concentrations of amino acids, the type of protein consumed by men was associated with their plasma amino acid profile. Men who consumed larger proportions of their daily energy intake from protein

had higher t4-OH-Pro and lower glycine plasma concentrations, than men who consumed lower proportions. Similar relationships were observed between these two amino acids and animal protein intake, which was further positively associated with the branched-chain amino acids leucine and valine, and the sum of branched-chain amino acids, while it was inversely related to citrulline, its ratio to arginine, and the sum of non-essential amino acids. A higher intake of dairy protein was associated with higher concentrations of several amino acids, including all branched-chain amino acids, several other essential amino acids (i.e. lysine, phenylalanine and tryptophan) and their sum, the non-essential amino acids proline and tyrosine, and the sum of ketogenic amino acids (which comprise several of the essential amino acids and tyrosine). Furthermore, intake of dairy protein was inversely associated with t4-OH-Pro concentration. In contrast to the other protein sub-types, intake of plant protein was in general not strongly associated with amino acid concentration besides a positive association with histidine concentration.

8.5.2 Previous studies and plausibility of findings

The associations between intake of protein types and plasma concentrations of amino acids should be interpreted in the light of findings from similar studies and how amino acid composition in dietary protein relates to amino acid composition in plasma. Plasma concentrations of amino acids are increased by intake, by *de novo* synthesis and by proteolysis from tissue, and are decreased by metabolism, incorporation into proteins and losses [129]. The influence of each of these factors varies between amino acids [129], making the interpretation of our results complex. For example, essential amino acids cannot be synthesised endogenously [128] and transformation into non-standard amino acids (e.g. proline to t4-OH-Pro) occurs after incorporation into proteins [232]. Furthermore, plasma concentrations of amino acids are under homeostatic control. Primarily the liver and hormones (including insulin and cortisol), regulate plasma concentrations [129]. The liver up- or downregulates ureagenesis to dispose of or retain nitrogen, respectively. Insulin and cortisol decrease and increase plasma concentrations, respectively, for example by affecting cellular uptake and protein breakdown.

8.5.2.1 Branched-chain amino acids

In the current study, positive associations between intake of animal protein especially from dairy products and the essential branched-chain amino acids leucine, isoleucine and valine were observed. Similar findings were previously reported for these amino acids and total protein intake [77]. Furthermore, lower concentrations of isoleucine or leucine, and valine have been previously reported in vegans than meat-eaters [77, 152, 134].

Branched-chain amino acids are abundant in dairy and other animal protein, comprising 18%, 20% and 21% protein content of beef, milk and egg, respectively, while somewhat sparser

in plant protein, e.g. 14%, 16% and 17% in wheat, nuts and soya, respectively [233]. This might partly contribute to the positive associations of animal and dairy protein intakes with branched-chain amino acid concentrations.

8.5.2.2 Other standard amino acids

Lysine, phenylalanine, tryptophan, proline, tyrosine and ketogenic amino acids: Intake of dairy protein was also positively associated with concentrations of the essential amino acids lysine, phenylalanine and tryptophan, the non-essential amino acids proline and tyrosine, and with the sum of ketogenic amino acids. In line with this, phenylalanine, tryptophan and tyrosine have been reported to be lower in vegans than in individuals consuming animal products, in one or more studies [77, 134, 152] and Table 4.5 on page 50.

As for the branched-chain amino acids, other essential amino acids are also more abundant in animal products than plant foods [128], and this might contribute to the current findings. The positive association of dairy protein intake with the sum of ketogenic amino acids (isoleucine, leucine, lysine, phenylalanine, threonine, tryptophan and tyrosine) was not surprising as each individual amino acid, except threonine ($p = 0.05$), was positively associated with intake of dairy protein.

Glycine: Inverse associations of total and animal protein intakes with glycine concentration were observed. This is in line with our previously reported results from the EPIC-Oxford (Table 4.7 on page 52), in which we observed a positive correlation between glycine and intake of plant protein, especially from soya, which is rich in glycine [120]. The lack of association between plant protein and glycine in the current analysis might partly reflect a narrower range of plant protein intake in the control participants from EPIC (2.4 to 8.9 percent of energy) than in the four diet groups in EPIC-Oxford (3.9 to 16.6 percent of energy).

In line with these results is our previous observation of higher glycine concentration in vegans than the other diet groups (Tables 4.5 on page 50 and 5.7 on page 89). So is the higher glycine concentration in vegetarians than in omnivores observed in a previous study [151]. However, no difference in glycine concentration between vegans and omnivores has been reported in a small study ($n = 36$) [134], possibly reflecting low statistical power.

Histidine: A possible reason for the positive association between plant protein intake and histidine concentration is not clear. However, the proportion of histidine in some plant proteins, including soya and nuts, is comparable to that of some animal proteins, e.g. fish and milk, but lower that of others, e.g. beef and cheese [233]. Previously, in the EPIC-Oxford diet group analysis, we did not find an association between this amino acid and intake of plant protein or

other protein types (Table 4.7 on page 52). Histidine was not investigated by Anderson and Blendis [77]. With respect to diet groups, lower histidine has been reported in vegans than in omnivores by others [134], while we observed no association (Tables 4.5 on page 50).

8.5.2.3 Non-standard amino acids

t4-OH-Pro: The strongest associations in this study were seen for t4-OH-Pro (one type of hydroxyproline) with total and especially animal protein intake, while an inverse association was seen with dairy protein intake. Hydroxyproline is found in connective tissue, such as collagen, and has been suggested as a marker of red and processed meat intakes [186, 187]. Hydroxyproline was not included in the other study reporting on habitual protein intake and amino acid concentrations [77], nor was it available for our analysis of amino acids by diet group (Chapter 4). However, in line with our results, omnivores have been reported to have higher plasma concentrations of hydroxyproline than vegans [152].

Citrulline: Intake of animal products was inversely associated with citrulline and its ratio to arginine. The latter association might mainly be explained by citrulline because arginine was less closely associated with animal protein intake. Neither citrulline nor arginine was correlated with intake of protein of any type in our previous analysis in the four diet groups (Table 4.7 on page 52) and neither was investigated by Anderson and Blendis [77].

8.5.3 Strengths and limitations

In contrast to previous analyses on the associations between habitual protein intake and circulating amino acid concentrations (Chapter 4, [17, 77]), this study utilises a large dataset including men (regardless of their diet group) from six European countries and goes beyond correlation coefficients to describe the associations. The large sample size and extensive data on dietary intake, covariates and metabolites are the main strengths of the current study. The large sample size and the wide range of dietary exposure, owing to differences in local dietary habits across EPIC, have improved the likelihood of finding associations between protein intake and amino acid concentrations, should they truly exist. Furthermore, the chosen metabolomics assay has provided data on 19 of 20 standard amino acids (not cysteine) and three non-standard amino acids. Finally, we could assess the effects of a wide range of potential confounding factors on the associations.

The main limitations of the current study are related to measurements of protein intake and amino acid concentrations. Such measurement errors are likely to be random and therefore, the associations may be attenuated to some degree.

Firstly, measurement error in estimations of dietary intake is inevitable. The recorded diet might not reflect the usual diet due to intentional or unintentional erroneous recording (e.g. recall bias or difficulties answering the questionnaire) [234] or if the measurement tool does not capture the diet consumed (e.g. if important foods are not included in the questionnaire). In order to capture local diets, validated country-specific food questionnaires were used in EPIC [99, 101, 102, 103, 104, 105, 106, 107]. Correlations between estimated dietary nitrogen intake and urinary nitrogen excretion ranged from 0.24 in the UK [107] to 0.58 in Spain [102], suggesting modest to good validity of protein intake in the cohort. The lack of associations between plant protein intake and amino acid concentrations might indicate that plant protein was measured with greater degree of measurement error than the other protein types. However, accounting for energy intake, by using protein density (percentage of energy) instead of absolute intake (g/d) and by further adjusting the statistical analyses for energy intake, might reduce the degree of measurement error as food frequency questionnaires are better at assessing composition of diet than absolute intakes [142].

Secondly, pre-analytical factors relating to blood sampling and processing, and analytical factors might have led to random measurement error in the metabolite data and thereby to attenuation the study results. Studies have found circulating concentrations of most metabolites not to be strongly affected by fasting status [147, 161, 171, 172], time of day at blood collection [144, 172], time from blood collection to processing and freeze-thaw cycles [144, 145]. However, some amino acids are among the metabolites that are influenced by fasting status (leucine/isoleucine, tryptophan and tyrosine) [143] and processing delay (plasma concentrations of two thirds of amino acids changed significantly after 24 hours at room temperature compared to concentrations measured in samples that were frozen immediately) [144]. However, EPIC-Oxford was the only centre to transport blood samples at ambient temperature (6.8% of samples in the current analysis). To account for pre-analytical and analytic factors in this analysis, all analyses were adjusted for centre and assay batch. Moreover, further adjustments for fasting status and time of day at blood collection did not materially change our results.

8.5.4 Future work

In Chapter 9, the association between concentrations and ratios amino acids and IGF-I concentration will be investigated to elucidate whether amino acid concentrations may play a role in the association between protein intake and IGF-I concentrations.

Several of the amino acids associated with protein intake in the current study might also be associated with risk of prostate cancer (i.e. lysine, phenylalanine, tryptophan, proline, citrulline and t4-OH-Pro; see references [19, 29, 82, 84] and Chapter 6). While it is not possible from this thesis to conclude if the observed associations of protein intake with plasma amino acids,

and of amino acids with prostate cancer risk, translate into associations between intake of protein and prostate cancer risk, men with higher intakes of dairy protein have been reported to have higher risk of overall and high grade prostate cancer [8]. Further large-scale research into the association between protein intake and prostate cancer risk might contribute to our understanding of modifiable risk factors for prostate cancer.

8.6 Conclusion

This first large study to investigate the associations between intake of protein and circulating concentrations of amino acids showed that the type of protein men consume is reflected in their plasma amino acid profile. Men who consumed a larger proportion of their daily energy intake from protein had higher t4-OH-Pro and lower glycine concentrations. Similar relationships were observed between these two amino acids and intake of animal protein, which was further positively related to concentrations of branched-chain amino acids and inversely to citrulline, its ratio to arginine and the sum of non-essential amino acids. A higher intake of dairy protein was associated with higher concentrations of several amino acids, including all branched-chain amino acids, several other essential amino acids, proline and tyrosine, and with lower t4-OH-Pro concentration. In contrast, in the current study, intake of plant protein was not strongly associated with plasma amino acids.

8.7 TABLES

Table 8.1: Characteristics of the study population by thirds of total protein intake (% of energy)[1]

Characteristics	Thirds of total protein intake (% of energy)[2,3]		
	Bottom n=565	Middle n=566	Top n=566
Age at blood collection, years	59.9 (59.4 - 60.5)	59.7 (59.2 - 60.2)	60.1 (59.6 - 60.6)
Height[4], cm	172.3 (171.8 - 172.9)	172.4 (171.8 - 172.9)	172.4 (171.8 - 173.0)
Weight[4], cm	78.3 (77.3 - 79.3)	79.5 (78.5 - 80.5)	82.3 (81.3 - 83.3)
BMI[4], kg/m²	26.3 (26.0 - 26.6)	26.8 (26.5 - 27.1)	27.6 (27.3 - 27.9)
Smoking, %			
Never	29.8 (26.1 - 33.6)	30.4 (26.6 - 34.2)	29.2 (25.5 - 32.8)
Former	49.2 (45.1 - 53.3)	42.6 (38.6 - 46.6)	46.9 (42.9 - 50.9)
Current	21.0 (17.7 - 24.3)	27.0 (23.4 - 30.6)	24.0 (20.6 - 27.4)
Alcohol consumption[4], %			
<10 g/day	39.5 (35.7 - 43.2)	41.3 (37.6 - 45.1)	50.7 (46.8 - 54.6)
10-19 g/day	17.0 (13.9 - 20.1)	17.1 (14.1 - 20.1)	23.3 (19.9 - 26.8)
20-40 g/day	21.8 (18.5 - 25.1)	25.1 (21.6 - 28.6)	16.8 (13.8 - 19.8)
≥40 g/day	21.8 (18.6 - 25.0)	16.4 (13.6 - 19.2)	9.2 (6.9 - 11.5)
Physical activity[4], %			
Inactive	27.1 (23.7 - 30.6)	23.9 (20.5 - 27.3)	30.3 (26.7 - 34.0)
Moderately inactive	30.6 (26.9 - 34.3)	31.6 (27.8 - 35.5)	34.1 (30.4 - 37.9)
Moderately active	19.5 (16.2 - 22.7)	24.2 (20.8 - 27.7)	23.2 (19.7 - 26.6)
Active	22.8 (19.3 - 26.2)	20.3 (17.0 - 23.5)	12.4 (9.7 - 15.1)
Marital status[4], %			
Not married or cohabiting	10.3 (7.7 - 12.9)	9.9 (7.1 - 12.6)	9.3 (6.2 - 12.3)
Married or cohabiting	89.7 (87.1 - 92.3)	90.1 (87.4 - 92.9)	90.7 (87.7 - 93.8)
Educational attainment[4], %			
Primary or equivalent	36.0 (32.3 - 39.8)	41.2 (37.4 - 44.9)	48.0 (44.2 - 51.9)
Secondary	37.5 (33.6 - 41.5)	38.8 (34.9 - 42.7)	34.0 (30.2 - 37.7)
Degree	26.4 (23.0 - 29.8)	20.0 (16.9 - 23.2)	18.0 (15.0 - 21.0)
Energy intake, MJ	10.59 (10.37 - 10.81)	9.90 (9.69 - 10.11)	8.88 (8.65 - 9.10)
Animal protein[4], % of energy	7.1 (6.9 - 7.3)	9.6 (9.5 - 9.8)	13.0 (12.8 - 13.1)
Dairy protein[4], % of energy	2.5 (2.3 - 2.6)	3.1 (3.0 - 3.3)	3.8 (3.6 - 3.9)
Plant protein[4], % of energy	5.0 (4.9 - 5.1)	5.1 (5.0 - 5.3)	5.3 (5.2 - 5.4)
Fasting status[4,5], %			
No	60.5	48.7	44.7
Yes	39.5	51.3	55.3
Time of blood collection[4,5], %			
06:00 - 08:59	10.8	16.1	13.1
09:00 - 11:59	44.6	48.9	53.8
12:00 - 14:59	23.2	15.4	13.8
15:00 - 20:00	21.4	19.6	19.3

[1] Adjusted for study centre

[2] Total protein intake: bottom third: 6.64 — 15.28% of energy; middle third: 15.28 — 17.78% of energy; top third: 17.78 — 29.32% of energy

[3] Numbers are mean (95%CI) for continuous variables and % (95%CI) for categorical variables

[4] Unknown values for some participants; the calculations of percentages exclude missing values

[5] No confidence intervals could be calculated due to zero observations in some cells

Table 8.2: Partial correlation coefficients between protein intake (% of energy) and amino acid concentrations and ratios in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Total protein			Animal protein			Dairy protein			Plant protein		
	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2
Branched-chain essential amino acids												
Isoleucine	0.07	**0.0081**	0.0172	0.07	**0.0059**	0.0128	0.05	0.0266	0.0906	-0.01	0.7713	0.9502
Leucine	0.09	**0.0003**	0.0017	0.09	**0.0003**	0.0013	0.08	**0.0010**	0.0038	0.00	0.9127	0.9995
Valine	0.11	**2.E-05**	0.0002	0.10	**2.E-05**	0.0002	0.10	**3.E-05**	0.0002	-0.02	0.4778	0.7756
Branched-chain amino acids	0.10	**0.0001**	0.0008	0.10	**0.0001**	0.0005	0.09	**0.0003**	0.0012	-0.01	0.6442	0.8761
Other essential amino acids												
Histidine	-0.01	0.6598	0.7457	-0.04	0.1012	0.1434	-0.02	0.3696	0.5252	0.08	**0.0018**	0.0411
Lysine	0.08	**0.0022**	0.0092	0.07	**0.0048**	0.0124	0.05	0.0304	0.0938	0.01	0.8104	0.9502
Methionine	-0.01	0.6267	0.7347	-0.02	0.5209	0.5366	0.04	0.0997	0.2118	0.03	0.2526	0.4771
Phenylalanine	0.05	0.0400	0.0755	0.03	0.1946	0.2363	0.11	**1.E-05**	0.0002	0.01	0.6077	0.8609
Threonine	-0.02	0.3450	0.4692	-0.02	0.4205	0.4468	0.04	0.0793	0.1930	0.00	0.9886	0.9995
Tryptophan	0.04	0.1342	0.2401	0.04	0.0917	0.1418	0.10	**0.0001**	0.0005	-0.02	0.4132	0.7394
Essential amino acids	0.08	**0.0013**	0.0062	0.07	**0.0024**	0.0069	0.09	**0.0003**	0.0013	0.00	0.9995	0.9995
Non-essential amino acids												
Alanine	-0.04	0.1525	0.2592	-0.05	0.0354	0.0668	-0.01	0.5982	0.7533	0.04	0.1169	0.2932
Arginine	0.01	0.6238	0.7347	0.01	0.6332	0.6332	0.00	0.9304	0.9304	-0.01	0.7952	0.9502
Asparagine	-0.07	**0.0076**	0.0172	-0.09	**0.0001**	0.0006	0.02	0.4017	0.5252	0.07	**0.0067**	0.0453
Aspartate	-0.03	0.3102	0.4692	-0.03	0.2966	0.3361	-0.04	0.1287	0.2446	-0.01	0.5698	0.8424
Glutamate	0.02	0.3393	0.4692	0.04	0.0841	0.1361	-0.01	0.7724	0.8207	-0.03	0.1925	0.3850
Glutamine	-0.07	**0.0039**	0.0128	-0.10	**0.0001**	0.0004	-0.02	0.3958	0.5252	0.07	**0.0039**	0.0438
Glycine	-0.12	**9.E-07**	2.E-05	-0.15	**4.E-09**	7.E-08	-0.04	0.0865	0.1962	0.07	**0.0065**	0.0453
Proline	-0.02	0.4790	0.6031	-0.04	0.0977	0.1434	0.05	0.0384	0.1088	0.04	0.1207	0.2932
Serine	-0.01	0.7557	0.7786	-0.04	0.1136	0.1544	0.01	0.7628	0.8207	0.05	0.0378	0.1286
Tyrosine	0.05	0.0348	0.0696	0.05	0.0401	0.0718	0.10	**2.E-05**	0.0002	0.01	0.7197	0.9412
Non-essential amino acids	-0.07	**0.0047**	0.0134	-0.10	**5.E-05**	0.0003	0.00	0.8822	0.9089	0.08	**0.0024**	0.0411

Continues...

Table 8.2 continued

Amino acids	Total protein			Animal protein			Dairy protein			Plant protein		
	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2
Non-standard amino acids												
Citrulline	-0.09	**0.0003**	0.0017	-0.12	**2.E-06**	3.E-05	-0.04	0.1295	0.2446	0.04	0.0829	0.2539
Ornithine	0.00	0.9536	0.9536	-0.03	0.2370	0.2779	-0.01	0.6654	0.8080	0.06	0.0128	0.0635
t4-OH-Pro	0.13	**2.E-07**	6.E-06	0.16	**1.E-10**	4.E-09	-0.11	**4.E-06**	0.0001	-0.04	0.0896	0.2539
Non-standard amino acids	-0.01	0.8033	0.8276	-0.04	0.1555	0.1958	-0.04	0.1012	0.2150	0.06	0.0086	0.0490
Breakdown pathways												
Glucogenic amino acids[3]	-0.03	0.1867	0.3023	-0.06	**0.0205**	0.0410	0.03	0.2637	0.4269	0.06	0.0223	0.0841
Ketogenic amino acids[4]	0.07	**0.0036**	0.0128	0.07	**0.0051**	0.0124	0.09	**0.0002**	0.0012	0.00	0.9827	0.9995
Ureagenic amino acids[5]	-0.02	0.3235	0.4692	-0.05	0.0428	0.0727	-0.03	0.2160	0.3673	0.06	0.0175	0.0745
Amino acid ratios												
Citrulline:arginine	-0.07	**0.0074**	0.0172	-0.08	**0.0008**	0.0028	-0.03	0.3099	0.4790	0.03	0.1922	0.3850
Citrulline:ornithine	-0.07	**0.0053**	0.0138	-0.07	**0.0060**	0.0128	-0.02	0.3996	0.5252	-0.02	0.5291	0.8177
Fischer ratio[6]	0.07	**0.0042**	0.0128	0.08	**0.0019**	0.0059	-0.01	0.7100	0.8207	-0.02	0.4790	0.7756
Ornithine:arginine	-0.01	0.6899	0.7457	-0.02	0.3245	0.3559	-0.01	0.7681	0.8207	0.04	0.1299	0.2944
Tyrosine:phenylal.	0.02	0.4228	0.5529	0.04	0.1520	0.1914	0.03	0.1626	0.2911	0.00	0.9867	0.9995

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.3: Geometric mean amino acid concentrations and ratios by fifths of total protein intake in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of total protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	65.7	(63.9-67.6)	68.3	(66.5-70.2)	66.8	(65.0-68.6)	69.5	(67.6-71.4)	69.1	(67.1-71.2)	0.0421	0.1023
Leucine	121.2	(118.2-124.3)	125.7	(122.7-128.8)	122.7	(119.8-125.6)	129.3	(126.2-132.4)	128.8	(125.5-132.2)	**0.0010**	0.0114
Valine	204.8	(200.4-209.3)	211.5	(207.1-216.0)	208.9	(204.7-213.2)	215.0	(210.6-219.4)	217.8	(213.0-222.8)	**0.0024**	0.0172
Branched-chain aa	393.1	(384.3-402.2)	406.8	(398.1-415.8)	399.7	(391.2-408.4)	414.9	(406.0-423.9)	416.9	(407.1-426.8)	**0.0025**	0.0172
Other essential amino acids												
Histidine	69.4	(68.2-70.5)	68.4	(67.4-69.5)	68.3	(67.2-69.3)	69.1	(68.1-70.2)	68.5	(67.3-69.6)	0.5489	0.6912
Lysine	177.9	(173.7-182.1)	180.1	(176.2-184.2)	177.8	(173.9-181.8)	184.9	(180.8-189.0)	185.0	(180.5-189.5)	0.0369	0.0964
Methionine	19.2	(18.7-19.8)	19.1	(18.6-19.6)	19.1	(18.6-19.6)	19.5	(19.0-20.0)	19.0	(18.5-19.6)	0.6599	0.8013
Phenylalanine	55.3	(54.3-56.3)	55.7	(54.8-56.7)	56.2	(55.2-57.1)	56.9	(55.9-57.9)	56.8	(55.8-57.9)	0.2125	0.3441
Threonine	103.6	(101.1-106.1)	102.4	(100.1-104.8)	102.2	(99.9-104.5)	103.3	(101.0-105.7)	101.5	(99.0-104.0)	0.7457	0.8743
Tryptophan	51.4	(50.4-52.5)	51.0	(50.0-52.0)	51.5	(50.5-52.5)	52.4	(51.4-53.4)	52.2	(51.1-53.3)	0.3526	0.4796
Essential aa	876.3	(860.9-892.1)	890.5	(875.4-905.8)	882.0	(867.3-896.9)	908.1	(892.9-923.6)	907.4	(890.9-924.3)	0.0196	0.0741
Non-essential amino acids												
Alanine	340.7	(332.1-349.4)	335.8	(327.7-344.0)	324.6	(317.0-332.5)	333.8	(325.9-341.9)	330.2	(321.7-339.0)	0.0905	0.1923
Arginine	47.2	(45.3-49.3)	47.9	(46.0-49.8)	46.1	(44.3-48.0)	49.1	(47.2-51.1)	48.2	(46.1-50.3)	0.2752	0.3899
Asparagine	41.0	(40.2-41.9)	39.9	(39.1-40.7)	39.9	(39.1-40.8)	39.8	(39.0-40.6)	39.2	(38.4-40.1)	0.0992	0.1984
Aspartate	6.2	(5.9-6.5)	6.1	(5.8-6.3)	6.2	(6.0-6.5)	6.1	(5.8-6.4)	6.0	(5.7-6.3)	0.8417	0.9232
Glutamate	52.4	(50.2-54.7)	53.1	(51.0-55.4)	54.1	(51.9-56.3)	54.2	(52.1-56.5)	53.7	(51.4-56.2)	0.8082	0.9159
Glutamine	615.6	(605.3-626.0)	602.3	(592.7-612.1)	593.6	(584.2-603.0)	594.3	(584.9-603.8)	596.0	(585.7-606.5)	0.0182	0.0741
Glycine	183.2	(178.9-187.6)	175.0	(171.1-179.0)	171.9	(168.1-175.8)	168.8	(165.1-172.6)	169.2	(165.1-173.3)	**1.E-05**	0.0003
Proline	186.0	(180.3-192.0)	185.0	(179.5-190.7)	177.8	(172.6-183.1)	185.5	(180.1-191.2)	184.4	(178.5-190.5)	0.1914	0.3253

Continues...

Table 8.3 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of total protein intake (% of energy)										p_{diff}[2]	p_{adj}[2]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	81.1	(79.2-83.0)	81.0	(79.2-82.8)	81.3	(79.6-83.1)	81.5	(79.7-83.4)	81.6	(79.7-83.6)	0.9887	0.9979
Tyrosine	60.5	(58.9-62.0)	60.9	(59.4-62.4)	61.9	(60.5-63.4)	63.8	(62.3-65.3)	62.3	(60.7-64.0)	0.0357	0.0964
Non-essential aa	1649	(1,626-1,673)	1621	(1,600-1,643)	1591	(1,570-1,612)	1612	(1,591-1,634)	1603	(1,580-1,627)	**0.0067**	0.0325
Non-standard amino acids												
Citrulline	30.1	(29.2-31.0)	29.7	(28.9-30.6)	29.0	(28.3-29.9)	29.0	(28.2-29.9)	28.0	(27.2-28.9)	0.0273	0.0843
Ornithine	79.3	(77.0-81.8)	79.2	(77.0-81.5)	79.7	(77.5-82.0)	78.9	(76.7-81.2)	79.4	(77.0-81.9)	0.9932	0.9979
t4-OH-Pro	7.6	(7.3-8.0)	8.2	(7.8-8.6)	8.0	(7.7-8.4)	8.4	(8.0-8.8)	9.2	(8.8-9.7)	**1.E-05**	0.0003
Non-standard aa	119.6	(116.8-122.5)	119.6	(116.9-122.4)	119.2	(116.5-121.9)	119.1	(116.4-121.8)	119.5	(116.6-122.5)	0.9981	0.9981
Breakdown pathways												
Glucogenic aa[3]	2228	(2,198-2,258)	2208	(2,179-2,237)	2173	(2,145-2,201)	2206	(2,178-2,235)	2197	(2,167-2,228)	0.1276	0.2283
Ketogenic aa[4]	641.6	(629.8-653.7)	650.4	(638.9-662.1)	645.8	(634.6-657.2)	666.3	(654.7-678.2)	662.6	(650.0-675.4)	0.0256	0.0843
Ureagenic aa[5]	165.5	(162.3-168.8)	166.2	(163.1-169.3)	162.3	(159.3-165.3)	166.5	(163.4-169.6)	163.5	(160.2-166.8)	0.2533	0.3899
Amino acid ratios												
Citrulline:arginine	0.6	(0.6-0.7)	0.6	(0.6-0.7)	0.6	(0.6-0.7)	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.0476	0.1079
Citrulline:ornithine	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.3-0.4)	0.1220	0.2283
Fischer ratio[6]	2.3	(2.3-2.4)	2.4	(2.4-2.5)	2.3	(2.3-2.4)	2.4	(2.3-2.4)	2.4	(2.4-2.5)	**0.0046**	0.0259
Ornithine:arginine	1.7	(1.6-1.8)	1.7	(1.6-1.7)	1.7	(1.6-1.8)	1.6	(1.5-1.7)	1.6	(1.6-1.7)	0.4392	0.5744
Tyrosine:phenylala.	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	0.2717	0.3899

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.4: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in total protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in total protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	4.83	(0.77 - 9.04)	**0.0191**	0.0392
Leucine	6.21	(2.57 - 9.98)	**0.0007**	0.0049
Valine	5.72	(2.59 - 8.95)	**0.0003**	0.0033
Branched-chain amino acids	5.64	(2.36 - 9.03)	**0.0007**	0.0049
Other essential amino acids				
Histidine	-0.61	(-2.82 - 1.65)	0.5911	0.6930
Lysine	4.28	(0.97 - 7.71)	**0.0110**	0.0313
Methionine	0.13	(-3.45 - 3.84)	0.9445	0.9666
Phenylalanine	3.06	(0.48 - 5.70)	**0.0196**	0.0392
Threonine	-1.23	(-4.44 - 2.09)	0.4632	0.6299
Tryptophan	2.24	(-0.56 - 5.12)	0.1178	0.2093
Essential amino acids	3.67	(1.15 - 6.25)	**0.0041**	0.0200
Non-essential amino acids				
Alanine	-2.72	(-6.07 - 0.75)	0.1231	0.2093
Arginine	2.70	(-3.12 - 8.87)	0.3702	0.5244
Asparagine	-3.59	(-6.40 - -0.69)	**0.0157**	0.0356
Aspartate	-2.01	(-8.28 - 4.69)	0.5466	0.6637
Glutamate	2.95	(-3.01 - 9.27)	0.3389	0.5009
Glutamine	-3.16	(-5.38 - -0.88)	**0.0069**	0.0233
Glycine	-7.66	(-10.63 - -4.59)	**2.E-06**	0.0001
Proline	-0.63	(-4.86 - 3.78)	0.7749	0.8499
Serine	0.84	(-2.36 - 4.14)	0.6127	0.6944
Tyrosine	4.56	(0.98 - 8.28)	**0.0122**	0.0320
Non-essential amino acids	-2.51	(-4.39 - -0.59)	**0.0108**	0.0313
Non-standard amino acids				
Citrulline	-6.26	(-9.97 - -2.41)	**0.0017**	0.0095
Ornithine	-0.09	(-4.12 - 4.12)	0.9666	0.9666
t4-OH-Pro	17.35	(9.69 - 25.54)	**4.E-06**	0.0001
Non-standard amino acids	-0.25	(-3.50 - 3.10)	0.8804	0.9354
Breakdown pathways				
Glucogenic amino acids[4]	-1.13	(-2.97 - 0.74)	0.2335	0.3781
Ketogenic amino acids[5]	3.66	(1.04 - 6.36)	**0.0060**	0.0230
Ureagenic amino acids[6]	-0.86	(-3.51 - 1.86)	0.5298	0.6637
Amino acid ratios				
Citrulline:arginine	-8.73	(-14.49 - -2.58)	**0.0061**	0.0230
Citrulline:ornithine	-6.18	(-10.89 - -1.22)	**0.0152**	0.0356
Fischer ratio[7]	2.10	(-0.17 - 4.43)	0.0701	0.1324
Ornithine:arginine	-2.72	(-9.99 - 5.15)	0.4871	0.6370
Tyrosine:phenylalanine	1.46	(-1.26 - 4.26)	0.2943	0.4549

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.5: Geometric mean amino acid concentrations and ratios by fifths of animal protein intake in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of animal protein intake (% of energy)										p_{diff}[2]	p_{adj}[2]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	64.9	(63.1 - 66.8)	67.6	(65.7 - 69.4)	69.4	(67.6 - 71.3)	68.1	(66.3 - 70.0)	69.5	(67.5 - 71.5)	**0.0091**	0.0310
Leucine	120.2	(117.3 - 123.3)	124.6	(121.6 - 127.7)	127.0	(124.0 - 130.1)	126.3	(123.3 - 129.4)	129.5	(126.2 - 132.8)	**0.0030**	0.0146
Valine	203.7	(199.4 - 208.2)	210.4	(206.0 - 214.8)	213.8	(209.4 - 218.2)	211.2	(206.8 - 215.6)	219.0	(214.2 - 223.9)	**0.0006**	0.0047
Branched-chain aa	390.3	(381.5 - 399.3)	403.7	(395.0 - 412.7)	411.7	(402.9 - 420.6)	406.6	(397.9 - 415.6)	419.1	(409.5 - 429.0)	**0.0010**	0.0070
Other essential amino acids												
Histidine	69.2	(68.1 - 70.3)	69.0	(68.0 - 70.1)	68.6	(67.5 - 69.6)	68.4	(67.3 - 69.5)	68.5	(67.3 - 69.6)	0.8429	0.8956
Lysine	177.2	(173.1 - 181.4)	178.6	(174.6 - 182.6)	182.2	(178.3 - 186.3)	181.4	(177.4 - 185.5)	186.3	(181.9 - 190.7)	0.0611	0.1485
Methionine	19.2	(18.7 - 19.7)	19.3	(18.8 - 19.8)	19.1	(18.7 - 19.6)	19.3	(18.8 - 19.8)	19.1	(18.6 - 19.6)	0.9778	0.9778
Phenylalanine	55.2	(54.2 - 56.3)	55.9	(55.0 - 56.9)	56.9	(55.9 - 57.9)	56.3	(55.3 - 57.3)	56.5	(55.5 - 57.6)	0.2395	0.4071
Threonine	103.1	(100.7 - 105.6)	101.6	(99.3 - 104.0)	103.7	(101.4 - 106.0)	102.7	(100.4 - 105.1)	101.9	(99.4 - 104.4)	0.7222	0.7921
Tryptophan	50.9	(49.9 - 52.0)	51.4	(50.4 - 52.4)	51.5	(50.5 - 52.5)	52.2	(51.2 - 53.2)	52.5	(51.4 - 53.6)	0.3394	0.5076
Essential aa	871.6	(856.2 - 887.2)	886.3	(871.3 - 901.6)	901.2	(886.2 - 916.5)	893.7	(878.6 - 909.1)	911.5	(895.2 - 928.2)	**0.0155**	0.0438
Non-essential amino acids												
Alanine	339.7	(331.2 - 348.4)	335.2	(327.2 - 343.5)	332.3	(324.4 - 340.3)	326.9	(319.1 - 335.0)	330.9	(322.5 - 339.6)	0.3434	0.5076
Arginine	46.8	(44.9 - 48.9)	47.9	(46.0 - 49.9)	46.8	(45.0 - 48.7)	48.3	(46.4 - 50.3)	48.6	(46.5 - 50.7)	0.6569	0.7672
Asparagine	41.0	(40.2 - 41.9)	40.1	(39.3 - 40.9)	39.8	(39.0 - 40.6)	39.7	(38.9 - 40.6)	39.3	(38.4 - 40.1)	0.1030	0.2002
Aspartate	6.2	(5.9 - 6.5)	6.2	(5.9 - 6.5)	6.0	(5.7 - 6.3)	6.3	(6.0 - 6.6)	6.1	(5.8 - 6.4)	0.6260	0.7672
Glutamate	51.0	(48.9 - 53.2)	53.7	(51.5 - 56.0)	55.0	(52.8 - 57.3)	53.7	(51.6 - 56.0)	54.2	(51.9 - 56.6)	0.1576	0.2820
Glutamine	621.0	(610.6 - 631.5)	600.2	(590.6 - 610.0)	593.6	(584.3 - 603.1)	591.5	(582.1 - 601.1)	595.6	(585.5 - 605.8)	**0.0005**	0.0047
Glycine	184.7	(180.4 - 189.1)	176.3	(172.4 - 180.4)	170.3	(166.6 - 174.2)	167.8	(164.0 - 171.6)	169.2	(165.2 - 173.3)	**1.E-07**	3.E-06
Proline	185.8	(180.0 - 191.7)	186.2	(180.6 - 191.9)	184.5	(179.1 - 190.1)	181.5	(176.1 - 187.0)	180.8	(175.1 - 186.7)	0.6650	0.7672

Continues...

Table 8.5 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of animal protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	81.6	(79.7 - 83.6)	82.0	(80.2 - 83.9)	80.7	(79.0 - 82.5)	81.2	(79.4 - 83.0)	81.0	(79.1 - 82.9)	0.8893	0.9162
Tyrosine	60.1	(58.6 - 61.6)	61.6	(60.1 - 63.1)	61.7	(60.3 - 63.2)	63.2	(61.7 - 64.7)	62.8	(61.2 - 64.4)	0.0873	0.1979
Non-essential aa	1652	(1,629 - 1,675)	1623	(1,601 - 1,645)	1607	(1,586 - 1,629)	1593	(1,572 - 1,615)	1602	(1,579 - 1,625)	**0.0069**	0.0262
Non-standard amino acids												
Citrulline	30.4	(29.6 - 31.3)	29.6	(28.8 - 30.4)	29.4	(28.6 - 30.3)	28.7	(27.9 - 29.5)	27.8	(27.0 - 28.6)	**0.0015**	0.0086
Ornithine	80.1	(77.8 - 82.6)	79.0	(76.8 - 81.3)	80.4	(78.1 - 82.7)	78.1	(75.9 - 80.4)	79.0	(76.6 - 81.4)	0.6606	0.7672
t4-OH-Pro	7.5	(7.1 - 7.8)	7.8	(7.5 - 8.2)	8.5	(8.1 - 8.9)	8.6	(8.2 - 9.0)	9.2	(8.7 - 9.7)	**3.E-07**	5.E-06
Non-standard aa	120.6	(117.8 - 123.5)	118.8	(116.1 - 121.6)	120.8	(118.1 - 123.6)	118.2	(115.5 - 120.9)	118.6	(115.7 - 121.5)	0.5725	0.7672
Breakdown pathways												
Glucogenic aa[3]	2227	(2,197 - 2,258)	2207	(2,178 - 2,236)	2200	(2,172 - 2,229)	2180	(2,152 - 2,208)	2198	(2,167 - 2,228)	0.3078	0.4984
Ketogenic aa[4]	637.8	(626.1 - 649.7)	647.4	(635.9 - 659.0)	659.3	(647.9 - 671.0)	656.4	(644.9 - 668.2)	665.8	(653.4 - 678.5)	0.0281	0.0735
Ureagenic aa[5]	166.0	(162.8 - 169.3)	164.7	(161.6 - 167.8)	166.2	(163.1 - 169.3)	163.7	(160.7 - 166.8)	163.3	(160.1 - 166.6)	0.6769	0.7672
Amino acid ratios												
Citrulline:arginine	0.6	(0.6 - 0.7)	0.6	(0.6 - 0.6)	0.6	(0.6 - 0.7)	0.6	(0.6 - 0.6)	0.6	(0.5 - 0.6)	**0.0064**	0.0262
Citrulline:ornithine	0.4	(0.4 - 0.4)	0.4	(0.4 - 0.4)	0.4	(0.4 - 0.4)	0.4	(0.4 - 0.4)	0.4	(0.3 - 0.4)	0.0996	0.2002
Fischer ratio[6]	2.3	(2.3 - 2.4)	2.4	(2.3 - 2.4)	2.4	(2.4 - 2.4)	2.4	(2.3 - 2.4)	2.4	(2.4 - 2.5)	**0.0107**	0.0330
Ornithine:arginine	1.7	(1.6 - 1.8)	1.6	(1.6 - 1.7)	1.7	(1.6 - 1.8)	1.6	(1.5 - 1.7)	1.6	(1.5 - 1.7)	0.4379	0.6203
Tyrosine:phenylala.	1.1	(1.1 - 1.1)	1.1	(1.1 - 1.1)	1.1	(1.1 - 1.1)	1.1	(1.1 - 1.1)	1.1	(1.1 - 1.1)	0.1060	0.2002

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.6: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in animal protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in animal protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	5.96	(1.84 - 10.24)	**0.0043**	0.0121
Leucine	6.71	(3.03 - 10.53)	**0.0003**	0.0016
Valine	6.12	(2.96 - 9.38)	**0.0001**	0.0010
Branched-chain amino acids	6.22	(2.90 - 9.64)	**0.0002**	0.0013
Other essential amino acids				
Histidine	-1.23	(-3.44 - 1.03)	0.2838	0.3554
Lysine	4.70	(1.35 - 8.17)	**0.0057**	0.0148
Methionine	-0.09	(-3.68 - 3.63)	0.9602	0.9602
Phenylalanine	2.17	(-0.40 - 4.81)	0.0983	0.1499
Threonine	-0.55	(-3.80 - 2.81)	0.7449	0.7914
Tryptophan	3.02	(0.18 - 5.94)	0.0367	0.0693
Essential amino acids	4.01	(1.47 - 6.62)	**0.0019**	0.0063
Non-essential amino acids				
Alanine	-3.08	(-6.44 - 0.40)	0.0820	0.1404
Arginine	3.20	(-2.68 - 9.45)	0.2927	0.3554
Asparagine	-3.79	(-6.62 - -0.88)	0.0110	0.0234
Aspartate	-0.93	(-7.32 - 5.89)	0.7821	0.8058
Glutamate	5.14	(-0.98 - 11.64)	0.1014	0.1499
Glutamine	-3.93	(-6.15 - -1.66)	**0.0008**	0.0033
Glycine	-8.69	(-11.64 - -5.64)	**7.E-08**	1.E-06
Proline	-3.11	(-7.25 - 1.23)	0.1574	0.2230
Serine	-1.04	(-4.20 - 2.23)	0.5286	0.5797
Tyrosine	4.67	(1.06 - 8.42)	**0.0109**	0.0234
Non-essential amino acids	-3.20	(-5.08 - -1.29)	**0.0011**	0.0043
Non-standard amino acids				
Citrulline	-8.08	(-11.73 - -4.29)	**5.E-05**	0.0005
Ornithine	-1.64	(-5.64 - 2.53)	0.4352	0.4933
t4-OH-Pro	22.74	(14.71 - 31.33)	**3.E-09**	1.E-07
Non-standard amino acids	-1.57	(-4.79 - 1.76)	0.3515	0.4121
Breakdown pathways				
Glucogenic amino acids[4]	-1.58	(-3.42 - 0.29)	0.0968	0.1499
Ketogenic amino acids[5]	4.10	(1.44 - 6.82)	**0.0023**	0.0072
Ureagenic amino acids[6]	-1.53	(-4.18 - 1.19)	0.2674	0.3497
Amino acid ratios				
Citrulline:arginine	-10.94	(-16.59 - -4.90)	**0.0005**	0.0027
Citrulline:ornithine	-6.55	(-11.27 - -1.58)	**0.0104**	0.0234
Fischer ratio[7]	2.71	(0.40 - 5.06)	**0.0211**	0.0421
Ornithine:arginine	-4.69	(-11.86 - 3.07)	0.2286	0.3110
Tyrosine:phenylalanine	2.45	(-0.31 - 5.29)	0.0826	0.1404

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red.

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.7: Geometric mean amino acid concentrations and ratios by fifths of dairy protein intake in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of dairy protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	66.5	(64.7-68.4)	67.1	(65.4-69.0)	67.0	(65.2-68.8)	68.7	(66.9-70.6)	70.1	(68.2-72.1)	0.0462	0.1316
Leucine	121.8	(118.8-124.8)	124.1	(121.2-127.1)	123.9	(121.0-126.9)	127.3	(124.3-130.3)	130.6	(127.5-133.8)	**0.0009**	0.0035
Valine	205.0	(200.8-209.4)	210.3	(206.1-214.7)	208.2	(204.0-212.5)	215.0	(210.6-219.4)	219.5	(215.0-224.1)	**0.0001**	0.0006
Branched-chain aa	394.6	(385.9-403.4)	402.8	(394.2-411.5)	400.4	(391.9-409.1)	412.2	(403.4-421.2)	421.5	(412.4-430.7)	**0.0003**	0.0023
Other essential amino acids												
Histidine	68.6	(67.5-69.7)	69.0	(67.9-70.1)	68.4	(67.4-69.5)	68.9	(67.9-70.0)	68.7	(67.7-69.8)	0.9445	0.9445
Lysine	177.3	(173.3-181.4)	179.4	(175.5-183.5)	181.9	(178.0-186.0)	181.5	(177.6-185.6)	185.4	(181.3-189.6)	0.0982	0.2568
Methionine	18.9	(18.4-19.3)	19.1	(18.7-19.6)	19.1	(18.6-19.5)	19.3	(18.8-19.8)	19.7	(19.2-20.2)	0.2287	0.4319
Phenylalanine	54.5	(53.5-55.4)	55.9	(55.0-56.9)	55.5	(54.5-56.4)	57.4	(56.4-58.4)	57.8	(56.8-58.8)	**1.E-05**	0.0004
Threonine	100.6	(98.3-103.0)	102.5	(100.2-104.8)	102.9	(100.6-105.3)	102.9	(100.6-105.3)	104.0	(101.7-106.4)	0.4063	0.6906
Tryptophan	50.4	(49.5-51.4)	51.0	(50.0-52.0)	51.0	(50.0-52.0)	52.5	(51.6-53.5)	53.5	(52.5-54.6)	**0.0001**	0.0006
Essential aa	871.3	(856.3-886.5)	886.7	(872.0-901.7)	886.8	(872.1-901.7)	901.5	(886.5-916.7)	918.3	(902.8-934.0)	**0.0006**	0.0027
Non-essential amino acids												
Alanine	341.1	(332.8-349.7)	331.1	(323.3-339.1)	328.0	(320.3-336.0)	329.5	(321.7-337.5)	335.3	(327.2-343.5)	0.1825	0.3878
Arginine	47.2	(45.3-49.2)	47.2	(45.4-49.1)	48.2	(46.3-50.2)	48.4	(46.5-50.4)	47.4	(45.5-49.4)	0.8443	0.8971
Asparagine	39.8	(39.0-40.6)	39.7	(38.9-40.6)	39.7	(38.9-40.5)	40.0	(39.2-40.8)	40.7	(39.8-41.5)	0.4805	0.7779
Aspartate	6.2	(6.0-6.5)	6.2	(6.0-6.5)	6.1	(5.8-6.4)	6.1	(5.8-6.4)	6.0	(5.7-6.3)	0.7295	0.8971
Glutamate	53.7	(51.5-56.1)	55.9	(53.7-58.2)	51.4	(49.4-53.5)	52.4	(50.3-54.5)	54.2	(52.1-56.5)	0.0464	0.1316
Glutamine	603.5	(593.6-613.5)	596.4	(587.0-605.9)	604.1	(594.6-613.7)	597.4	(587.9-607.0)	600.2	(590.6-609.9)	0.7262	0.8971
Glycine	175.4	(171.3-179.5)	175.1	(171.2-179.1)	174.3	(170.4-178.3)	169.0	(165.3-172.9)	174.1	(170.2-178.1)	0.1495	0.3388
Proline	182.4	(176.9-188.1)	180.5	(175.2-185.9)	179.1	(173.9-184.4)	185.6	(180.2-191.2)	191.3	(185.7-197.1)	0.0193	0.0656

Continues...

Table 8.7 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of dairy protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	80.6	(78.8-82.5)	81.6	(79.8-83.4)	81.0	(79.2-82.8)	81.2	(79.4-83.0)	82.1	(80.3-84.0)	0.8110	0.8971
Tyrosine	60.3	(58.9-61.8)	60.3	(58.9-61.8)	60.8	(59.4-62.3)	63.6	(62.1-65.1)	64.3	(62.8-65.9)	**0.0001**	0.0006
Non-essential aa	1626	(1,604-1,649)	1608	(1,587-1,629)	1607	(1,585-1,628)	1606	(1,585-1,628)	1630	(1,608-1,652)	0.3715	0.6648
Non-standard amino acids												
Citrulline	29.3	(28.4-30.1)	29.4	(28.6-30.2)	29.0	(28.2-29.8)	29.5	(28.7-30.3)	28.8	(28.0-29.7)	0.8087	0.8971
Ornithine	79.9	(77.6-82.2)	79.6	(77.4-81.9)	78.3	(76.2-80.6)	78.7	(76.5-80.9)	80.1	(77.9-82.4)	0.7765	0.8971
t4-OH-Pro	8.8	(8.4-9.2)	8.5	(8.1-8.9)	8.3	(7.9-8.7)	8.4	(8.0-8.8)	7.6	(7.2-7.9)	**0.0004**	0.0023
Non-standard aa	120.6	(117.8-123.4)	120.4	(117.7-123.2)	118.1	(115.4-120.8)	119.0	(116.4-121.8)	118.9	(116.2-121.7)	0.6928	0.8971
Breakdown pathways												
Glucogenic aa[3]	2199	(2,170-2,228)	2192	(2,165-2,221)	2188	(2,160-2,216)	2200	(2,172-2,228)	2232	(2,204-2,261)	0.2268	0.4319
Ketogenic aa[4]	637.5	(626.0-649.1)	646.7	(635.4-658.1)	650.0	(638.8-661.5)	660.1	(648.7-671.8)	672.7	(660.9-684.7)	**0.0006**	0.0027
Ureagenic aa[5]	165.3	(162.2-168.5)	165.2	(162.2-168.3)	163.6	(160.6-166.6)	165.6	(162.5-168.7)	164.2	(161.1-167.3)	0.8841	0.9109
Amino acid ratios												
Citrulline:arginine	0.6	(0.6-0.6)	0.6	(0.6-0.7)	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.8417	0.8971
Citrulline:ornithine	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.3-0.4)	0.6271	0.8971
Fischer ratio[6]	2.4	(2.3-2.4)	2.4	(2.4-2.4)	2.4	(2.3-2.4)	2.4	(2.3-2.4)	2.4	(2.3-2.4)	0.7741	0.8971
Ornithine:arginine	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.6	(1.5-1.7)	1.6	(1.5-1.7)	1.7	(1.6-1.8)	0.6604	0.8971
Tyrosine:phenylala.	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	0.1305	0.3170

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.8: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in dairy protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in dairy protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	5.43	(1.72 - 9.28)	**0.0038**	0.0129
Leucine	6.89	(3.56 - 10.33)	**4.E-05**	0.0002
Valine	6.62	(3.75 - 9.57)	**4.E-06**	5.E-05
Branched-chain amino acids	6.47	(3.47 - 9.56)	**2.E-05**	0.0001
Other essential amino acids				
Histidine	0.14	(-1.88 - 2.21)	0.8919	0.9190
Lysine	4.08	(1.06 - 7.18)	**0.0077**	0.0239
Methionine	3.76	(0.39 - 7.24)	0.0286	0.0747
Phenylalanine	5.93	(3.54 - 8.38)	**8.E-07**	3.E-05
Threonine	2.84	(-0.20 - 5.97)	0.0678	0.1647
Tryptophan	6.26	(3.63 - 8.97)	**2.E-06**	4.E-05
Essential amino acids	5.01	(2.70 - 7.38)	**2.E-05**	0.0001
Non-essential amino acids				
Alanine	-1.43	(-4.52 - 1.76)	0.3758	0.6084
Arginine	1.36	(-3.88 - 6.88)	0.6189	0.8093
Asparagine	2.04	(-0.67 - 4.83)	0.1409	0.2994
Aspartate	-3.92	(-9.53 - 2.03)	0.1919	0.3624
Glutamate	-1.68	(-6.87 - 3.79)	0.5387	0.7632
Glutamine	-0.38	(-2.47 - 1.75)	0.7230	0.8325
Glycine	-1.98	(-4.87 - 1.00)	0.1902	0.3624
Proline	5.24	(1.17 - 9.47)	**0.0112**	0.0319
Serine	1.28	(-1.65 - 4.29)	0.3949	0.6102
Tyrosine	7.60	(4.25 - 11.06)	**6.E-06**	5.E-05
Non-essential amino acids	0.20	(-1.57 - 1.99)	0.8293	0.9096
Non-standard amino acids				
Citrulline	-0.97	(-4.55 - 2.74)	0.6022	0.8093
Ornithine	-0.16	(-3.83 - 3.65)	0.9328	0.9328
t4-OH-Pro	-11.71	(-16.97 - -6.11)	**0.0001**	0.0003
Non-standard amino acids	-1.51	(-4.42 - 1.50)	0.3215	0.5465
Breakdown pathways				
Glucogenic amino acids[4]	1.40	(-0.32 - 3.14)	0.1106	0.2506
Ketogenic amino acids[5]	5.30	(2.88 - 7.77)	**1.E-05**	0.0001
Ureagenic amino acids[6]	-0.43	(-2.85 - 2.06)	0.7346	0.8325
Amino acid ratios				
Citrulline:arginine	-2.30	(-7.93 - 3.69)	0.4434	0.6555
Citrulline:ornithine	-0.81	(-5.36 - 3.95)	0.7330	0.8325
Fischer ratio[7]	-0.15	(-2.18 - 1.92)	0.8831	0.9190
Ornithine:arginine	-1.50	(-8.22 - 5.72)	0.6758	0.8325
Tyrosine:phenylalanine	1.58	(-0.90 - 4.11)	0.2142	0.3834

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.9: Geometric mean amino acid concentrations and ratios by fifths of plant protein intake in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of plant protein intake (% of energy)										p_{diff}[2]	p_{adj}[2]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	66.9	(65.1-68.9)	68.9	(67.1-70.8)	67.6	(65.8-69.4)	68.4	(66.5-70.3)	67.5	(65.6-69.5)	0.5964	0.8956
Leucine	124.6	(121.6-127.8)	125.7	(122.7-128.8)	124.6	(121.7-127.6)	127.2	(124.1-130.3)	125.4	(122.2-128.6)	0.7735	0.8956
Valine	210.8	(206.3-215.4)	212.2	(207.9-216.7)	210.2	(205.9-214.5)	213.4	(209.0-217.8)	211.3	(206.7-216.0)	0.8691	0.9111
Branched-chain aa	403.7	(394.7-412.9)	408.1	(399.3-417.0)	403.7	(395.0-412.5)	410.1	(401.2-419.1)	405.4	(396.2-414.8)	0.8166	0.8956
Other essential amino acids												
Histidine	67.1	(66.0-68.1)	68.3	(67.3-69.4)	69.0	(68.0-70.1)	69.3	(68.3-70.4)	70.0	(68.8-71.1)	0.0073	0.1875
Lysine	178.5	(174.4-182.6)	182.6	(178.6-186.7)	180.6	(176.6-184.6)	182.4	(178.3-186.5)	181.6	(177.3-185.9)	0.6284	0.8956
Methionine	18.9	(18.4-19.4)	19.3	(18.8-19.8)	19.3	(18.8-19.7)	19.1	(18.7-19.6)	19.4	(18.9-19.9)	0.6459	0.8956
Phenylalanine	55.7	(54.7-56.7)	56.5	(55.6-57.5)	55.9	(55.0-56.9)	56.2	(55.2-57.2)	56.6	(55.5-57.6)	0.7270	0.8956
Threonine	101.7	(99.3-104.1)	103.1	(100.7-105.4)	103.3	(101.0-105.6)	102.6	(100.3-105.0)	102.3	(99.9-104.8)	0.8843	0.9111
Tryptophan	51.7	(50.6-52.7)	51.6	(50.7-52.6)	51.8	(50.9-52.8)	51.5	(50.5-52.5)	51.8	(50.8-52.9)	0.9847	0.9847
Essential aa	884.3	(868.8-900.0)	896.2	(881.1-911.5)	890.8	(875.9-905.9)	898.5	(883.3-913.9)	894.2	(878.3-910.3)	0.7484	0.8956
Non-essential amino acids												
Alanine	323.3	(315.3-331.5)	337.8	(329.8-346.0)	336.5	(328.5-344.6)	333.4	(325.5-341.6)	334.1	(325.7-342.7)	0.1026	0.4053
Arginine	47.9	(45.9-49.9)	47.7	(45.8-49.7)	46.8	(45.0-48.7)	48.5	(46.6-50.5)	47.5	(45.5-49.6)	0.7917	0.8956
Asparagine	39.1	(38.3-39.9)	40.0	(39.2-40.9)	40.0	(39.2-40.8)	40.0	(39.2-40.8)	40.8	(40.0-41.7)	0.1073	0.4053
Aspartate	6.1	(5.8-6.4)	6.3	(6.0-6.6)	6.0	(5.7-6.3)	6.1	(5.8-6.4)	6.2	(5.9-6.5)	0.5860	0.8956
Glutamate	54.4	(52.1-56.8)	54.2	(52.0-56.4)	51.8	(49.7-54.0)	54.2	(52.0-56.5)	53.0	(50.7-55.3)	0.4182	0.7997
Glutamine	591.3	(581.6-601.2)	598.1	(588.6-607.7)	604.4	(594.9-614.0)	598.1	(588.6-607.8)	609.8	(599.5-620.2)	0.1348	0.4584
Glycine	169.4	(165.5-173.4)	173.1	(169.3-177.1)	177.2	(173.2-181.2)	171.5	(167.7-175.4)	176.7	(172.5-181.0)	0.0341	0.2321
Proline	178.8	(173.3-184.4)	184.9	(179.5-190.5)	182.8	(177.5-188.3)	186.4	(180.9-192.1)	185.9	(180.1-191.8)	0.3589	0.7997

Continues...

Table 8.9 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of plant protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	79.9	(78.1-81.8)	81.1	(79.3-82.9)	81.3	(79.5-83.1)	81.7	(79.9-83.5)	82.6	(80.7-84.6)	0.4469	0.7997
Tyrosine	61.2	(59.7-62.8)	61.6	(60.1-63.0)	62.2	(60.7-63.7)	61.8	(60.3-63.3)	62.6	(61.0-64.2)	0.7967	0.8956
Non-essential aa	1584	(1,562-1,606)	1617	(1,596-1,639)	1624	(1,602-1,645)	1617	(1,596-1,639)	1636	(1,613-1,659)	0.0317	0.2321
Non-standard amino acids												
Citrulline	28.6	(27.8-29.4)	28.8	(28.0-29.6)	29.2	(28.4-30.0)	29.3	(28.5-30.2)	30.0	(29.1-30.9)	0.2596	0.7356
Ornithine	75.8	(73.6-78.0)	79.9	(77.7-82.2)	80.2	(78.0-82.5)	79.5	(77.3-81.8)	81.3	(78.9-83.8)	0.0156	0.1875
t4-OH-Pro	8.3	(7.9-8.7)	8.5	(8.1-8.9)	8.5	(8.1-8.9)	8.2	(7.9-8.6)	7.9	(7.6-8.3)	0.3512	0.7997
Non-standard aa	115.1	(112.4-117.8)	119.8	(117.1-122.6)	120.4	(117.7-123.1)	119.7	(117.0-122.4)	122.2	(119.3-125.2)	0.0131	0.1763
Breakdown pathways												
Glucogenic aa[3]	2166	(2,137-2,195)	2206	(2,178-2,234)	2210	(2,182-2,238)	2207	(2,179-2,236)	2224	(2,194-2,254)	0.0993	0.4053
Ketogenic aa[4]	647.1	(635.3-659.0)	656.2	(644.7-667.8)	652.4	(641.0-663.9)	656.7	(645.1-668.5)	654.1	(642.0-666.5)	0.8006	0.8956
Ureagenic aa[5]	160.7	(157.6-163.8)	164.8	(161.8-167.9)	164.6	(161.6-167.7)	165.9	(162.8-169.0)	167.9	(164.6-171.2)	0.0523	0.2965
Amino acid ratios												
Citrulline:arginine	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.6	(0.6-0.7)	0.6	(0.6-0.6)	0.6	(0.6-0.7)	0.4359	0.7997
Citrulline:ornithine	0.4	(0.4-0.4)	0.4	(0.3-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4379	0.7997
Fischer ratio[6]	2.4	(2.3-2.4)	2.4	(2.4-2.4)	2.4	(2.3-2.4)	2.4	(2.4-2.4)	2.4	(2.3-2.4)	0.3181	0.7997
Ornithine:arginine	1.6	(1.5-1.7)	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.6	(1.6-1.7)	1.7	(1.6-1.8)	0.2285	0.7063
Tyrosine:phenylala.	1.1	(1.1-1.1)	1.1	(1.1 - 1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	0.6244	0.8956

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.10: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in plant protein intake (% of energy) in 1,697 men in EPIC – minimally-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in plant protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	0.42	(-3.36 - 4.35)	0.8318	0.8838
Leucine	0.94	(-2.43 - 4.44)	0.5888	0.8007
Valine	0.41	(-2.49 - 3.40)	0.7821	0.8838
Branched-chain amino acids	0.55	(-2.50 - 3.69)	0.7286	0.8838
Other essential amino acids				
Histidine	4.09	(1.85 - 6.38)	**0.0003**	0.0106
Lysine	1.33	(-1.81 - 4.57)	0.4093	0.6710
Methionine	1.93	(-1.61 - 5.59)	0.2894	0.5467
Phenylalanine	0.94	(-1.51 - 3.46)	0.4540	0.6710
Threonine	0.38	(-2.79 - 3.66)	0.8148	0.8838
Tryptophan	0.11	(-2.56 - 2.85)	0.9374	0.9658
Essential amino acids	1.01	(-1.38 - 3.46)	0.4105	0.6710
Non-essential amino acids				
Alanine	2.17	(-1.25 - 5.71)	0.2165	0.4601
Arginine	0.05	(-5.46 - 5.89)	0.9853	0.9853
Asparagine	3.48	(0.55 - 6.51)	0.0197	0.1056
Aspartate	-0.69	(-6.88 - 5.90)	0.8318	0.8838
Glutamate	-2.10	(-7.61 - 3.74)	0.4719	0.6710
Glutamine	2.48	(0.18 - 4.82)	0.0341	0.1287
Glycine	3.07	(-0.18 - 6.41)	0.0641	0.2010
Proline	3.51	(-0.77 - 7.97)	0.1097	0.3107
Serine	2.99	(-0.18 - 6.27)	0.0650	0.2010
Tyrosine	1.97	(-1.43 - 5.50)	0.2592	0.5183
Non-essential amino acids	2.62	(0.69 - 4.58)	0.0075	0.0510
Non-standard amino acids				
Citrulline	4.56	(0.54 - 8.75)	0.0258	0.1098
Ornithine	5.61	(1.47 - 9.92)	0.0075	0.0510
t4-OH-Pro	-4.75	(-10.83 - 1.74)	0.1480	0.3872
Non-standard amino acids	4.86	(1.55 - 8.27)	0.0037	0.0510
Breakdown pathways				
Glucogenic amino acids[4]	2.15	(0.31 - 4.03)	0.0217	0.1056
Ketogenic amino acids[5]	0.92	(-1.57 - 3.47)	0.4736	0.6710
Ureagenic amino acids[6]	3.81	(1.12 - 6.57)	0.0053	0.0510
Amino acid ratios				
Citrulline:arginine	4.51	(-1.92 - 11.36)	0.1735	0.3932
Citrulline:ornithine	-0.99	(-5.83 - 4.10)	0.6970	0.8777
Fischer ratio[7]	-0.49	(-2.65 - 1.71)	0.6597	0.8627
Ornithine:arginine	5.55	(-2.12 - 13.83)	0.1605	0.3899
Tyrosine:phenylalanine	1.02	(-1.61 - 3.72)	0.4504	0.6710

[1] Adjusted for energy intake (fourths of the distribution), study centre and metabolite assay batch

[2] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.11: Partial correlation coefficients between protein intake (% of energy) and amino acid concentrations and ratios in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Total protein			Animal protein			Dairy protein			Plant protein		
	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2	r	p^2	p_{adj}^2
Branched-chain essential amino acids												
Isoleucine	0.05	0.0454	0.1102	0.05	0.0400	0.0756	0.06	0.0264	0.0899	0.00	0.9869	0.9869
Leucine	0.07	**0.0033**	0.0190	0.07	**0.0052**	0.0161	0.08	**0.0010**	0.0038	0.01	0.8239	0.9099
Valine	0.08	**0.0007**	0.0083	0.08	**0.0011**	0.0074	0.11	**2.E-05**	0.0002	-0.01	0.7414	0.9099
Branched-chain amino acids	0.08	**0.0022**	0.0147	0.07	**0.0027**	0.0112	0.09	**0.0002**	0.0012	0.00	0.9182	0.9637
Other essential amino acids												
Histidine	-0.01	0.6454	0.7567	-0.04	0.0769	0.1377	-0.03	0.2565	0.4172	0.08	**0.0010**	0.0348
Lysine	0.07	**0.0072**	0.0305	0.06	**0.0145**	0.0410	0.05	0.0365	0.1034	0.01	0.7178	0.9099
Methionine	-0.02	0.4941	0.6865	-0.02	0.3712	0.4507	0.04	0.1558	0.3006	0.03	0.2057	0.4372
Phenylalanine	0.04	0.1372	0.2592	0.02	0.4601	0.5394	0.11	**1.E-05**	0.0002	0.02	0.5317	0.8412
Threonine	-0.01	0.6731	0.7628	-0.01	0.7766	0.7766	0.04	0.0766	0.1999	0.00	0.9354	0.9637
Tryptophan	0.02	0.3734	0.5770	0.02	0.3366	0.4401	0.09	**0.0001**	0.0009	-0.01	0.6175	0.8749
Essential amino acids	0.07	**0.0088**	0.0316	0.06	**0.0175**	0.0426	0.09	**0.0003**	0.0015	0.01	0.7777	0.9099
Non-essential amino acids												
Alanine	-0.04	0.0791	0.1581	-0.06	**0.0170**	0.0426	-0.01	0.5891	0.7418	0.04	0.1248	0.3465
Arginine	0.02	0.4692	0.6865	0.02	0.4955	0.5435	0.00	0.9948	0.9948	-0.01	0.8045	0.9099
Asparagine	-0.05	0.0526	0.1191	-0.08	**0.0017**	0.0085	0.02	0.5160	0.6747	0.06	0.0089	0.0783
Aspartate	-0.02	0.3164	0.5140	-0.03	0.3071	0.4177	-0.04	0.1102	0.2378	-0.02	0.5443	0.8412
Glutamate	-0.01	0.7863	0.8420	0.01	0.6033	0.6410	-0.01	0.7270	0.8562	-0.02	0.3215	0.6072
Glutamine	-0.05	0.0596	0.1266	-0.07	**0.0030**	0.0112	-0.02	0.3917	0.5490	0.06	0.0092	0.0783
Glycine	-0.09	**0.0002**	0.0028	-0.12	**2.E-06**	4.E-05	-0.04	0.0823	0.1999	0.06	0.0167	0.0944
Proline	-0.02	0.5441	0.6865	-0.04	0.1271	0.2058	0.05	0.0317	0.0981	0.04	0.1420	0.3465
Serine	0.02	0.5452	0.6865	-0.02	0.4775	0.5412	0.01	0.8109	0.8893	0.05	0.0562	0.1912
Tyrosine	0.02	0.3175	0.5140	0.02	0.3580	0.4507	0.11	**2.E-05**	0.0002	0.02	0.5148	0.8412
Non-essential amino acids	-0.06	0.0221	0.0577	-0.09	**0.0005**	0.0041	0.00	0.8846	0.9114	0.07	0.0048	0.0783

Continues...

Table 8.11 continued

Amino acids	Total protein			Animal protein			Dairy protein			Plant protein		
	r	p^2	$p_{adj}{}^2$	r	p^2	$p_{adj}{}^2$	r	p^2	$p_{adj}{}^2$	r	p^2	$p_{adj}{}^2$
Non-standard amino acids												
Citrulline	-0.08	**0.0015**	0.0126	-0.10	**4.E-05**	0.0005	-0.03	0.1680	0.3006	0.03	0.1633	0.3702
Ornithine	0.00	0.9699	0.9699	-0.03	0.2588	0.3768	-0.01	0.7693	0.8719	0.06	0.0160	0.0944
t4-OH-Pro	0.12	**1.E-06**	4.E-05	0.15	**7.E-10**	2.E-08	-0.11	**5.E-06**	0.0002	-0.04	0.1156	0.3465
Non-standard amino acids	0.00	0.8920	0.9318	-0.03	0.2228	0.3443	-0.04	0.1438	0.3006	0.06	0.0133	0.0810
Breakdown pathways												
Glucogenic amino acids[3]	-0.03	0.2724	0.4875	-0.05	0.0369	0.0738	0.03	0.2700	0.4172	0.05	0.0273	0.1035
Ketogenic amino acids[4]	0.06	0.0193	0.0548	0.05	0.0285	0.0606	0.09	**0.0002**	0.0012	0.01	0.8296	0.9099
Ureagenic amino acids[5]	-0.02	0.5095	0.6865	-0.04	0.1047	0.1780	-0.03	0.2643	0.4172	0.05	0.0274	0.1035
Amino acid ratios												
Citrulline:arginine	-0.06	**0.0093**	0.0316	-0.08	**0.0018**	0.0085	-0.02	0.3941	0.5490	0.03	0.2810	0.5620
Citrulline:ornithine	-0.06	**0.0148**	0.0458	-0.06	0.0231	0.0523	-0.02	0.4037	0.5490	-0.02	0.3937	0.7045
Fischer ratio[6]	0.07	**0.0063**	0.0304	0.07	**0.0034**	0.0115	-0.01	0.7303	0.8562	-0.01	0.6028	0.8749
Ornithine:arginine	-0.01	0.5729	0.6957	-0.03	0.2660	0.3768	0.00	0.8723	0.9114	0.04	0.1427	0.3465
Tyrosine:phenylalanine	0.00	0.9044	0.9318	0.01	0.6322	0.6513	0.03	0.1629	0.3006	0.01	0.8075	0.9099

[1] Adjusted for energy intake (fourths of the distribution), study centre, metabolite assay batch, age group (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years) and BMI (fourths of the distribution; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.12: Geometric mean amino acid concentrations and ratios by fifths of total protein intake in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of total protein intake (% of energy)										p_{diff}[2]	p_{adj}[2]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	66.2	(64.3-68.1)	68.6	(66.7-70.5)	67.0	(65.2-68.8)	69.0	(67.2-70.9)	68.7	(66.7-70.8)	0.1735	0.3537
Leucine	121.9	(118.9-125.0)	126.2	(123.2-129.2)	123.0	(120.2-126.0)	128.4	(125.4-131.5)	128.1	(124.8-131.4)	0.0133	0.0953
Valine	206.2	(201.8-210.6)	212.5	(208.1-216.9)	209.6	(205.4-213.8)	213.4	(209.1-217.8)	216.4	(211.6-221.2)	0.0386	0.1876
Branched-chain aa	395.6	(386.8-404.6)	408.5	(399.7-417.4)	400.9	(392.5-409.5)	411.9	(403.2-420.8)	414.3	(404.7-424.1)	0.0350	0.1876
Other essential amino acids												
Histidine	69.4	(68.3-70.6)	68.3	(67.3-69.4)	68.3	(67.3-69.4)	69.1	(68.0-70.2)	68.5	(67.4-69.7)	0.5098	0.6844
Lysine	178.3	(174.2-182.6)	180.4	(176.4-184.6)	178.1	(174.2-182.0)	184.3	(180.2-188.4)	184.5	(180.1-189.0)	0.1068	0.2922
Methionine	19.3	(18.8-19.8)	19.1	(18.6-19.6)	19.1	(18.7-19.6)	19.5	(19.0-20.0)	19.0	(18.5-19.5)	0.7121	0.8289
Phenylalanine	55.5	(54.5-56.5)	55.9	(54.9-56.9)	56.2	(55.3-57.2)	56.7	(55.7-57.6)	56.7	(55.6-57.7)	0.5234	0.6844
Threonine	103.2	(100.8-105.7)	102.0	(99.7-104.4)	102.1	(99.8-104.4)	103.8	(101.4-106.2)	101.9	(99.4-104.4)	0.7314	0.8289
Tryptophan	51.7	(50.6-52.7)	51.1	(50.1-52.1)	51.6	(50.6-52.6)	52.1	(51.2-53.1)	52.0	(50.9-53.1)	0.6485	0.7874
Essential aa	879.4	(864.0-895.2)	892.2	(877.1-907.5)	883.6	(869.0-898.4)	904.4	(889.4-919.8)	904.5	(888.0-921.2)	0.1117	0.2922
Non-essential amino acids												
Alanine	341.4	(332.8-350.1)	336.6	(328.5-344.9)	325.0	(317.4-332.8)	332.7	(324.8-340.8)	329.5	(320.9-338.2)	0.0734	0.2329
Arginine	47.1	(45.2-49.2)	47.6	(45.7-49.6)	46.0	(44.2-47.9)	49.3	(47.4-51.3)	48.4	(46.3-50.5)	0.1848	0.3537
Asparagine	40.8	(40.0-41.7)	39.7	(38.9-40.5)	39.9	(39.1-40.7)	40.0	(39.2-40.8)	39.5	(38.6-40.3)	0.2260	0.3659
Aspartate	6.2	(5.9-6.5)	6.0	(5.8-6.3)	6.2	(6.0-6.5)	6.1	(5.8-6.4)	6.1	(5.8-6.4)	0.8115	0.8900
Glutamate	53.3	(51.1-55.6)	53.9	(51.7-56.1)	54.5	(52.4-56.6)	53.2	(51.1-55.3)	52.8	(50.5-55.1)	0.8516	0.9048
Glutamine	612.4	(602.3-622.7)	599.4	(589.9-609.0)	592.4	(583.2-601.7)	597.9	(588.5-607.4)	599.5	(589.3-609.9)	0.0753	0.2329
Glycine	181.7	(177.5-185.9)	173.5	(169.7-177.4)	171.4	(167.7-175.1)	170.6	(166.9-174.4)	170.8	(166.8-174.9)	**0.0009**	0.0161
Proline	185.8	(180.0-191.8)	184.7	(179.2-190.4)	177.8	(172.6-183.2)	185.8	(180.3-191.5)	184.7	(178.7-190.8)	0.1995	0.3570

Continues...

Table 8.12 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of total protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	80.6	(78.7-82.5)	80.4	(78.6-82.2)	81.1	(79.4-82.9)	82.2	(80.4-84.0)	82.3	(80.3-84.2)	0.5830	0.7341
Tyrosine	60.9	(59.4-62.4)	61.4	(60.0-62.9)	62.1	(60.7-63.6)	63.1	(61.7-64.7)	61.7	(60.1-63.3)	0.3168	0.4896
Non-essential aa	1646	(1,623-1,669)	1618	(1,596-1,640)	1589	(1,569-1,611)	1617	(1,595-1,638)	1608	(1,584-1,631)	0.0140	0.0953
Non-standard amino acids												
Citrulline	29.9	(29.1-30.8)	29.7	(28.9-30.6)	29.0	(28.2-29.8)	29.2	(28.4-30.0)	28.1	(27.3-29.0)	0.0630	0.2329
Ornithine	79.3	(76.9-81.7)	79.3	(77.1-81.6)	79.7	(77.5-82.0)	78.9	(76.7-81.2)	79.4	(77.0-81.9)	0.9928	0.9991
t4-OH-Pro	7.7	(7.3-8.0)	8.3	(7.9-8.7)	8.0	(7.7-8.4)	8.4	(8.0-8.8)	9.2	(8.7-9.7)	**0.0001**	0.0024
Non-standard aa	119.4	(116.6-122.3)	119.7	(117.0-122.5)	119.1	(116.5-121.8)	119.2	(116.5-122.0)	119.5	(116.6-122.6)	0.9981	0.9981
Breakdown pathways												
Glucogenic aa[3]	2226	(2,196-2,256)	2205	(2,176-2,234)	2173	(2,145-2,201)	2209	(2,180-2,237)	2200	(2,169-2,231)	0.1467	0.3326
Ketogenic aa[4]	643.7	(631.9-655.8)	651.8	(640.3-663.5)	646.9	(635.8-658.3)	663.7	(652.1-675.5)	660.5	(647.9-673.3)	0.1283	0.3117
Ureagenic aa[5]	165.1	(161.9-168.4)	166.0	(162.9-169.2)	162.1	(159.2-165.1)	166.8	(163.8-170.0)	163.8	(160.5-167.1)	0.2215	0.3659
Amino acid ratios												
Citrulline:arginine	0.6	(0.6-0.7)	0.6	(0.6-0.7)	0.6	(0.6-0.7)	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.0523	0.2223
Citrulline:ornithine	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.3-0.4)	0.1872	0.3537
Fischer ratio[6]	2.3	(2.3-2.4)	2.4	(2.4-2.4)	2.3	(2.3-2.4)	2.4	(2.3-2.4)	2.4	(2.4-2.5)	0.0091	0.0953
Ornithine:arginine	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.6	(1.5-1.7)	1.6	(1.5-1.7)	0.3499	0.5173
Tyrosine:phenylala.	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	0.5174	0.6844

[1] Adjusted for energy intake (fourths of the distribution), study centre, metabolite assay batch, age group (< 55; $55-59$; $60-64$; $65-69$; ≥ 70 years) and BMI (fourths; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.13: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in total protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in total protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	3.38	(-0.63 - 7.56)	0.0994	0.2088
Leucine	4.78	(1.18 - 8.50)	0.0089	0.0503
Valine	4.10	(1.04 - 7.25)	0.0084	0.0503
Branched-chain amino acids	4.12	(0.89 - 7.44)	0.0119	0.0578
Other essential amino acids				
Histidine	-0.61	(-2.84 - 1.67)	0.5982	0.7823
Lysine	3.62	(0.29 - 7.06)	0.0331	0.1125
Methionine	-0.23	(-3.84 - 3.51)	0.9003	0.9566
Phenylalanine	2.27	(-0.30 - 4.91)	0.0839	0.2039
Threonine	-0.28	(-3.54 - 3.10)	0.8707	0.9566
Tryptophan	1.35	(-1.44 - 4.22)	0.3451	0.5229
Essential amino acids	2.85	(0.34 - 5.42)	0.0259	0.1101
Non-essential amino acids				
Alanine	-3.32	(-6.67 - 0.16)	0.0614	0.1606
Arginine	3.60	(-2.34 - 9.89)	0.2405	0.3894
Asparagine	-2.40	(-5.25 - 0.53)	0.1076	0.2088
Aspartate	-1.85	(-8.19 - 4.94)	0.5848	0.7823
Glutamate	-1.19	(-6.80 - 4.76)	0.6876	0.8658
Glutamine	-1.87	(-4.11 - 0.43)	0.1105	0.2088
Glycine	-5.61	(-8.60 - -2.51)	**0.0005**	0.0079
Proline	-0.32	(-4.61 - 4.17)	0.8879	0.9566
Serine	2.56	(-0.69 - 5.91)	0.1238	0.2215
Tyrosine	2.32	(-1.17 - 5.92)	0.1950	0.3315
Non-essential amino acids	-1.93	(-3.85 - 0.02)	0.0519	0.1470
Non-standard amino acids				
Citrulline	-5.36	(-9.09 - -1.47)	0.0074	0.0503
Ornithine	-0.07	(-4.15 - 4.20)	0.9752	0.9755
t4-OH-Pro	15.88	(8.24 - 24.05)	**2.E-05**	0.0008
Non-standard amino acids	-0.09	(-3.38 - 3.30)	0.9562	0.9755
Breakdown pathways				
Glucogenic amino acids[4]	-0.89	(-2.76 - 1.01)	0.3545	0.5229
Ketogenic amino acids[5]	2.86	(0.23 - 5.55)	0.0326	0.1125
Ureagenic amino acids[6]	-0.40	(-3.10 - 2.36)	0.7716	0.9370
Amino acid ratios				
Citrulline:arginine	-8.64	(-14.44 - -2.45)	0.0069	0.0503
Citrulline:ornithine	-5.30	(-10.08 - -0.25)	0.0398	0.1231
Fischer ratio[7]	1.95	(-0.34 - 4.29)	0.0959	0.2088
Ornithine:arginine	-3.53	(-10.82 - 4.35)	0.3691	0.5229
Tyrosine:phenylalanine	0.04	(-2.64 - 2.80)	0.9755	0.9755

[1] Adjusted for energy intake (fourths), study centre, metabolite assay batch, age group (<55; $55-59$; $60-64$; $65-69$; ≥ 70 years) and BMI (fourths; unknown)

[2] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.14: Geometric mean amino acid concentrations and ratios by fifths of animal protein intake in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of animal protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	65.4	(63.6-67.3)	68.1	(66.3-70.0)	69.4	(67.5-71.2)	67.6	(65.8-69.5)	69.0	(67.0 - 71.0)	0.0364	0.1234
Leucine	121.1	(118.1-124.2)	125.5	(122.5-128.6)	126.9	(123.9-129.9)	125.4	(122.4-128.4)	128.6	(125.4 - 132.0)	0.0258	0.0976
Valine	205.4	(201.0-209.8)	212.2	(207.8-216.6)	213.4	(209.2-217.8)	209.5	(205.3-213.9)	217.4	(212.7 - 222.2)	0.0060	0.0507
Branched-chain aa	393.3	(384.5-402.2)	406.9	(398.2-415.9)	411.2	(402.5-420.0)	403.5	(394.9-412.3)	416.3	(406.9 - 426.0)	0.0102	0.0614
Other essential amino acids												
Histidine	69.3	(68.2-70.5)	69.0	(67.9-70.1)	68.6	(67.6-69.7)	68.3	(67.2-69.3)	68.5	(67.4 - 69.6)	0.7524	0.8778
Lysine	177.7	(173.6-181.9)	179.3	(175.3-183.4)	182.1	(178.1-186.2)	180.7	(176.7-184.8)	185.8	(181.4 - 190.3)	0.1353	0.3286
Methionine	19.2	(18.7-19.7)	19.3	(18.8-19.8)	19.2	(18.7-19.6)	19.2	(18.8-19.7)	19.1	(18.6 - 19.6)	0.9869	0.9869
Phenylalanine	55.4	(54.4-56.4)	56.2	(55.2-57.2)	56.8	(55.9-57.8)	56.1	(55.2-57.1)	56.3	(55.3 - 57.4)	0.4128	0.6564
Threonine	102.7	(100.3-105.2)	101.2	(98.9-103.6)	103.6	(101.3-106.0)	103.0	(100.6-105.4)	102.5	(100.0 - 105.0)	0.6825	0.8778
Tryptophan	51.2	(50.2-52.3)	51.5	(50.6-52.5)	51.5	(50.6-52.5)	51.9	(50.9-52.9)	52.3	(51.2 - 53.3)	0.7745	0.8778
Essential aa	875.3	(859.9-891.0)	890.2	(875.1-905.5)	900.6	(885.7-915.7)	889.6	(874.6-904.8)	908.4	(892.1 - 925.0)	0.0659	0.1868
Non-essential amino acids												
Alanine	340.3	(331.8-349.0)	336.7	(328.6-345.0)	331.8	(323.9-339.8)	326.4	(318.6-334.5)	329.9	(321.5 - 338.5)	0.2214	0.4704
Arginine	46.8	(44.8-48.8)	47.6	(45.7-49.6)	46.9	(45.0-48.8)	48.4	(46.5-50.4)	48.8	(46.7 - 51.0)	0.6012	0.8517
Asparagine	40.8	(40.0-41.7)	39.8	(39.0-40.6)	39.9	(39.1-40.7)	39.9	(39.1-40.7)	39.5	(38.7 - 40.4)	0.2900	0.5189
Aspartate	6.2	(5.9-6.5)	6.2	(5.9-6.5)	6.0	(5.7-6.2)	6.3	(6.0-6.6)	6.1	(5.8 - 6.4)	0.6285	0.8547
Glutamate	51.9	(49.8-54.1)	55.0	(52.8-57.2)	54.9	(52.8-57.1)	52.7	(50.6-54.8)	53.2	(50.9 - 55.5)	0.1779	0.4033
Glutamine	617.2	(607.0-627.5)	596.4	(586.9-606.0)	593.8	(584.6-603.1)	594.9	(585.5-604.4)	599.5	(589.5 - 609.7)	0.0060	0.0507
Glycine	182.7	(178.5-187.0)	174.5	(170.7-178.4)	170.4	(166.7-174.1)	169.4	(165.7-173.2)	171.1	(167.1 - 175.1)	**4.E-05**	0.0008
Proline	185.4	(179.6-191.3)	186.3	(180.7-192.1)	184.1	(178.7-189.6)	181.9	(176.5-187.5)	181.1	(175.3 - 187.0)	0.7503	0.8778

Continues...

Table 8.14 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of animal protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	81.1	(79.2-83.0)	81.3	(79.5-83.1)	80.8	(79.1-82.6)	81.7	(79.9-83.5)	81.6	(79.7 - 83.6)	0.9619	0.9869
Tyrosine	60.6	(59.1-62.2)	62.2	(60.8-63.8)	61.7	(60.3-63.2)	62.7	(61.2-64.2)	62.1	(60.5 - 63.7)	0.4247	0.6564
Non-essential aa	1647	(1,624-1,671)	1619	(1,597-1,641)	1607	(1,585-1,628)	1598	(1,576-1,619)	1606	(1,583 - 1,629)	0.0399	0.1234
Non-standard amino acids												
Citrulline	30.2	(29.3-31.1)	29.6	(28.7-30.4)	29.4	(28.6-30.2)	28.9	(28.1-29.7)	28.0	(27.2 - 28.8)	0.0161	0.0684
Ornithine	80.0	(77.7-82.5)	79.2	(76.9-81.5)	80.3	(78.1-82.6)	78.2	(76.0-80.5)	78.9	(76.6 - 81.4)	0.7472	0.8778
t4-OH-Pro	7.5	(7.1-7.9)	7.9	(7.5-8.2)	8.4	(8.1-8.8)	8.6	(8.2-9.0)	9.1	(8.7 - 9.6)	**2.E-06**	0.0001
Non-standard aa	120.3	(117.4-123.2)	119.0	(116.2-121.7)	120.7	(118.0-123.4)	118.4	(115.7-121.2)	118.7	(115.8 - 121.6)	0.7493	0.8495
Breakdown pathways												
Glucogenic aa[3]	2225	(2,195-2,256)	2205	(2,177-2,235)	2199	(2,171-2,228)	2182	(2,153-2,210)	2200	(2,170 - 2,231)	0.4135	0.6564
Ketogenic aa[4]	640.3	(628.5-652.3)	650.2	(638.7-661.9)	658.9	(647.5-670.5)	653.7	(642.2-665.4)	663.5	(651.1 - 676.2)	0.1189	0.3110
Ureagenic aa[5]	165.4	(162.2-168.7)	164.5	(161.5-167.7)	166.0	(163.0-169.1)	164.1	(161.0-167.2)	163.7	(160.5 - 167.0)	0.8568	0.9103
Amino acid ratios												
Citrulline:arginine	0.6	(0.6-0.7)	0.6	(0.6-0.6)	0.6	(0.6-0.7)	0.6	(0.6-0.6)	0.6	(0.5 - 0.6)	0.0158	0.0684
Citrulline:ornithine	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.3 - 0.4)	0.2367	0.4734
Fischer ratio[6]	2.3	(2.3-2.4)	2.4	(2.3-2.4)	2.4	(2.4-2.4)	2.3	(2.3-2.4)	2.4	(2.4 - 2.5)	0.0108	0.0614
Ornithine:arginine	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.6	(1.5-1.7)	1.6	(1.5 - 1.7)	0.4500	0.6652
Tyrosine:phenylala.	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1 - 1.1)	0.2696	0.5093

[1] Adjusted for energy intake (fourths of the distribution), study centre, metabolite assay batch, age group (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years) and BMI (fourths; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red.

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.15: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in animal protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in animal protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	4.21	(0.12 - 8.46)	0.0433	0.0945
Leucine	4.95	(1.31 - 8.72)	**0.0074**	0.0386
Valine	4.21	(1.12 - 7.40)	**0.0073**	0.0386
Branched-chain amino acids	4.39	(1.13 - 7.76)	**0.0079**	0.0386
Other essential amino acids				
Histidine	-1.40	(-3.64 - 0.88)	0.2271	0.3218
Lysine	3.98	(0.61 - 7.47)	0.0204	0.0631
Methionine	-0.57	(-4.20 - 3.20)	0.7632	0.8073
Phenylalanine	1.34	(-1.24 - 3.98)	0.3126	0.4252
Threonine	0.48	(-2.84 - 3.92)	0.7793	0.8073
Tryptophan	1.84	(-0.99 - 4.75)	0.2044	0.3184
Essential amino acids	3.02	(0.48 - 5.62)	0.0196	0.0631
Non-essential amino acids				
Alanine	-3.68	(-7.06 - -0.18)	0.0396	0.0945
Arginine	4.09	(-1.93 - 10.49)	0.1868	0.3176
Asparagine	-2.56	(-5.44 - 0.40)	0.0889	0.1778
Aspartate	-0.94	(-7.41 - 5.98)	0.7836	0.8073
Glutamate	0.45	(-5.31 - 6.55)	0.8827	0.8827
Glutamine	-2.50	(-4.75 - -0.19)	0.0338	0.0945
Glycine	-6.39	(-9.39 - -3.30)	**0.0001**	0.0012
Proline	-2.77	(-7.00 - 1.65)	0.2154	0.3184
Serine	0.74	(-2.48 - 4.07)	0.6550	0.7423
Tyrosine	2.25	(-1.27 - 5.89)	0.2126	0.3184
Non-essential amino acids	-2.56	(-4.48 - -0.61)	**0.0105**	0.0445
Non-standard amino acids				
Citrulline	-6.73	(-10.44 - -2.86)	**0.0008**	0.0090
Ornithine	-1.58	(-5.64 - 2.66)	0.4598	0.5791
t4-OH-Pro	21.56	(13.50 - 30.19)	**3.E-08**	9.E-07
Non-standard amino acids	-1.24	(-4.52 - 2.15)	0.4689	0.5905
Breakdown pathways				
Glucogenic amino acids[4]	-1.34	(-3.21 - 0.57)	0.1681	0.3009
Ketogenic amino acids[5]	3.15	(0.49 - 5.87)	0.0200	0.0631
Ureagenic amino acids[6]	-0.93	(-3.64 - 1.85)	0.5059	0.6109
Amino acid ratios				
Citrulline:arginine	-10.40	(-16.13 - -4.26)	**0.0012**	0.0099
Citrulline:ornithine	-5.23	(-10.07 - -0.13)	0.0445	0.0945
Fischer ratio[7]	2.43	(0.11 - 4.81)	0.0400	0.0945
Ornithine:arginine	-5.45	(-12.66 - 2.36)	0.1662	0.3009
Tyrosine:phenylalanine	0.90	(-1.83 - 3.71)	0.5210	0.6109

[1] Adjusted for energy intake (fourths), study centre, metabolite assay batch, age group (<55; $55-59$; $60-64$; $65-69$; ≥ 70 years) and BMI (fourths; unknown)

[2] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.16: Geometric mean amino acid concentrations and ratios by fifths of dairy protein intake in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of dairy protein intake (% of energy)										p_{diff}[2]	p_{adj}[2]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	66.6	(64.8-68.5)	67.2	(65.4-69.0)	66.9	(65.2-68.7)	68.6	(66.8-70.5)	70.1	(68.3-72.1)	0.0519	0.1470
Leucine	122.0	(119.1-125.0)	124.1	(121.2-127.0)	123.8	(120.9-126.7)	127.1	(124.1-130.1)	130.6	(127.5-133.8)	**0.0011**	0.0042
Valine	205.4	(201.2-209.7)	210.2	(206.0-214.4)	208.3	(204.2-212.5)	214.6	(210.3-219.0)	219.5	(215.1-224.1)	**0.0001**	0.0010
Branched-chain aa	395.3	(386.7-404.0)	402.7	(394.3-411.3)	400.4	(392.0-408.9)	411.5	(402.9-420.4)	421.5	(412.6-430.6)	**0.0004**	0.0029
Other essential amino acids												
Histidine	68.7	(67.6-69.8)	69.0	(68.0-70.1)	68.3	(67.3-69.4)	69.0	(67.9-70.0)	68.6	(67.6-69.7)	0.9004	0.9256
Lysine	177.6	(173.6-181.7)	179.4	(175.5-183.4)	182.0	(178.0-186.0)	181.3	(177.4-185.4)	185.3	(181.2-189.5)	0.1250	0.3268
Methionine	18.9	(18.4-19.4)	19.2	(18.7-19.6)	19.0	(18.6-19.5)	19.3	(18.8-19.8)	19.6	(19.1-20.1)	0.3135	0.5922
Phenylalanine	54.5	(53.5-55.5)	55.9	(55.0-56.9)	55.5	(54.6-56.5)	57.3	(56.3-58.3)	57.8	(56.8-58.8)	**2.E-05**	0.0007
Threonine	100.6	(98.3-103.0)	102.4	(100.2-104.8)	102.6	(100.4-105.0)	103.3	(101.0-105.6)	104.0	(101.6-106.4)	0.3984	0.7130
Tryptophan	50.5	(49.6-51.5)	51.0	(50.1-52.0)	50.9	(50.0-51.9)	52.5	(51.5-53.5)	53.5	(52.5-54.6)	**0.0001**	0.0010
Essential aa	872.6	(857.6-887.7)	886.7	(872.1-901.6)	886.4	(871.8-901.2)	900.8	(885.9-916.0)	918.1	(902.7-933.7)	**0.0009**	0.0040
Non-essential amino acids												
Alanine	341.3	(333.0-349.9)	331.1	(323.3-339.1)	328.6	(320.8-336.5)	329.0	(321.2-337.0)	335.1	(327.1-343.3)	0.1783	0.4041
Arginine	47.2	(45.3-49.2)	47.2	(45.3-49.1)	48.2	(46.3-50.1)	48.5	(46.6-50.5)	47.3	(45.5-49.3)	0.8234	0.9216
Asparagine	39.8	(39.0-40.7)	39.8	(39.0-40.6)	39.6	(38.8-40.4)	40.1	(39.3-40.9)	40.6	(39.8-41.4)	0.4771	0.7725
Aspartate	6.3	(6.0-6.6)	6.2	(6.0-6.5)	6.1	(5.8-6.3)	6.1	(5.8-6.4)	6.0	(5.7-6.3)	0.6501	0.9210
Glutamate	54.0	(51.8-56.2)	55.9	(53.8-58.2)	51.5	(49.5-53.6)	52.0	(49.9-54.1)	54.3	(52.2-56.5)	0.0256	0.0791
Glutamine	602.8	(593.1-612.7)	596.5	(587.2-605.9)	603.4	(594.0-612.9)	598.9	(589.6-608.4)	599.9	(590.4-609.5)	0.8403	0.9216
Glycine	174.9	(171.0-178.9)	175.2	(171.4-179.1)	174.0	(170.2-177.8)	169.8	(166.1-173.5)	173.9	(170.1-177.8)	0.2823	0.5647
Proline	182.2	(176.6-187.9)	180.5	(175.2-185.9)	179.1	(173.9-184.5)	185.9	(180.5-191.5)	191.3	(185.6-197.1)	0.0194	0.0661

Continues...

Table 8.16 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of dairy protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	80.6	(78.8-82.4)	81.7	(79.9-83.5)	80.8	(79.0-82.5)	81.4	(79.7-83.2)	82.1	(80.3-83.9)	0.7631	0.9216
Tyrosine	60.4	(59.0-61.9)	60.3	(58.9-61.7)	61.0	(59.6-62.4)	63.3	(61.9-64.8)	64.4	(62.9-65.9)	**0.0001**	0.0011
Non-essential aa	1625	(1,603-1,648)	1608	(1,587-1,630)	1606	(1,584-1,627)	1608	(1,587-1,630)	1629	(1,607-1,651)	0.4299	0.7309
Non-standard amino acids												
Citrulline	29.2	(28.4-30.0)	29.3	(28.5-30.1)	29.1	(28.3-29.9)	29.5	(28.7-30.4)	28.8	(28.1-29.7)	0.8237	0.9216
Ornithine	79.7	(77.4-82.1)	79.6	(77.4-81.9)	78.4	(76.2-80.6)	78.7	(76.5-81.0)	80.2	(77.9-82.5)	0.7808	0.9216
t4-OH-Pro	8.8	(8.4-9.2)	8.5	(8.1-8.9)	8.3	(7.9-8.7)	8.3	(8.0-8.7)	7.6	(7.2-7.9)	**0.0005**	0.0029
Non-standard aa	120.4	(117.6-123.2)	120.3	(117.6-123.0)	118.2	(115.6-120.9)	119.1	(116.4-121.8)	119.0	(116.3-121.8)	0.8089	0.9216
Breakdown pathways												
Glucogenic aa[3]	2199	(2,170-2,228)	2193	(2,165-2,221)	2187	(2,159-2,215)	2202	(2,174-2,230)	2231	(2,203-2,261)	0.2371	0.5039
Ketogenic aa[4]	638.3	(626.8-649.9)	646.7	(635.5-658.0)	649.8	(638.6-661.2)	659.6	(648.2-671.2)	672.6	(660.9-684.6)	**0.0009**	0.0040
Ureagenic aa[5]	165.1	(162.0-168.3)	165.0	(162.0-168.1)	163.7	(160.7-166.7)	165.8	(162.7-168.9)	164.3	(161.2-167.4)	0.8947	0.9256
Amino acid ratios												
Citrulline:arginine	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.9256	0.9256
Citrulline:ornithine	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.3-0.4)	0.5385	0.8323
Fischer ratio[6]	2.4	(2.3-2.4)	2.4	(2.4-2.4)	2.4	(2.3-2.4)	2.4	(2.3-2.4)	2.4	(2.3-2.4)	0.8255	0.9216
Ornithine:arginine	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.6	(1.5-1.7)	1.6	(1.5-1.7)	1.7	(1.6-1.8)	0.6485	0.9210
Tyrosine:phenylala.	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	0.1386	0.3367

[1] Adjusted for energy intake (fourths of the distribution), study centre, metabolite assay batch, age group (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years) and BMI (fourths; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.17: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in dairy protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in dairy protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	5.25	(1.55 - 9.08)	**0.0050**	0.0171
Leucine	6.70	(3.39 - 10.11)	**0.0001**	0.0002
Valine	6.42	(3.60 - 9.32)	**6.E-06**	0.0001
Branched-chain amino acids	6.28	(3.31 - 9.33)	**3.E-05**	0.0001
Other essential amino acids				
Histidine	-0.10	(-2.13 - 1.97)	0.9244	0.9524
Lysine	3.88	(0.85 - 6.99)	**0.0116**	0.0329
Methionine	3.35	(-0.03 - 6.84)	0.0523	0.1368
Phenylalanine	5.86	(3.46 - 8.31)	**1.E-06**	4.E-05
Threonine	2.96	(-0.09 - 6.11)	0.0569	0.1382
Tryptophan	5.99	(3.37 - 8.68)	**6.E-06**	0.0001
Essential amino acids	4.85	(2.54 - 7.21)	**3.E-05**	0.0002
Non-essential amino acids				
Alanine	-1.57	(-4.67 - 1.63)	0.3313	0.5517
Arginine	1.24	(-4.02 - 6.79)	0.6496	0.8835
Asparagine	1.93	(-0.76 - 4.70)	0.1617	0.3234
Aspartate	-4.25	(-9.86 - 1.71)	0.1589	0.3234
Glutamate	-2.30	(-7.33 - 3.00)	0.3877	0.5732
Glutamine	-0.23	(-2.30 - 1.88)	0.8273	0.9273
Glycine	-1.72	(-4.55 - 1.20)	0.2454	0.4590
Proline	5.41	(1.30 - 9.68)	**0.0094**	0.0291
Serine	1.38	(-1.53 - 4.37)	0.3570	0.5517
Tyrosine	7.37	(4.08 - 10.77)	**8.E-06**	0.0001
Non-essential amino acids	0.25	(-1.52 - 2.05)	0.7856	0.9210
Non-standard amino acids				
Citrulline	-0.59	(-4.15 - 3.11)	0.7516	0.9210
Ornithine	0.10	(-3.61 - 3.95)	0.9590	0.9590
t4-OH-Pro	-11.76	(-17.04 - -6.14)	**0.0001**	0.0003
Non-standard amino acids	-1.23	(-4.17 - 1.80)	0.4216	0.6232
Breakdown pathways				
Glucogenic amino acids[4]	1.40	(-0.32 - 3.16)	0.1115	0.2528
Ketogenic amino acids[5]	5.15	(2.73 - 7.63)	**2.E-05**	0.0001
Ureagenic amino acids[6]	-0.22	(-2.66 - 2.28)	0.8617	0.9273
Amino acid ratios				
Citrulline:arginine	-1.81	(-7.47 - 4.20)	0.5469	0.7748
Citrulline:ornithine	-0.69	(-5.24 - 4.09)	0.7739	0.9210
Fischer ratio[7]	-0.17	(-2.20 - 1.91)	0.8728	0.9273
Ornithine:arginine	-1.13	(-7.91 - 6.15)	0.7536	0.9210
Tyrosine:phenylalanine	1.43	(-1.03 - 3.95)	0.2565	0.4590

[1] Adjusted for energy intake (fourths), study centre, metabolite assay batch, age group (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years) and BMI (fourths; unknown)

[2] Calculated as $(e^{\beta} - 1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.18: Percentage difference in amino acid concentrations and ratios for an 80 percentile increment in plant protein intake (% of energy) in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Difference in amino acids per 80 percentile increment in plant protein intake[2], %		p_{trend}[3]	p_{adj}[3]
	Estimate	(95% CI)		
Brancehd-chain essential amino acids				
Isoleucine	0.85	(-2.91 - 4.77)	0.6609	0.8025
Leucine	1.39	(-1.96 - 4.86)	0.4202	0.5953
Valine	0.92	(-1.94 - 3.86)	0.5325	0.6963
Branched-chain amino acids	1.02	(-1.99 - 4.12)	0.5109	0.6949
Other essential amino acids				
Histidine	4.12	(1.89 - 6.41)	**0.0003**	0.0092
Lysine	1.54	(-1.60 - 4.78)	0.3392	0.5320
Methionine	2.00	(-1.54 - 5.67)	0.2726	0.5149
Phenylalanine	1.17	(-1.28 - 3.68)	0.3508	0.5320
Threonine	0.13	(-3.02 - 3.39)	0.9356	0.9640
Tryptophan	0.39	(-2.26 - 3.12)	0.7740	0.8489
Essential amino acids	1.28	(-1.10 - 3.71)	0.2949	0.5214
Non-essential amino acids				
Alanine	2.34	(-1.08 - 5.87)	0.1825	0.3760
Arginine	-0.08	(-5.59 - 5.75)	0.9785	0.9785
Asparagine	3.10	(0.21 - 6.08)	0.0353	0.1501
Aspartate	-0.66	(-6.85 - 5.93)	0.8389	0.8913
Glutamate	-1.03	(-6.43 - 4.69)	0.7185	0.8224
Glutamine	2.09	(-0.15 - 4.39)	0.0681	0.2573
Glycine	2.39	(-0.75 - 5.62)	0.1365	0.3574
Proline	3.45	(-0.83 - 7.92)	0.1157	0.3574
Serine	2.53	(-0.59 - 5.75)	0.1126	0.3574
Tyrosine	2.56	(-0.80 - 6.03)	0.1367	0.3574
Non-essential amino acids	2.45	(0.53 - 4.40)	0.0123	0.0834
Non-standard amino acids				
Citrulline	4.26	(0.30 - 8.37)	0.0348	0.1501
Ornithine	5.59	(1.45 - 9.91)	0.0078	0.0660
t4-OH-Pro	-4.32	(-10.42 - 2.19)	0.1880	0.3760
Non-standard amino acids	4.81	(1.50 - 8.22)	0.0041	0.0660
Breakdown pathways				
Glucogenic amino acids[4]	2.09	(0.25 - 3.97)	0.0261	0.1478
Ketogenic amino acids[5]	1.17	(-1.32 - 3.71)	0.3599	0.5320
Ureagenic amino acids[6]	3.69	(1.00 - 6.45)	0.0069	0.0660
Amino acid ratios				
Citrulline:arginine	4.34	(-2.04 - 11.14)	0.1870	0.3760
Citrulline:ornithine	-1.26	(-6.07 - 3.79)	0.6173	0.7774
Fischer ratio[7]	-0.39	(-2.54 - 1.81)	0.7256	0.8224
Ornithine:arginine	5.67	(-2.01 - 13.96)	0.1515	0.3680
Tyrosine:phenylalanine	1.37	(-1.24 - 4.05)	0.3067	0.5214

[1] Adjusted for energy intake (fourths), study centre, metabolite assay batch, age group (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years) and BMI (fourths; unknown)

[2] Calculated as $(e^{\beta} - 1) \times 100$; β is the regression coefficient from the linear regression

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 8.19: Geometric mean amino acid concentrations and ratios by fifths of plant protein intake in 1,697 men in EPIC – multivariable-adjusted model[1]

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of plant protein intake (% of energy)											
	1		2		3		4		5		p_{diff}[2]	p_{adj}[2]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	66.8	(65.0-68.7)	68.9	(67.1-70.8)	67.6	(65.8-69.4)	68.3	(66.4-70.1)	67.9	(65.9-69.8)	0.5946	0.8451
Leucine	124.4	(121.4-127.5)	125.6	(122.7-128.7)	124.5	(121.6-127.5)	127.0	(124.0-130.0)	125.9	(122.8-129.1)	0.7758	0.8852
Valine	210.3	(205.9-214.8)	212.0	(207.8-216.4)	210.2	(206.0-214.4)	213.0	(208.7-217.4)	212.3	(207.7-216.9)	0.8730	0.8995
Branched-chain aa	402.9	(394.0-411.9)	407.8	(399.2-416.6)	403.6	(395.1-412.2)	409.4	(400.7-418.3)	407.2	(398.1-416.6)	0.8210	0.8995
Other essential amino acids												
Histidine	67.1	(66.0-68.1)	68.3	(67.3-69.4)	69.0	(67.9-70.0)	69.3	(68.2-70.4)	70.0	(68.9-71.2)	0.0067	0.1936
Lysine	178.3	(174.3-182.5)	182.5	(178.5-186.6)	180.6	(176.6-184.6)	182.2	(178.2-186.3)	181.9	(177.7-186.2)	0.6200	0.8451
Methionine	18.9	(18.4-19.4)	19.3	(18.8-19.8)	19.2	(18.8-19.7)	19.1	(18.7-19.6)	19.4	(18.9-20.0)	0.6172	0.8451
Phenylalanine	55.7	(54.7-56.7)	56.5	(55.5-57.4)	55.9	(55.0-56.9)	56.2	(55.2-57.2)	56.6	(55.6-57.7)	0.6907	0.8852
Threonine	101.7	(99.3-104.1)	103.2	(100.9-105.6)	103.3	(101.0-105.7)	102.5	(100.2-104.9)	102.2	(99.8-104.7)	0.8507	0.8995
Tryptophan	51.6	(50.6-52.6)	51.6	(50.7-52.6)	51.8	(50.8-52.8)	51.4	(50.5-52.4)	52.0	(50.9-53.0)	0.9551	0.9551
Essential aa	883.3	(868.0-898.8)	895.9	(881.0-911.1)	890.6	(875.9-905.6)	897.5	(882.4-912.7)	896.6	(880.8-912.7)	0.7035	0.8852
Non-essential amino acids												
Alanine	323.2	(315.2-331.3)	337.3	(329.3-345.5)	336.8	(328.8-344.9)	333.6	(325.7-341.7)	334.3	(325.9-342.9)	0.1026	0.4610
Arginine	47.9	(46.0-49.9)	47.7	(45.9-49.7)	46.8	(44.9-48.7)	48.5	(46.6-50.5)	47.5	(45.5-49.5)	0.7810	0.8852
Asparagine	39.1	(38.3-40.0)	40.1	(39.3-40.9)	40.0	(39.2-40.8)	39.9	(39.1-40.8)	40.8	(39.9-41.6)	0.1484	0.5607
Aspartate	6.1	(5.8-6.4)	6.3	(6.0-6.6)	6.0	(5.7-6.3)	6.1	(5.8-6.4)	6.2	(5.9-6.5)	0.6214	0.8451
Glutamate	54.2	(52.0-56.5)	54.0	(51.9-56.2)	51.8	(49.8-53.9)	54.1	(52.0-56.3)	53.5	(51.3-55.8)	0.4805	0.8451
Glutamine	592.2	(582.6-601.9)	598.7	(589.4-608.2)	604.5	(595.1-614.0)	598.2	(588.8-607.7)	608.0	(598.0-618.2)	0.2278	0.7041
Glycine	169.8	(166.0-173.7)	173.5	(169.7-177.4)	177.2	(173.4-181.1)	171.4	(167.7-175.3)	175.9	(171.9-180.1)	0.0532	0.3619
Proline	178.8	(173.3-184.4)	184.9	(179.5-190.6)	182.9	(177.6-188.5)	186.2	(180.7-191.9)	185.9	(180.1-191.9)	0.3773	0.8451

Continues...

Table 8.19 continued

Amino acids	Geometric mean amino acid concentrations (µmol/L) and ratios by fifths of plant protein intake (% of energy)										p_{diff}[2]	p_{adj}[2]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	80.0	(78.2-81.9)	81.2	(79.5-83.0)	81.2	(79.5-83.0)	81.7	(79.9-83.5)	82.4	(80.5-84.3)	0.5774	0.8451
Tyrosine	61.1	(59.6-62.6)	61.4	(60.0-62.9)	62.1	(60.7-63.6)	61.8	(60.4-63.3)	62.8	(61.3-64.4)	0.6056	0.8451
Non-essential aa	1585	(1,563-1,607)	1618	(1,596-1,639)	1624	(1,602-1,646)	1617	(1,595-1,639)	1634	(1,611-1,657)	0.0452	0.3619
Non-standard amino acids												
Citrulline	28.6	(27.8-29.5)	28.8	(28.0-29.6)	29.3	(28.5-30.1)	29.4	(28.6-30.2)	29.9	(29.0-30.7)	0.3155	0.8451
Ornithine	75.8	(73.6-78.0)	79.9	(77.7-82.2)	80.3	(78.1-82.6)	79.5	(77.3-81.8)	81.3	(78.9-83.8)	0.0149	0.1936
t4-OH-Pro	8.3	(7.9-8.7)	8.5	(8.1-8.9)	8.5	(8.1-8.9)	8.2	(7.9-8.6)	8.0	(7.6-8.4)	0.4051	0.8451
Non-standard aa	115.1	(112.4-117.8)	119.8	(117.1-122.5)	120.5	(117.9-123.3)	119.7	(117.0-122.5)	122.0	(119.2-125.0)	0.0132	0.1692
Breakdown pathways												
Glucogenic aa[3]	2166	(2,138-2,196)	2206	(2,178-2,235)	2210	(2,182-2,239)	2206	(2,178-2,235)	2223	(2,193-2,254)	0.1085	0.4610
Ketogenic aa[4]	646.4	(634.7-658.3)	656.0	(644.5-667.6)	652.2	(640.9-663.7)	656.1	(644.6-667.8)	655.8	(643.7-668.1)	0.7624	0.8852
Ureagenic aa[5]	160.7	(157.7-163.9)	164.8	(161.8-167.9)	164.8	(161.8-167.9)	165.9	(162.8-169.0)	167.6	(164.4-171.0)	0.0651	0.3687
Amino acid ratios												
Citrulline:arginine	0.6	(0.6-0.6)	0.6	(0.6-0.6)	0.6	(0.6-0.7)	0.6	(0.6-0.6)	0.6	(0.6-0.7)	0.4271	0.8451
Citrulline:ornithine	0.4	(0.4-0.4)	0.4	(0.3-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4	(0.4-0.4)	0.4040	0.8451
Fischer ratio[6]	2.4	(2.3-2.4)	2.4	(2.4-2.4)	2.4	(2.3-2.4)	2.4	(2.4-2.4)	2.4	(2.3-2.4)	0.4271	0.8451
Ornithine:arginine	1.6	(1.5-1.7)	1.7	(1.6-1.8)	1.7	(1.6-1.8)	1.6	(1.6-1.7)	1.7	(1.6-1.8)	0.2081	0.7041
Tyrosine:phenylala.	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	1.1	(1.1-1.1)	0.5513	0.8451

[1] Adjusted for energy intake (fourths of the distribution), study centre, metabolite assay batch, age group (< 55; $55-59$; $60-64$; $65-69$; ≥ 70 years) and BMI (fourths; unknown)

[2] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[3] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[4] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[5] Ureagenic: arg + cit + orn

[6] Fischer ratio: Branched-chain/(phe + trp + tyr)

8.8 Figures

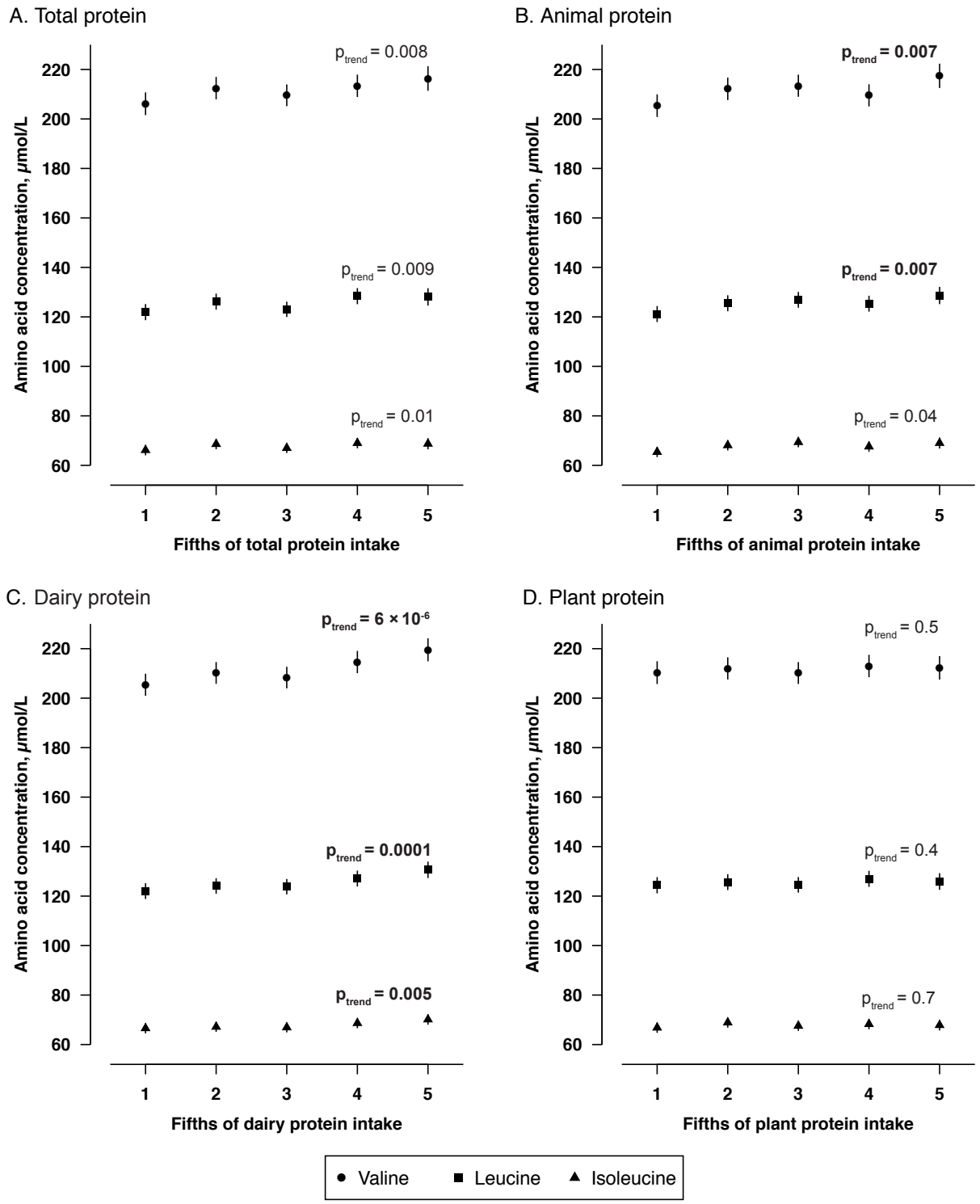


Figure 8.1: Mean concentrations of branched-chain amino acids by fifths of protein intakes in 1,697 men in EPIC

A. Total protein intake. B. Animal protein intake. C. Dairy protein intake. D. Plant protein intake. Included amino acids were significantly associated with intake of one or more protein types after correcting for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg); such significance is marked in bold.

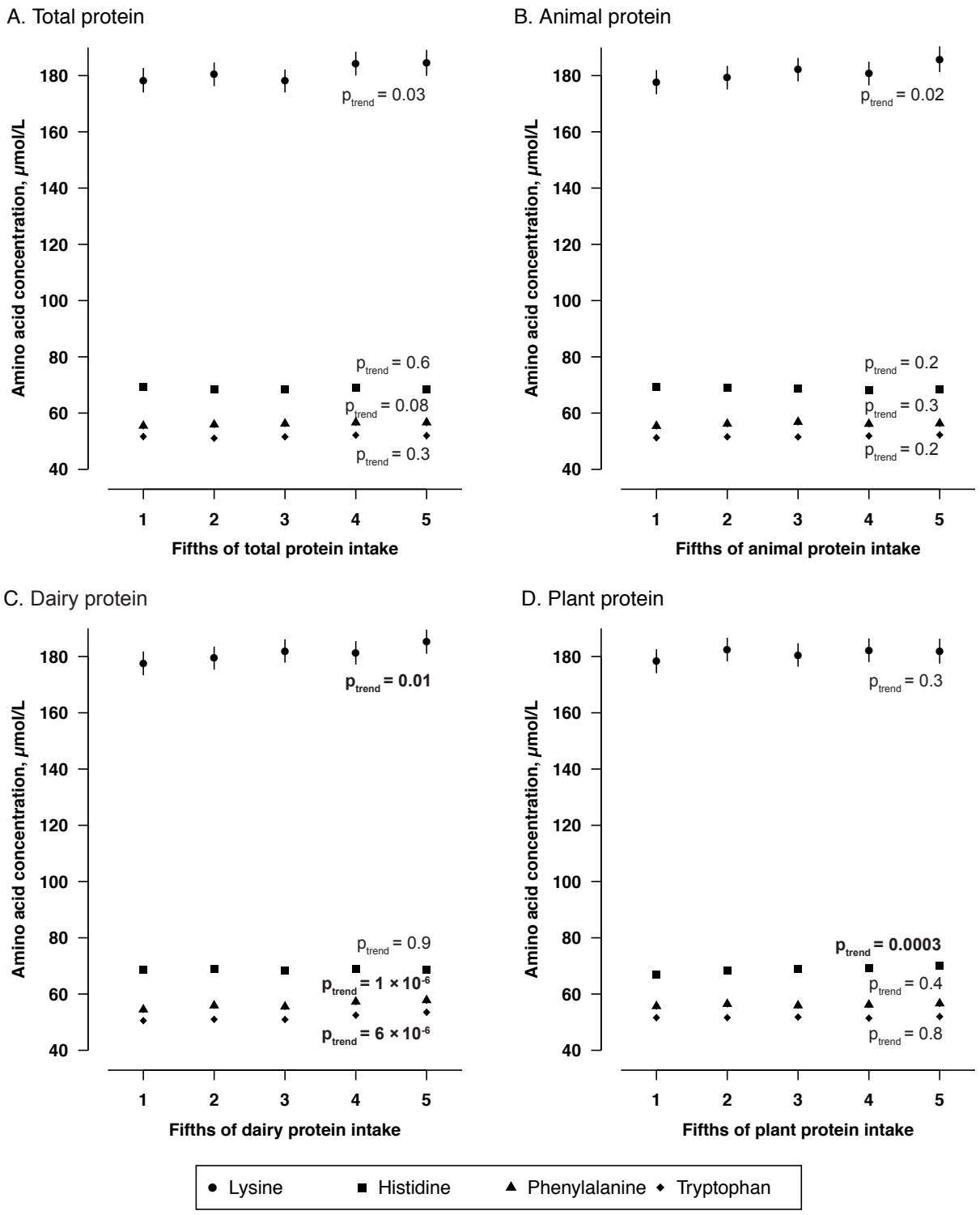


Figure 8.2: Mean concentrations of selected essential amino acids by fifths of protein intakes in 1,697 men in EPIC

A. Total protein intake. B. Animal protein intake. C. Dairy protein intake. D. Plant protein intake. Included amino acids were significantly associated with intake of one or more protein types after correcting for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg); such significance is marked in bold.

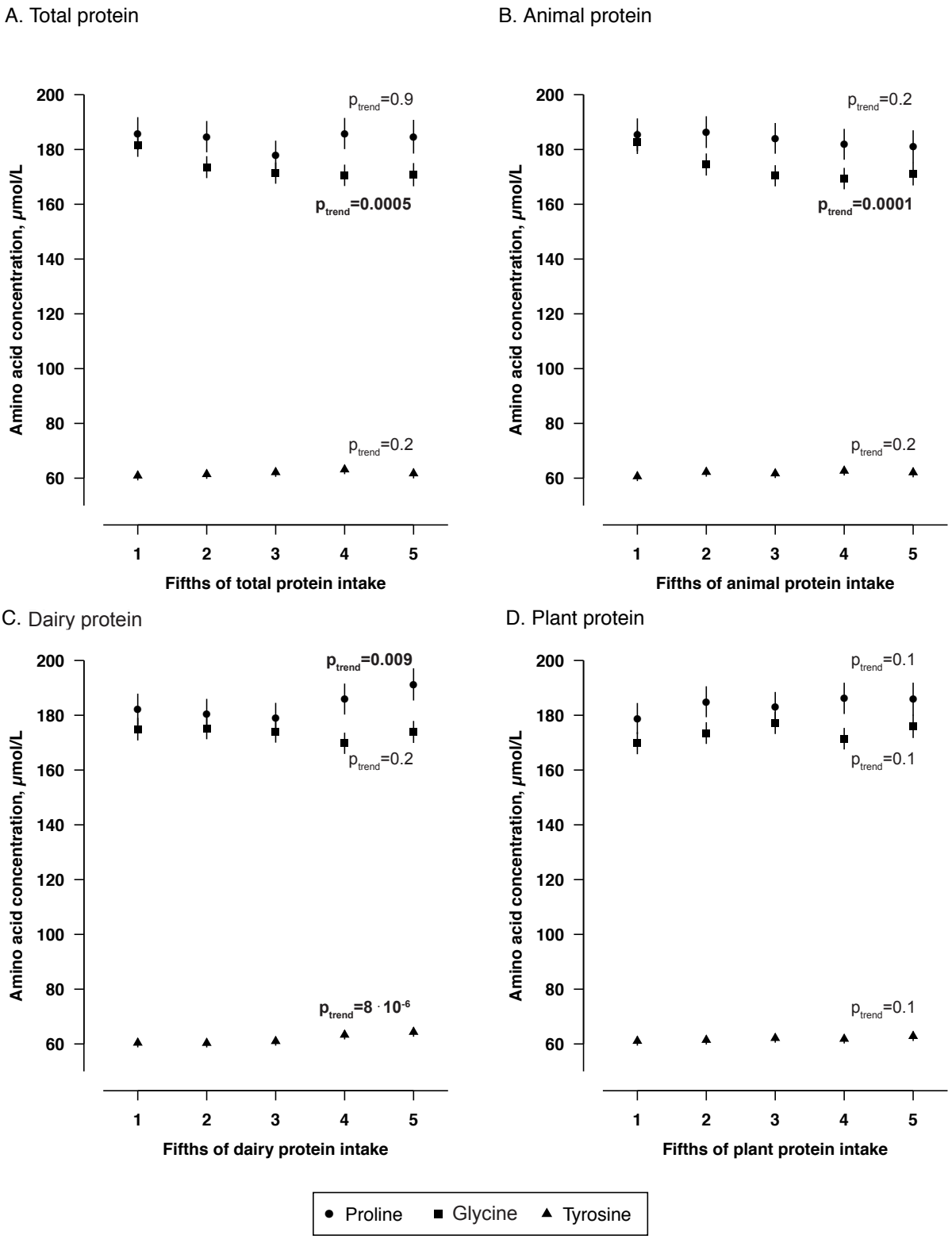


Figure 8.3: Mean concentrations of selected non-essential amino acids by fifths of protein intakes in 1,697 men in EPIC

A. Total protein intake. B. Animal protein intake. C. Dairy protein intake. D. Plant protein intake. Included amino acids were significantly associated with intake of one or more protein types after correcting for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg); such significance is marked in bold.

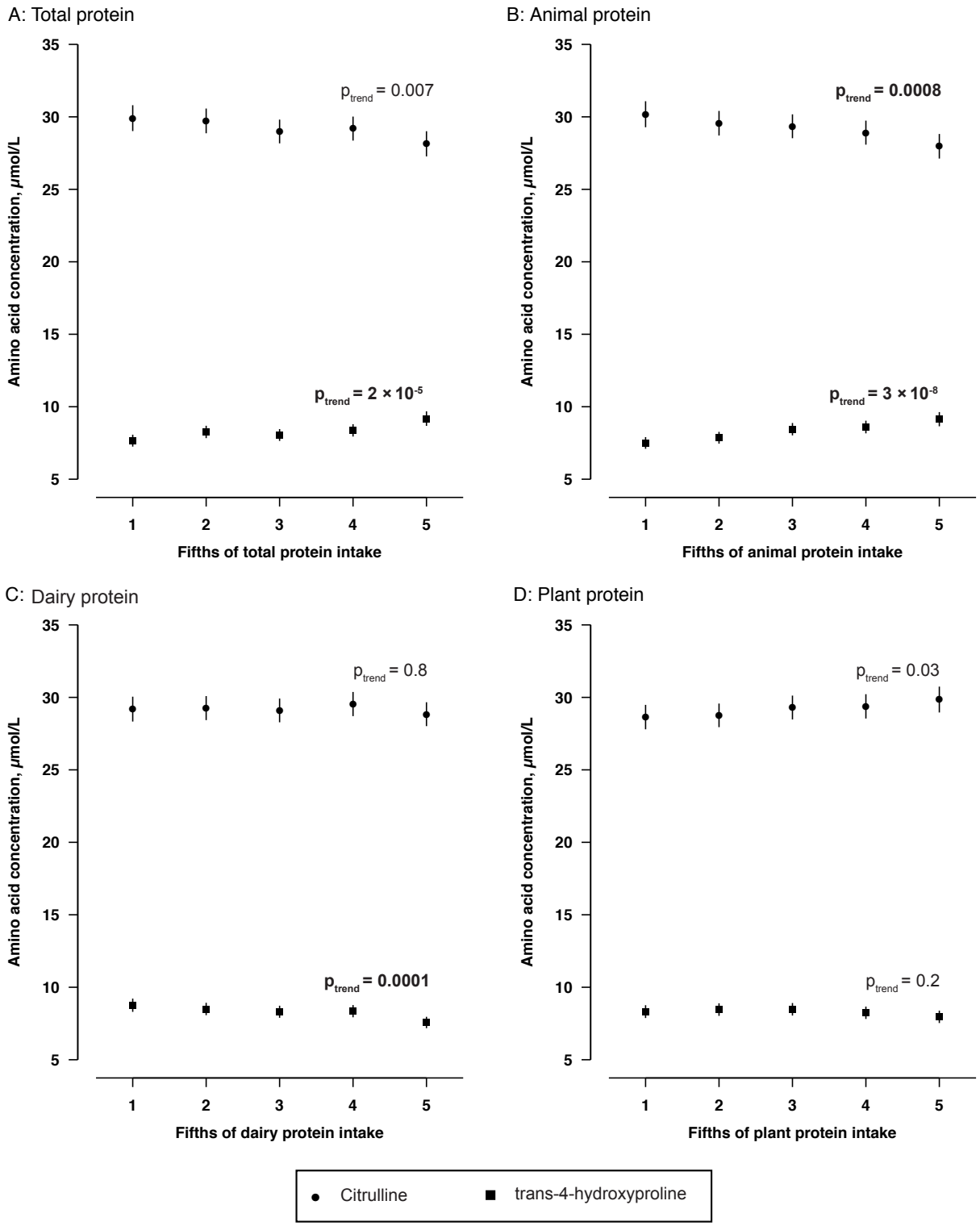


Figure 8.4: Mean concentrations of selected non-standard amino acids by fifths of protein intakes in 1,697 men in EPIC

A. Total protein intake. B. Animal protein intake. C. Dairy protein intake. D. Plant protein intake. Included amino acids were significantly associated with intake of one or more protein types after correcting for multiple testing using a false discovery rate controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg); such significance is marked in bold.

9. Concentrations of amino acids and insulin-like growth factor-I

9.1 Introduction

Circulating concentration of insulin-like growth factor-I (IGF-I) is the only potentially modifiable risk factor for prostate cancer [2], and it is also positively related to risk of other cancers, including breast cancer [235]. IGF-I is a hormone with mitogenic and anti-apoptotic effects exerted through the PI3K-Akt-mTOR pathway [3, 89]. This pathway is upregulated in 30 to 50% of prostate cancers and may be important for prostate cancer initiation and progression [4, 89].

Studies have found that higher intake of protein especially from dairy products is related to a higher circulating concentration of IGF-I [6, 7, 75, 76], but it is not clear if this association is explained by certain amino acids. The associations between protein intake and plasma amino acid concentrations were investigated in Chapter 8. Among other findings, we observed associations of animal protein intake with glycine (inverse) and t4-OH-Pro (positive), and of dairy protein intake with branched-chain and other essential amino acids (all positive). The current chapter investigates the relationships between concentrations of amino acids and IGF-I. Taken together, this work might help to inform the development of strategies for prevention of IGF-I-related cancers, including prostate cancer.

9.1.1 Previous studies

Studies of intake of essential amino acids, which are more abundant in animal protein than plant protein [128], support a role of these amino acids in the association between protein intake and IGF-I. Higher circulating concentrations of IGF-I have been observed in individuals with higher intake of protein rich in essential amino acids (i.e. protein from animals and soya) [79]. Moreover, experimental studies have found that intake of essential amino acids might increase plasma IGF-I concentration and IGF-I gene expression in humans and rodents [86, 87, 88]. In contrast, habitual intake of protein low in essential amino acids (i.e. non-soya plant protein) was associated with lower IGF-I concentration [79]. Furthermore, vegans who do not consume animal products (including dairy), and thus have a diet lower in essential amino acids [79, 80], have been observed to have lower IGF-I concentrations than meat-eaters and vegetarians [79, 81].

Less is known about the association of IGF-I with circulating concentrations of amino acids, which are affected by dietary, lifestyle, genetic and homeostatic factors [14, 129], and might thus better reflect the endogenous exposure than estimated intake. To my knowledge, only a single observational study on the association between circulating concentrations of amino acids and

IGF-I has been published. It was conducted in 229 pre-menopausal women and reported inverse correlations of IGF-I with arginine and the sum of ureagenic amino acids (i.e. arginine, ornithine and citrulline) [78]. In contrast, no results for men, that I am aware of, have been published.

9.2 Aim

The aim of the current chapter is to investigate associations of plasma amino acid concentrations, sums and ratios with serum IGF-I in men from the European Prospective Investigation into Cancer and Nutrition (EPIC).

9.3 Methods

9.3.1 Study population

The EPIC study is a prospective cohort aiming to investigate associations between diet and cancer risk in men and women from ten European countries; details of the cohort have been described previously (Section 3.1 on page 21) [99].

All participants gave written informed consent to participate in the EPIC cohort. Approval for the current study was obtained from the Internal Review Board of the International Agency for Research on Cancer (IARC), Lyon, France (Project No. 14-09) and participating centres as required.

Participants were eligible for the current cross-sectional analysis if they were control participants from the nested case-control study of metabolites patterns and risk of prostate cancer (Chapter 7). These were 1,699 men from 15 centres in Germany, Greece, Italy, Netherlands, Spain and United Kingdom, with no cancer (except non-melanoma skin cancer) at blood collection or at diagnosis of their index case. Assays of IGF-I have not been conducted for follow-up phase III and thus, 557 men without IGF-I data were excluded, leaving 1,142 men for further analyses.

9.3.2 Blood collection and laboratory analysis

The procedure for collection of blood samples has been described previously (Section 3.2.1 on page 23).

9.3.3 Measurement of amino acids

As detailed in Section 8.3.3 on page 273, concentrations of 22 amino acids were measured using the Absolute*IDQ*® p180 Kit (Biocrates Life Sciences AG, Innsbrcuk, Austria) and ultra-high

performance liquid chromatography coupled with mass spectrometry. All 22 measured amino acid concentrations were included in the current analysis.

9.3.3.1 Measurement of IGF-I

Concentrations of IGF-I were measured for a previous nested case-control study of IGF-I and prostate cancer risk [15, 16]. Samples from follow-up phase I and II were measured at IARC [15] and the Cancer Epidemiology Unit laboratory, Oxford, UK [16], respectively. All samples were measured in serum using the DSL-10-5600 ACTIVE® ELISA (Diagnostic System Laboratories, Webster, Texas, USA). To avoid interference of IGF binding proteins, the IGF-I assay included an acid-ethanol precipitation of IGF-I binding proteins. Duplicate quality control samples were included in each batch; the intra- and inter-batch coefficients of variation (CV) were 3.0% and 13.7% for phase I and 4.4% and 12.4% for phase II, respectively [15, 16].

9.3.4 Statistical analysis

9.3.4.1 Preparation of variables

From the 22 amino acid concentrations, seven variables summing up amino acids from various classes (e.g. branched-chain amino acids) and five ratios between amino acids were calculated (see details in Section 8.3.4.1 on page 273).

Concentrations of amino acids, sums and ratios, and IGF-I were logarithmically-transformed to approximate the normal distribution and log-transformed variables were used for all analyses.

To control for any effect on the measured IGF-I concentrations of differences in blood collection procedures between centres and the use of two laboratories, correction of the logarithmically transformed IGF-I concentrations was applied. A multiple linear regression of centre, IGF-I batch and measured IGF-I was fitted, and the corrected IGF-I was then calculated by adding the difference between the mean measured IGF-I concentration and the concentration predicted from the regression to the measured IGF-I concentration (i.e. corrected IGF-I = measured IGF-I + mean measured IGF-I − predicted IGF-I). The corrected IGF-I values were used to investigate the associations between IGF-I and amino acid concentrations.

9.3.4.2 Participant characteristics

Participant characteristics, selected nutrient intakes and factors related to blood collection were presented as means and percentages, for continuous and categorical variables, respectively, with 95% confidence intervals (CI) adjusted for study centre by thirds of IGF-I concentration.

9.3.4.3 Controlling for confounding

For all following analyses, a minimally-adjusted model, including study centre and metabolite batch, was fitted to account for any variation in the measurement of amino acid concentration which might arise from differences in blood handling between centres and any batch effects.

Furthermore, two multivariable-adjusted models were fitted; the first was further adjusted for age ($< 55; 55-59; 60-64; 65-69; \geq 70$ years) and body mass index (BMI; fourths of the distribution, i.e. $< 24.6; 24.6-26.5; 26.6-28.9; > 28.9$; unknown), and the second was yet further adjusted for fasting status (no; yes (> 6 hours since last meal); unknown) and time of day at blood collection (6:00 – 8:59; 9:00 – 11:59; 12:00 – 14:59; 15:00 – 20:00; unknown). Results of the two multivariable-adjusted models did not materially differ, so only results from the former, including study centre, metabolite batch, age and BMI, are shown.

9.3.4.4 Correlation analyses

Partial correlations between concentrations of amino acids and IGF-I were computed.

9.3.4.5 IGF-I by fifths of amino acid concentration

Adjusted geometric mean IGF-I concentrations were calculated by fifths of amino acid concentrations and ratios. Multiple linear regression models of IGF-I and amino acid concentrations and ratios were fitted to test for linear trend and to calculate the percentage increase in IGF-I concentration for an 80 percentile increment in amino acid concentrations and ratios (details are found in Section 8.3.4.5 on page 275).

9.3.4.6 Statistical significance

All tests for statistical significance were two-sided. The Benjamini–Hochberg false discovery rate (FDR)-controlling procedure was used to account for multiple testing [154] with $\alpha = 0.05$. This allows 5% of statistically significant findings to be false, on average. For further details on this method, please refer to Section 6.3.4.5 on page 124.

9.3.4.7 Statistical power

With a sample size of 1,142 men, we had more than 80% power to detect a difference in IGF-I concentration for an 80 percentile increase in amino acid concentrations of approximately 8 to 12%.

9.3.4.8 Statistical software

All statistical analyses were conducted in Stata Statistical Software Package, version 14 (Stata Corporation, College Station, Texas, USA), while figures were drawn using R: A language and environment for statistical computing, version 3.3.3 (Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/). The power calculation was done in Quanto, version 1.2.4.

9.4 Results

9.4.1 Correction of IGF-I concentration

The mean concentration of measured IGF-I varied by follow-up phase, most likely due to the fact that assays were done in two different laboratories (as described in Section 9.3.3.1 on page 321). The geometric mean concentrations of measured IGF-I were 21.3 nmol/L (95%CI: 20.5−22.1 nmol/L) and 19.7 nmol/L (95%CI: 19.3−20.1 nmol/L), in phase I and II, respectively ($p_{diff} = 0.0002$; Table 9.1 on page 330). After correction for study centre and IGF-I assay batch, the geometric means were 20.4 nmol/L (95%CI: 19.7 − 21.2 nmol/L) and 20.4 nmol/L (95%CI: 20.1 − 20.8 nmol/L) in the two phases. All further statistical analyses were computed using corrected IGF-I concentrations.

9.4.2 Participant characteristics

Characteristics varied little between men with different IGF-I concentrations (Table 9.2 on page 331). The mean age at blood collection was 61 years. On average, men had a BMI of 27.0 kg/m^2 and consumed 9.7 MJ/d, 16.6% of which was from protein, including 9.7%, 3.1% and 5.1% of energy from animal, dairy and plant protein respectively. A quarter of participants (24.1%) were current smokers, alcohol intake and physical activity were limited in many participants (45.2% drank less than 10 g/d and 60.0% were inactive or moderately inactive), the vast majority was married or cohabiting (89.5%), and primary education was the highest educational attainment for many men (41.3%). At blood collection, nearly half (45.2%) were fasted and most blood samples were collected in the morning (58.1%).

9.4.3 Concentrations of amino acids and IGF-I

9.4.3.1 Correlation analysis

In the minimally-adjusted model (including study centre and metabolite batch), arginine and the Fischer ratio were positively correlated with IGF-I concentration ($r = 0.07$ for both; $0.01 \leq p \leq 0.02$; Table 9.3 on page 332), while tyrosine, ornithine, and the citrulline to arginine, the

ornithine to arginine and the tyrosine to phenylalanine ratios were inversely correlated with IGF-I concentrations ($-0.09 \leq r \leq -0.07$; $0.002 \leq p \leq 0.02$), all at a conventionally significance level ($p < 0.05$), but not after controlling the FDR.

Additional adjustments for age and BMI did in general not materially change the results. However, the correlation of IGF-I with the citrulline to arginine ratio was slightly weakened and no longer significant, while the opposite was the case for the citrulline to ornithine ratio.

9.4.3.2 Analysis by fifths of amino acid concentration

Minimally-adjusted model: After adjustments for study centre and metabolite assay batch, the geometric mean concentration of corrected IGF-I by fifths of amino acid concentrations and ratios is presented in Table 9.4 on page 333. Table 9.5 on page 335 shows the percentage difference in IGF-I concentration for an 80 percentile increase in amino acid concentrations and ratios.

The IGF-I concentration differed by fifths of the ornithine to arginine and the tyrosine to phenylalanine ratios ($p_{diff} = 0.0004$ for both), and there was evidence of inverse linear associations, after correction for multiple testing. The strongest association was with the ornithine to arginine ratio, for which IGF-I concentration was 14.4% (95%CI: $8.3 - 20.1$; $p_{trend} = 1 \times 10^{-5}$) lower for an 80 percentile increase in this ratio.

While IGF-I differed by fifths of arginine, ornithine and the citrulline to arginine ratio only at a conventional significance level, the linear trends remained statistically significant after controlling the FDR. IGF-I was higher in men with higher arginine concentrations, while lower in men with higher ornithine concentration and citrulline to arginine ratio (Table 9.5 on page 335).

Additionally, at a conventional significance level, IGF-I differed by fifths of valine and methionine concentrations, and for tyrosine and the Fischer ratio, the linear trend was inverse and positive, respectively.

Multivariable-adjusted model: After further adjustments for age and BMI, associations were in general slightly attenuated (Tables 9.6 on page 336 and 9.7 on page 338). However, the linear associations of IGF-I with arginine, ornithine, the ornithine to arginine ratio and the tyrosine to phenylalanine ratio remained statistically significant, in the multivariable-adjusted model after correction for multiple testing (Figures 9.2 on page 339 and 9.3 on page 340, A and B panels). Men with the highest arginine concentration had 9.4% (95%CI: $2.8 - 16.5$; $p_{trend} = 0.005$) higher IGF-I concentration than men with the lowest arginine concentration. The other three associations were all inverse, and the strongest was still with the ornithine to arginine ratio, for which IGF-I was 13.8% (95%CI: $7.8 - 19.4$; $p_{trend} = 2 \times 10^{-5}$) lower per an 80 percentile increase in the ratio.

Additionally, we observed inverse associations with methionine, tyrosine and the citrulline to arginine ratio, at a conventional significance level (Figures 9.2 on page 339 and 9.3 on page 340, C and D panels).

9.5 Discussion

9.5.1 Main findings

In this first observational study of circulating concentrations of amino acids and IGF-I in men, men with higher arginine concentration had higher IGF-I concentration, while a lower IGF-I concentration was observed in men with higher concentration of ornithine, and higher ratios of ornithine to arginine and tyrosine to phenylalanine. In contrast to other studies suggesting positive associations of IGF-I with intakes of essential amino acids [79, 86, 87, 88], there was little evidence of associations with plasma concentrations of essential amino acids, in the current study.

9.5.2 Other studies

To my knowledge, only one previous observational study has been published on circulating concentrations of amino acids and IGF-I. The authors reported an inverse correlation between plasma arginine and IGF-I concentration ($r = -0.28$; $p < 0.0004$), based on data from 229 pre-menopausal women [78]. This association was in the opposite direction to our result. While IGF-I is associated with with age, sex and sex steroid hormones [236, 237], it is not clear if the differences in study populations (e.g. younger to middle-aged women versus middle-aged to older men) and/or chance might help explain the discrepancy in results. The relationship between arginine and IGF-I concentrations, and the potential heterogeneity by sex, requires further investigation.

The previous study also found that the sum of ureagenic amino acids (arginine, citrulline and ornithine) was inversely associated with IGF-I ($r = -0.26$, $p < 0.002$). As no association of IGF-I with citrulline nor ornithine was observed, the association with the sum of ureagenic amino acids is likely to be driven by the association with arginine. In the current study, no association between the sum of ureagenic amino acids and IGF-I was found.

9.5.3 Plausibility of findings

9.5.3.1 Arginine, ornithine and their ratio

In the current study, IGF-I was positively associated with arginine, and inversely with ornithine and the ornithine to arginine ratio. Observing associations with IGF-I for both arginine

and ornithine in opposite directions may not be surprising, because arginine is transformed to ornithine by the enzyme arginase [228].

In line with our results, a review, including mainly small trials in humans, found that arginine increases resting growth hormone concentration, which in turn increases IGF-I [238]. On the other hand, arginine is a non-essential amino acid, while previous studies suggest that it is essential amino acids that might increase IGF-I concentrations [86, 87, 88]. A long list of beneficial effects of arginine supplementation has been suggested, including effects on immune function, blood pressure, insulin sensitivity and lean body mass, but the effect on cancers remains unclear [216]. While arginine deprivation has been used to treat prostate cancer [216, 239] due to potential toxic effects on prostate cancer cells [239], limited evidence of a tumour-enhancing effect of arginine supplementation is available [216].

The inverse association of IGF-I with the ornithine to arginine ratio is in line with the directions of the associations for the individual amino acids; an inverse association with ornithine (the numerator) and a positive association with arginine (the denominator). This ratio reflects the activity of arginase [228], which has been reported to be over-expressed in prostate cancer cells and might help protect prostate tumour cells from the immune response [229].

9.5.3.2 Tyrosine to phenylalanine ratio

Men with a higher tyrosine to phenylalanine ratio had lower IGF-I concentration in the current study. This association might primarily be driven by an inverse association with tyrosine; we observed an inverse association with tyrosine, albeit at the conventional significance level, while no association was found for phenylalanine. The non-essential amino acid tyrosine is synthesised from the essential phenylalanine, by phenylalanine hydroxylase. To my knowledge, little is known about how these amino acids or the enzyme might affect IGF-I concentrations.

9.5.4 The role of amino acid concentrations in the association between dietary protein intake and IGF-I concentration

Taken together, Chapter 8 and current chapter provide some evidence that plasma amino acids might mediate the association between intake of protein and IGF-I concentrations.

Of the four amino acids which remained significantly associated with IGF-I after correcting for multiple testing in this study, only ornithine was also associated with protein intake, albeit at a conventional significance level. Ornithine was positively associated with intake of plant protein (Table 8.18 on page 311), while inversely with IGF-I concentration. If these observed associations are true, it might indicate that men who consume more plant protein have lower IGF-I potentially via higher ornithine concentration (Figure 9.1 on the next page, panel A).

While some studies suggest that plant protein intake might be related to lower IGF-I [81, 79], others have found a positive association [75, 224] or no association [6, 7].

Moreover, the citrulline to arginine ratio was suggested to be inversely related to IGF-I ($p < 0.05$) and was inversely associated with intake of animal protein (Table 8.15 on page 307). This might imply that men with higher intakes of animal protein have higher IGF-I concentrations (Figure 9.1, panel B), which is in line with previously published results [6, 7, 75, 224]. This thesis provides evidence that this association might partly be explained by a lower citrulline to arginine ratio.

Finally, the current results for tyrosine (positively associated with dairy protein intake and suggestively inversely associated with IGF-I) might imply an inverse association between dairy protein intake and IGF-I concentration (Figure 9.1, panel C), which is in contrast with earlier evidence [6, 7, 75, 76].

As the logic presented above is based on the first analyses in men and some of the associations did not remain after correction for multiple testing, more research is needed.

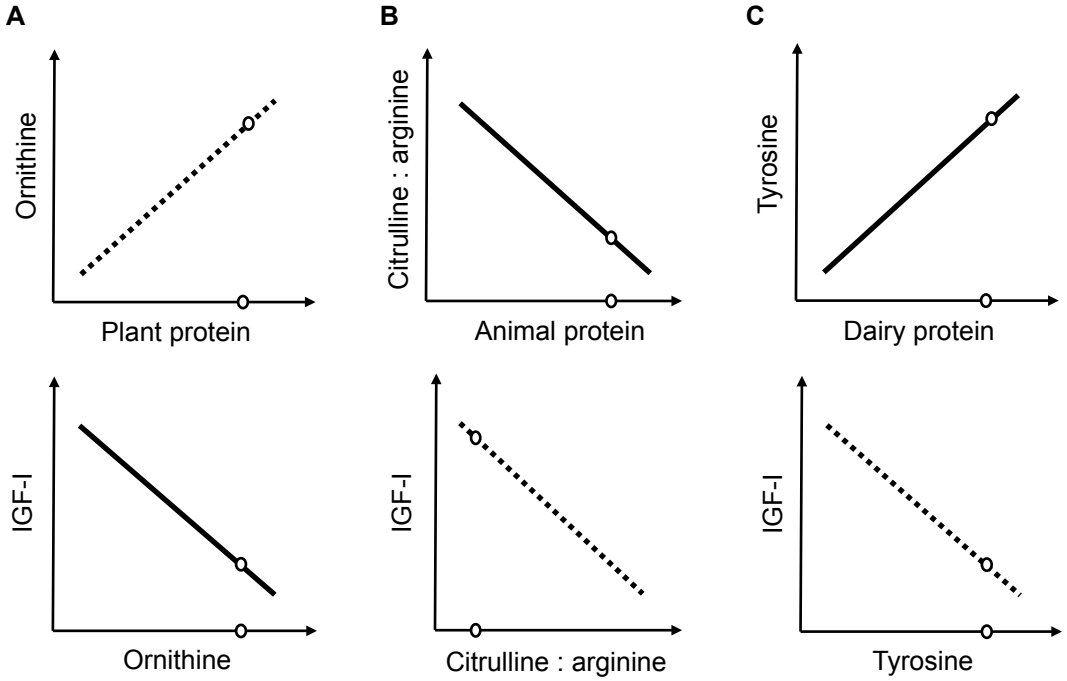


Figure 9.1: Schematic overview of the associations of selected amino acids with protein intake and circulating IGF-I concentration observed in this thesis
A. Ornithine. B. Citrulline to arginine ratio. C. Tyrosine. The full lines represent associations remaining statistically significant after correction for multiple testing using a false discovery rate-controlling procedure (Benjamini–Hochberg) at $\alpha = 0.05$, while the dashed lines represent association at the conventional significance level ($\alpha = 0.05$).

9.5.5 Strengths and limitations

The main strength of the current study is the large sample size comprising men from a well-characterised cohort. This allowed us to report the association between amino acids and IGF-I

concentrations in free-living men and account for a wide range of potential confounding factors. Furthermore, an extensive list of amino acids, sums and ratios were included.

There are also several limitations to the study. Firstly, the cross-sectional design of the study does not allow us to infer that the observed associations between amino acids and IGF-I concentrations are causal, and if so, we cannot determine whether amino acids concentrations affect IGF-I concentration or if the relationship could be in the other direction. Secondly, pre-analytical and analytical factors related to measurements of amino acids and IGF-I might have introduced measurement error into our results. IGF-I was measured in two different laboratories which yielded differences in mean concentrations. To account for this systematic difference and any differences in handling and processing of blood between centres, IGF-I concentrations corrected for IGF-I assay batch and centre were used in all analyses. Pre-analytical and analytical factors, which might have affected the measurement of amino acids have been detailed in Section 8.5.3 on page 282. To control for such factors, all analyses were adjusted for study centre and metabolite assay batch, and results were not materially changed after further adjustments for fasting status and time of day at blood collection.

9.5.6 Further work

Our results require replication given that this is the first study in men and that the association with arginine was in the opposite direction of that reported in women [78]. Furthermore, neither the current nor the previous study [78] provided much evidence of an association between concentrations of IGF-I and essential amino acids, intake of which might be related to IGF-I concentration [79, 86, 87, 88].

The potential mediating role of some amino acids in the association between protein intake and IGF-I also needs further investigations in large studies.

In light of the positive association between arginine and IGF-I reported here, the positive association of a metabolite pattern including high arginine with risk of high grade prostate cancer (Section 7.4.3 on page 218), and the limited of data on arginine supplementation in relation to cancer outcomes [216], more research on the role of arginine in cancer development is warranted. Furthermore, research identifying modifiable determinants of circulating arginine and ornithine concentrations might potentially inform preventive strategies for high IGF-I concentrations and thereby possibly for risk of prostate cancer and other IGF-I-related cancers.

9.6 Conclusion

In summary, men with higher plasma arginine concentration, lower ornithine concentration and lower ratios of ornithine to arginine and tyrosine to phenylalanine might have higher circulating

IGF-I concentrations.

In combination with the previous analysis, this study provides some evidence of a potential mediating role of ornithine in an inverse association between plant protein intake and IGF-I, and of the citrulline to arginine ratio in a positive association between intake of animal protein and IGF-I, but further investigations are required before firm conclusions can be drawn.

9.7 Tables

Table 9.1: Measured and corrected IGF-I concentration by follow-up phase in 1,142 men in EPIC

	Phase I, n = 537		Phase II, n = 605		p_{diff}[1]
	Geometric mean, nmol/L	95% CI	Geometric mean, nmol/L	95% CI	
IGF-I measured	21.3	(20.5 - 22.1)	19.7	(19.3 - 20.1)	0.0002
IGF-I corrected[2]	20.4	(19.7 - 21.2)	20.4	(20.1 - 20.8)	-

[1] t-test

[2] Corrected for for study centre and IGF-I assay batch

Table 9.2: Characteristics of the study population by thirds of corrected IGF-I concentration[1]

Characteristics	Thirds of corrected IGF-I concentration[2,3]		
	Bottom, n=380	Middle, n=381	Top, n=381
Age at blood collection, years	62.0 (61.4 - 62.6)	60.7 (60.2 - 61.3)	59.3 (58.7 - 59.9)
Height[4], cm	171.7 (171.0 - 172.3)	172.3 (171.6 - 172.9)	173.3 (172.7 - 174.0)
Weight[4], cm	80.4 (79.2 - 81.7)	79.9 (78.7 - 81.2)	80.6 (79.4 - 81.9)
BMI[4], kg/m²	27.3 (26.9 - 27.6)	26.9 (26.6 - 27.3)	26.8 (26.5 - 27.2)
Smoking, %			
Never	27.4 (22.9 - 31.9)	31.0 (26.4 - 35.6)	31.8 (27.2 - 36.4)
Former	43.8 (38.9 - 48.7)	49.6 (44.7 - 54.5)	43.9 (39.0 - 48.7)
Current	28.8 (24.4 - 33.2)	19.4 (15.6 - 23.2)	24.3 (20.1 - 28.6)
Alcohol consumption[4], %			
<10 g/day	43.7 (39.1 - 48.3)	47.1 (42.4 - 51.8)	44.7 (40.0 - 49.5)
10-19 g/day	16.8 (13.1 - 20.5)	18.2 (14.4 - 22.0)	23.4 (19.2 - 27.6)
20-40 g/day	20.3 (16.3 - 24.2)	20.3 (16.4 - 24.1)	19.5 (15.6 - 23.3)
≥40 g/day	19.2 (15.6 - 22.8)	14.5 (11.1 - 17.8)	12.4 (9.3 - 15.5)
Physical activity[4], %			
Inactive	30.5 (26.2 - 34.8)	28.9 (24.5 - 33.3)	27.1 (22.7 - 31.4)
Moderately inactive	31.8 (27.3 - 36.3)	31.9 (27.3 - 36.5)	30.0 (25.4 - 34.5)
Moderately active	21.9 (17.8 - 26.0)	19.7 (15.8 - 23.7)	23.1 (18.9 - 27.2)
Active	15.8 (12.2 - 19.3)	19.5 (15.6 - 23.4)	19.9 (15.9 - 23.8)
Marital status[4], %			
Not married or cohabiting	12.2 (8.6 - 15.8)	9.9 (6.6 - 13.3)	9.4 (6.2 - 12.7)
Married or cohabiting	87.8 (84.2 - 91.4)	90.1 (86.7 - 93.4)	90.6 (87.3 - 93.8)
Educational attainment[4], %			
Primary or equivalent	45.4 (40.9 - 49.9)	40.3 (35.7 - 44.9)	38.2 (33.6 - 42.8)
Secondary	35.0 (30.4 - 39.6)	34.3 (29.6 - 38.9)	34.1 (29.4 - 38.7)
Degree	19.6 (15.9 - 23.3)	25.4 (21.2 - 29.6)	27.7 (23.6 - 31.9)
Energy intake, MJ	9.61 (9.35 - 9.87)	9.67 (9.40 - 9.93)	9.71 (9.45 - 9.97)
Total protein[4], % of energy	16.4 (16.1 - 16.6)	16.5 (16.2 - 16.7)	16.9 (16.6 - 17.1)
Animal protein[4], % of energy	9.5 (9.2 - 9.7)	9.6 (9.3 - 9.9)	10.0 (9.7 - 10.3)
Dairy protein[4], % of energy	3.1 (3.0 - 3.3)	3.1 (3.0 - 3.3)	3.1 (2.9 - 3.3)
Plant protein[4], % of energy	5.1 (5.0 - 5.2)	5.0 (4.8 - 5.1)	5.1 (5.0 - 5.3)
Fasting status[4], %			
No	56.8 (52.2 - 61.3)	53.2 (48.7 - 57.8)	54.9 (50.2 - 59.5)
Yes	43.2 (38.7 - 47.8)	46.8 (42.2 - 51.3)	45.1 (40.5 - 49.8)
Time of blood collection[4,5], %			
06:00 - 08:59	12.1	11.2	13.0
09:00 - 11:59	42.3	46.4	49.2
12:00 - 14:59	22.3	19.1	17.4
15:00 - 20:00	23.4	23.2	20.4

[1] Adjusted for centre

[2] IGF-I concentration (corrected for centre and IGF-I assay batch): bottom third: 4.62 − 18.18 nmol/L; middle third: 18.19 − 23.45 nmol/L; top third: 23.45 − 72.74 nmol/L

[3] Numbers are mean (95%CI) for continuous variables and percentages (95%CI) for categorical variables

[4] Unknown values for some participants; the calculations of percentages exclude missing values

[5] No confidence intervals could be calculated due to zero observations in some cells

Table 9.3: Partial correlations of amino acid concentrations and ratios with corrected IGF-I in 1,142 men in EPIC[1]

Amino acids	Minimally-adj. model[2]			Multivariable-adj. model[3]		
	r	p[4]	p_{adj}[4]	r	p[4]	p_{adj}[4]
Branched-chain essential amino acids						
Isoleucine	-0.01	0.6229	0.8982	-0.02	0.4508	0.8116
Leucine	0.02	0.6102	0.8982	0.00	0.8769	0.9573
Valine	0.02	0.4608	0.8982	0.01	0.7196	0.9573
Branched-chain amino acids	0.01	0.6562	0.8982	0.00	0.9346	0.9629
Other essential amino acids						
Histidine	0.04	0.1673	0.5170	0.03	0.3614	0.7679
Lysine	0.05	0.1240	0.4217	0.05	0.1274	0.3442
Methionine	-0.04	0.2313	0.6552	-0.05	0.1091	0.3442
Phenylalanine	0.00	0.9339	0.9771	0.01	0.8666	0.9573
Threonine	-0.02	0.5032	0.8982	-0.03	0.2968	0.6727
Tryptophan	-0.03	0.3302	0.7484	-0.05	0.1280	0.3442
Essential amino acids	0.02	0.6185	0.8982	0.00	0.8739	0.9573
Non-essential amino acids						
Alanine	0.01	0.7995	0.9474	0.02	0.5154	0.8762
Arginine	0.07	0.0202	0.1070	0.06	0.0330	0.1872
Asparagine	0.00	0.8889	0.9771	-0.01	0.6223	0.9573
Aspartate	0.01	0.6446	0.8982	0.02	0.4535	0.8116
Glutamate	0.00	0.9688	0.9771	0.00	0.9010	0.9573
Glutamine	-0.03	0.2842	0.6902	-0.03	0.2835	0.6727
Glycine	-0.01	0.7365	0.9275	-0.01	0.8296	0.9573
Proline	0.00	0.9309	0.9771	0.00	0.8962	0.9573
Serine	-0.03	0.2828	0.6902	-0.05	0.1316	0.3442
Tyrosine	-0.07	0.0220	0.1070	-0.07	0.0300	0.1872
Non-essential amino acids	-0.01	0.6387	0.8982	-0.01	0.7828	0.9573
Non-standard amino acids						
Citrulline	-0.01	0.6604	0.8982	0.01	0.7002	0.9573
Ornithine	-0.08	0.0126	0.1070	-0.07	0.0166	0.1872
t4-OH-Pro	0.03	0.3560	0.7565	0.02	0.4455	0.8116
Non-standard amino acids	-0.06	0.0542	0.2305	-0.05	0.1053	0.3442
Breakdown pathways						
Glucogenic amino acids[5]	-0.01	0.7195	0.9275	-0.01	0.7224	0.9573
Ketogenic amino acids[6]	0.00	0.9771	0.9771	-0.01	0.7999	0.9573
Ureagenic amino acids[7]	-0.01	0.8081	0.9474	0.00	0.9985	0.9985
Amino acid ratios						
Citrulline:arginine	-0.07	0.0175	0.1070	-0.05	0.0944	0.3442
Citrulline:ornithine	0.05	0.1056	0.3991	0.07	0.0283	0.1872
Fischer ratio[8]	0.07	0.0133	0.1070	0.06	0.0470	0.2285
Ornithine:arginine	-0.09	0.0018	0.0607	-0.09	0.0034	0.0616
Tyrosine:phenylalanine	-0.09	0.0043	0.0736	-0.09	0.0036	0.0616

[1] IGF-I concentrations were corrected for study centre and IGF-I assay batch

[2] Adjusted for study centre and metabolite batch

[3] Additionally adjusted for age group (< 55; $55-59$; $60-64$; $65-69$; ≥ 70) and BMI (fourths; unknown)

[4] p-values marked in red were conventionally statistically significant ($p < 0.05$). No tests were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg)

[5] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[6] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[7] Ureagenic: arg + cit + orn

[8] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 9.4: Geometric mean corrected IGF-I concentration[1] by fifths of amino acid concentrations and ratios in 1,142 men in EPIC – minimally-adjusted model[2]

Amino acids	Geometric mean corrected IGF-I concentration by fifths of amino acid concentrations and ratios, nmol/L											
	1		2		3		4		5		p_{diff}[3]	p_{adj}[3]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	21.1	(20.2 - 22.0)	20.3	(19.4 - 21.2)	19.7	(18.9 - 20.6)	20.9	(20.0 - 21.8)	20.2	(19.3 - 21.1)	0.2356	0.5007
Leucine	20.2	(19.4 - 21.2)	20.8	(19.9 - 21.7)	19.8	(18.9 - 20.7)	20.8	(19.9 - 21.8)	20.5	(19.6 - 21.5)	0.5168	0.8349
Valine	19.7	(18.9 - 20.6)	21.2	(20.2 - 22.1)	19.8	(19.0 - 20.7)	21.2	(20.4 - 22.2)	20.2	(19.3 - 21.2)	0.0394	0.2152
Branched-chain aa	20.4	(19.5 - 21.3)	20.2	(19.3 - 21.1)	20.5	(19.6 - 21.4)	20.8	(19.9 - 21.7)	20.2	(19.3 - 21.2)	0.9108	0.9692
Other essential amino acids												
Histidine	20.4	(19.5 - 21.4)	19.7	(18.8 - 20.6)	20.6	(19.7 - 21.5)	20.3	(19.5 - 21.3)	21.1	(20.2 - 22.2)	0.3022	0.5829
Lysine	19.8	(18.9 - 20.7)	20.6	(19.7 - 21.5)	20.7	(19.8 - 21.6)	20.6	(19.6 - 21.5)	20.5	(19.6 - 21.5)	0.6732	0.9168
Methionine	20.8	(19.9 - 21.8)	20.8	(20.0 - 21.8)	20.9	(20.0 - 21.9)	19.2	(18.4 - 20.0)	20.4	(19.5 - 21.4)	0.0362	0.2152
Phenylalanine	20.5	(19.6 - 21.5)	20.9	(20.0 - 21.8)	20.5	(19.6 - 21.4)	19.6	(18.7 - 20.5)	20.7	(19.8 - 21.7)	0.3086	0.5829
Threonine	20.8	(19.9 - 21.8)	20.5	(19.6 - 21.4)	20.2	(19.3 - 21.0)	20.1	(19.2 - 21.1)	20.5	(19.6 - 21.5)	0.8212	0.9692
Tryptophan	20.2	(19.4 - 21.2)	20.7	(19.8 - 21.7)	20.8	(19.9 - 21.7)	20.9	(20.0 - 21.9)	19.5	(18.6 - 20.4)	0.1456	0.3809
Essential aa	20.0	(19.1 - 20.9)	20.6	(19.7 - 21.5)	21.3	(20.4 - 22.2)	19.9	(19.0 - 20.8)	20.4	(19.5 - 21.3)	0.2160	0.4896
Non-essential amino acids												
Alanine	20.1	(19.2 - 21.0)	20.9	(20.0 - 21.8)	20.0	(19.2 - 20.9)	20.8	(19.9 - 21.7)	20.3	(19.3 - 21.2)	0.5738	0.8482
Arginine	19.3	(18.4 - 20.3)	19.9	(19.0 - 20.8)	20.6	(19.7 - 21.5)	21.4	(20.4 - 22.4)	21.1	(20.1 - 22.0)	0.0349	0.2152
Asparagine	20.7	(19.7 - 21.6)	20.1	(19.2 - 21.0)	20.6	(19.7 - 21.5)	20.2	(19.4 - 21.2)	20.6	(19.6 - 21.6)	0.8949	0.9692
Aspartate	20.1	(19.1 - 21.2)	20.6	(19.6 - 21.7)	20.3	(19.4 - 21.3)	20.6	(19.6 - 21.6)	20.6	(19.5 - 21.7)	0.9700	0.9700
Glutamate	20.5	(19.6 - 21.5)	20.6	(19.7 - 21.5)	20.3	(19.4 - 21.2)	20.1	(19.2 - 21.1)	20.6	(19.6 - 21.7)	0.9407	0.9692
Glutamine	20.2	(19.3 - 21.1)	20.8	(19.9 - 21.8)	20.5	(19.6 - 21.4)	20.8	(19.9 - 21.7)	19.8	(19.0 - 20.8)	0.5188	0.8349
Glycine	20.8	(19.9 - 21.8)	19.5	(18.6 - 20.3)	21.1	(20.1 - 22.0)	20.6	(19.7 - 21.5)	20.3	(19.3 - 21.3)	0.1137	0.3716
Proline	20.2	(19.3 - 21.1)	20.3	(19.4 - 21.2)	20.2	(19.3 - 21.1)	21.5	(20.6 - 22.5)	19.9	(19.0 - 20.9)	0.1169	0.3716

Continues...

Table 9.4 continued

Amino acids	Geometric mean corrected IGF-I concentration by fifths of amino acid concentrations and ratios, nmol/L										p_{diff}[3]	p_{adj}[3]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	20.6	(19.7 - 21.5)	20.4	(19.5 - 21.3)	20.2	(19.4 - 21.1)	21.1	(20.2 - 22.1)	19.8	(18.9 - 20.8)	0.3934	0.7041
Tyrosine	21.1	(20.2 - 22.1)	20.6	(19.7 - 21.5)	20.7	(19.8 - 21.6)	20.3	(19.4 - 21.3)	19.4	(18.6 - 20.3)	0.1380	0.3809
Non-essential aa	20.3	(19.4 - 21.2)	20.3	(19.4 - 21.2)	20.6	(19.7 - 21.5)	20.8	(19.9 - 21.7)	20.3	(19.3 - 21.2)	0.9134	0.9692
Non-standard amino acids												
Citrulline	20.6	(19.7 - 21.6)	20.2	(19.3 - 21.1)	20.9	(20.0 - 21.8)	20.1	(19.3 - 21.0)	20.3	(19.4 - 21.3)	0.7598	0.9692
Ornithine	21.3	(20.4 - 22.4)	21.2	(20.3 - 22.2)	20.3	(19.4 - 21.2)	19.5	(18.7 - 20.4)	19.8	(18.8 - 20.8)	0.0443	0.2152
t4-OH-Pro	19.7	(18.8 - 20.6)	20.1	(19.3 - 21.0)	21.1	(20.2 - 22.1)	21.0	(20.1 - 22.0)	20.2	(19.3 - 21.1)	0.1202	0.3716
Non-standard aa	20.6	(19.7 - 21.7)	21.5	(20.6 - 22.5)	20.0	(19.1 - 20.9)	20.1	(19.3 - 21.1)	19.9	(18.9 - 20.9)	0.1227	0.3793
Breakdown pathways												
Glucogenic aa[4]	20.1	(19.2 - 21.0)	20.4	(19.6 - 21.4)	20.6	(19.7 - 21.5)	20.8	(19.9 - 21.8)	20.2	(19.3 - 21.1)	0.8112	0.9692
Ketogenic aa[5]	20.2	(19.3 - 21.1)	20.6	(19.8 - 21.6)	20.9	(20.0 - 21.8)	20.4	(19.5 - 21.3)	20.0	(19.2 - 20.9)	0.6741	0.9168
Ureagenic aa[6]	20.5	(19.6 - 21.5)	20.4	(19.5 - 21.4)	20.7	(19.8 - 21.7)	20.3	(19.5 - 21.3)	20.1	(19.2 - 21.1)	0.9225	0.9692
Amino acid ratios												
Citrulline:arginine	21.1	(20.2 - 22.1)	21.7	(20.7 - 22.6)	20.3	(19.4 - 21.2)	19.9	(19.0 - 20.8)	19.2	(18.3 - 20.2)	0.0122	0.1388
Citrulline:ornithine	20.2	(19.2 - 21.2)	20.3	(19.4 - 21.2)	19.9	(19.1 - 20.8)	20.8	(19.9 - 21.7)	21.0	(20.0 - 22.0)	0.5402	0.8349
Fischer ratio[7]	19.7	(18.8 - 20.6)	20.1	(19.2 - 21.0)	20.6	(19.7 - 21.5)	20.5	(19.7 - 21.5)	21.3	(20.3 - 22.3)	0.1996	0.4847
Ornithine:arginine	21.9	(20.9 - 23.0)	21.4	(20.5 - 22.4)	20.4	(19.5 - 21.3)	19.9	(19.0 - 20.8)	18.7	(17.7 - 19.7)	**0.0004**	0.0073
Tyrosine:phenylala.	21.0	(20.1 - 22.0)	21.0	(20.1 - 22.0)	21.4	(20.5 - 22.4)	19.0	(18.2 - 19.8)	19.7	(18.9 - 20.6)	**0.0004**	0.0073

[1] IGF-I concentrations were corrected for study centre and IGF-I assay batch

[2] Adjusted for study centre and metabolite batch

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 9.5: Percentage difference in corrected IGF-I concentration[1] for an 80 percentile increase in amino acid concentrations and ratios in 1,142 men in EPIC – minimally-adjusted[2]

Amino acids	Difference in corrected IGF-I per 80 percentile increment in amino acids[3], %		p_{trend}[4]	p_{adj}[4]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	-2.49	(-8.03 - 3.39)	0.3987	0.8935
Leucine	1.14	(-4.56 - 7.18)	0.7022	0.8935
Valine	2.19	(-3.70 - 8.43)	0.4743	0.8935
Branched-chain amino acids	0.51	(-5.16 - 6.51)	0.8644	0.9480
Other essential amino acids				
Histidine	4.13	(-2.07 - 10.73)	0.1962	0.5779
Lysine	2.98	(-3.21 - 9.57)	0.3527	0.8935
Methionine	-4.85	(-10.45 - 1.10)	0.1081	0.4084
Phenylalanine	-1.91	(-7.72 - 4.27)	0.5362	0.8935
Threonine	-1.91	(-7.70 - 4.25)	0.5352	0.8935
Tryptophan	-2.62	(-8.21 - 3.32)	0.3792	0.8935
Essential amino acids	-0.05	(-5.84 - 6.10)	0.9877	0.9877
Non-essential amino acids				
Alanine	0.53	(-5.33 - 6.75)	0.8627	0.9480
Arginine	9.99	(3.24 - 17.19)	**0.0032**	0.0276
Asparagine	-0.15	(-6.15 - 6.24)	0.9628	0.9877
Aspartate	1.74	(-5.80 - 9.87)	0.6609	0.8935
Glutamate	-0.93	(-7.48 - 6.09)	0.7897	0.9258
Glutamine	-1.59	(-7.18 - 4.34)	0.5917	0.8935
Glycine	0.32	(-5.63 - 6.63)	0.9193	0.9767
Proline	1.51	(-4.33 - 7.70)	0.6201	0.8935
Serine	-1.58	(-7.33 - 4.53)	0.6036	0.8935
Tyrosine	-6.91	(-12.18 - -1.33)	0.0161	0.0886
Non-essential amino acids	1.03	(-4.90 - 7.33)	0.7403	0.8989
Non-standard amino acids				
Citrulline	-1.15	(-6.97 - 5.04)	0.7096	0.8935
Ornithine	-9.23	(-15.16 - -2.89)	**0.0050**	0.0341
t4-OH-Pro	3.89	(-1.90 - 10.01)	0.1918	0.5779
Non-standard amino acids	-5.60	(-11.70 - 0.91)	0.0902	0.3834
Breakdown pathways				
Glucogenic amino acids[5]	1.24	(-4.71 - 7.55)	0.6906	0.8935
Ketogenic amino acids[6]	-1.19	(-6.93 - 4.91)	0.6960	0.8935
Ureagenic amino acids[7]	-1.70	(-7.53 - 4.49)	0.5808	0.8935
Amino acid ratios				
Citrulline:arginine	-10.17	(-15.96 - -3.98)	**0.0016**	0.0227
Citrulline:ornithine	4.18	(-2.20 - 10.98)	0.2040	0.5779
Fischer ratio[8]	7.31	(1.21 - 13.79)	0.0182	0.0886
Ornithine:arginine	-14.38	(-20.05 - -8.31)	**1.E-05**	0.0003
Tyrosine:phenylalanine	-8.87	(-14.07 - -3.35)	**0.0020**	0.0227

[1] IGF-I concentrations were corrected for study centre and IGF-I assay batch

[2] Analyses were adjusted for study centre and metabolite assay batch

[3] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[4] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[5] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[6] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[7] Ureagenic: arg + cit + orn

[8] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 9.6: Geometric mean corrected IGF-I concentration[1] by fifths of amino acid concentrations and ratios in 1,142 men in EPIC – multivariable-adjusted model[2]

Amino acids	Geometric mean corrected IGF-I concentration by fifths of amino acid concentrations and ratios, nmol/L											
	1		2		3		4		5		p_{diff}[3]	p_{adj}[3]
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Branched-chain essential amino acids												
Isoleucine	21.1	(20.2 - 22.1)	20.3	(19.5 - 21.2)	19.8	(19.0 - 20.7)	20.8	(19.9 - 21.7)	20.1	(19.2 - 21.0)	0.2414	0.5697
Leucine	20.3	(19.5 - 21.2)	20.8	(19.9 - 21.7)	19.9	(19.1 - 20.8)	20.7	(19.8 - 21.6)	20.5	(19.6 - 21.4)	0.6811	0.9378
Valine	19.9	(19.0 - 20.8)	21.2	(20.3 - 22.1)	19.7	(18.9 - 20.6)	21.1	(20.2 - 22.0)	20.2	(19.3 - 21.1)	0.0550	0.3119
Branched-chain aa	20.5	(19.6 - 21.5)	20.2	(19.4 - 21.1)	20.5	(19.7 - 21.4)	20.6	(19.7 - 21.5)	20.2	(19.3 - 21.1)	0.9455	0.9741
Other essential amino acids												
Histidine	20.6	(19.7 - 21.5)	19.7	(18.9 - 20.6)	20.7	(19.9 - 21.7)	20.3	(19.5 - 21.2)	20.8	(19.9 - 21.8)	0.4018	0.6505
Lysine	19.8	(19.0 - 20.8)	20.7	(19.8 - 21.6)	20.6	(19.7 - 21.5)	20.4	(19.5 - 21.3)	20.6	(19.7 - 21.6)	0.7016	0.9378
Methionine	21.1	(20.1 - 22.0)	20.8	(19.9 - 21.7)	20.8	(19.9 - 21.7)	19.2	(18.4 - 20.0)	20.4	(19.5 - 21.3)	0.0235	0.2659
Phenylalanine	20.5	(19.6 - 21.4)	20.8	(19.9 - 21.7)	20.6	(19.7 - 21.5)	19.6	(18.7 - 20.4)	20.7	(19.8 - 21.7)	0.2827	0.5697
Threonine	21.0	(20.1 - 22.0)	20.5	(19.6 - 21.4)	20.1	(19.3 - 21.0)	20.1	(19.2 - 21.0)	20.4	(19.5 - 21.4)	0.5987	0.8850
Tryptophan	20.5	(19.6 - 21.4)	20.8	(19.9 - 21.7)	20.6	(19.8 - 21.5)	20.8	(19.9 - 21.7)	19.4	(18.6 - 20.3)	0.1930	0.5309
Essential aa	20.2	(19.3 - 21.1)	20.6	(19.7 - 21.5)	21.2	(20.3 - 22.2)	19.9	(19.1 - 20.8)	20.2	(19.3 - 21.1)	0.2964	0.5697
Non-essential amino acids												
Alanine	19.9	(19.0 - 20.8)	21.1	(20.2 - 22.0)	20.1	(19.2 - 20.9)	20.8	(19.9 - 21.7)	20.3	(19.4 - 21.2)	0.3105	0.5697
Arginine	19.5	(18.5 - 20.4)	19.8	(18.9 - 20.6)	20.7	(19.8 - 21.6)	21.3	(20.4 - 22.3)	21.0	(20.1 - 21.9)	0.0376	0.3119
Asparagine	20.8	(19.9 - 21.8)	20.2	(19.3 - 21.1)	20.5	(19.6 - 21.4)	20.1	(19.2 - 21.0)	20.5	(19.6 - 21.5)	0.8027	0.9718
Aspartate	20.1	(19.1 - 21.1)	20.7	(19.6 - 21.7)	20.3	(19.4 - 21.2)	20.6	(19.7 - 21.5)	20.7	(19.6 - 21.8)	0.9196	0.9741
Glutamate	20.6	(19.7 - 21.7)	20.5	(19.6 - 21.4)	20.0	(19.2 - 20.9)	20.2	(19.3 - 21.1)	20.7	(19.7 - 21.8)	0.8331	0.9718
Glutamine	20.3	(19.4 - 21.2)	20.9	(20.0 - 21.8)	20.4	(19.5 - 21.3)	20.7	(19.9 - 21.7)	19.8	(18.9 - 20.7)	0.4524	0.6991
Glycine	21.0	(20.1 - 22.0)	19.4	(18.6 - 20.2)	20.9	(20.0 - 21.8)	20.6	(19.7 - 21.5)	20.3	(19.4 - 21.3)	0.0753	0.3511
Proline	20.0	(19.2 - 20.9)	20.5	(19.6 - 21.4)	20.1	(19.3 - 21.0)	21.5	(20.6 - 22.5)	20.0	(19.1 - 20.9)	0.1002	0.3784

Continues...

Table 9.6 continued

Amino acids	Geometric mean corrected IGF-I concentration by fifths of amino acid concentrations and ratios, nmol/L										p_{diff}[3]	p_{adj}[3]
	1		2		3		4		5			
	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)	Mean	(95% CI)		
Serine	20.7	(19.8 - 21.7)	20.3	(19.5 - 21.2)	20.3	(19.5 - 21.2)	21.0	(20.1 - 22.0)	19.7	(18.8 - 20.6)	0.3043	0.5697
Tyrosine	21.0	(20.1 - 22.0)	20.7	(19.8 - 21.6)	20.7	(19.8 - 21.6)	20.3	(19.4 - 21.2)	19.5	(18.6 - 20.3)	0.1624	0.5021
Non-essential aa	20.2	(19.3 - 21.1)	20.5	(19.6 - 21.4)	20.5	(19.6 - 21.4)	20.6	(19.7 - 21.5)	20.3	(19.4 - 21.3)	0.9781	0.9781
Non-standard amino acids												
Citrulline	20.3	(19.4 - 21.2)	20.2	(19.3 - 21.0)	20.9	(20.0 - 21.8)	20.3	(19.4 - 21.2)	20.6	(19.7 - 21.5)	0.8177	0.9718
Ornithine	21.3	(20.3 - 22.3)	21.2	(20.3 - 22.2)	20.3	(19.5 - 21.3)	19.5	(18.7 - 20.4)	19.8	(18.9 - 20.8)	0.0466	0.3119
t4-OH-Pro	19.8	(18.9 - 20.6)	20.2	(19.4 - 21.1)	21.0	(20.1 - 21.9)	21.0	(20.1 - 21.9)	20.2	(19.3 - 21.1)	0.2030	0.5309
Non-standard aa	20.4	(19.5 - 21.4)	21.5	(20.6 - 22.4)	20.1	(19.2 - 21.0)	20.2	(19.3 - 21.1)	20.0	(19.0 - 21.0)	0.1716	0.5303
Breakdown pathways												
Glucogenic aa[4]	20.2	(19.3 - 21.1)	20.4	(19.5 - 21.3)	20.7	(19.8 - 21.6)	20.6	(19.7 - 21.5)	20.3	(19.4 - 21.2)	0.9378	0.9741
Ketogenic aa[5]	20.3	(19.4 - 21.3)	20.7	(19.8 - 21.6)	20.7	(19.8 - 21.6)	20.5	(19.6 - 21.4)	19.9	(19.1 - 20.8)	0.7171	0.9378
Ureagenic aa[6]	20.3	(19.4 - 21.2)	20.6	(19.7 - 21.5)	20.8	(19.9 - 21.7)	20.4	(19.6 - 21.3)	20.1	(19.2 - 21.0)	0.8574	0.9718
Amino acid ratios												
Citrulline:arginine	20.9	(19.9 - 21.9)	21.4	(20.5 - 22.4)	20.4	(19.5 - 21.3)	20.1	(19.2 - 21.0)	19.4	(18.5 - 20.4)	0.0826	0.3511
Citrulline:ornithine	20.0	(19.1 - 21.0)	20.2	(19.3 - 21.1)	20.0	(19.2 - 20.9)	20.9	(20.0 - 21.8)	21.1	(20.2 - 22.1)	0.3575	0.6077
Fischer ratio[7]	19.9	(19.0 - 20.8)	20.1	(19.2 - 21.0)	20.6	(19.8 - 21.5)	20.4	(19.5 - 21.3)	21.2	(20.3 - 22.2)	0.3184	0.5697
Ornithine:arginine	21.8	(20.8 - 22.8)	21.4	(20.5 - 22.4)	20.4	(19.5 - 21.3)	19.9	(19.1 - 20.8)	18.7	(17.8 - 19.7)	**0.0006**	0.0107
Tyrosine:phenylala.	21.0	(20.1 - 21.9)	21.0	(20.1 - 22.0)	21.5	(20.6 - 22.5)	19.0	(18.2 - 19.8)	19.7	(18.8 - 20.6)	**0.0002**	0.0062

[1] IGF-I concentrations were corrected for study centre and IGF-I assay batch

[2] Analyses were adjusted for study centre, metabolite assay batch, age group (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70) and BMI (fourths; unknown)

[3] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at α =0.05 (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[4] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[5] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[6] Ureagenic: arg + cit + orn

[7] Fischer ratio: Branched-chain/(phe + trp + tyr)

Table 9.7: Percentage difference in corrected IGF-I concentration[1] for an 80 percentile increment in amino acid concentrations and ratios in 1,142 men in EPIC – multivariable-adjusted model[2]

Amino acids	Difference in corrected IGF-I per 80 percentile increment in amino acids[3], %		p_{trend}[4]	p_{adj}[4]
	Estimate	(95% CI)		
Branched-chain essential amino acids				
Isoleucine	-3.20	(-8.64 - 2.56)	0.2701	0.7238
Leucine	0.21	(-5.40 - 6.15)	0.9432	0.9614
Valine	0.79	(-5.02 - 6.95)	0.7952	0.8989
Branched-chain amino acids	-0.57	(-6.14 - 5.33)	0.8460	0.8989
Other essential amino acids				
Histidine	2.25	(-3.75 - 8.63)	0.4700	0.8411
Lysine	2.57	(-3.50 - 9.02)	0.4153	0.8411
Methionine	-5.95	(-11.39 - -0.17)	0.0438	0.2126
Phenylalanine	-1.62	(-7.37 - 4.47)	0.5931	0.8989
Threonine	-3.06	(-8.70 - 2.94)	0.3105	0.7542
Tryptophan	-4.03	(-9.47 - 1.74)	0.1674	0.5175
Essential amino acids	-1.27	(-6.94 - 4.73)	0.6702	0.8989
Non-essential amino acids				
Alanine	1.10	(-4.72 - 7.27)	0.7179	0.8989
Arginine	9.44	(2.84 - 16.47)	**0.0045**	0.0496
Asparagine	-1.40	(-7.27 - 4.85)	0.6536	0.8989
Aspartate	2.26	(-5.21 - 10.32)	0.5634	0.8989
Glutamate	-0.69	(-7.31 - 6.41)	0.8445	0.8989
Glutamine	-2.37	(-7.94 - 3.55)	0.4249	0.8411
Glycine	0.15	(-5.85 - 6.53)	0.9614	0.9614
Proline	1.83	(-3.93 - 7.93)	0.5422	0.8989
Serine	-2.59	(-8.25 - 3.43)	0.3913	0.8411
Tyrosine	-6.74	(-12.01 - -1.15)	0.0188	0.1065
Non-essential amino acids	0.90	(-4.92 - 7.09)	0.7669	0.8989
Non-standard amino acids				
Citrulline	1.37	(-4.57 - 7.68)	0.6579	0.8989
Ornithine	-8.91	(-14.75 - -2.67)	**0.0058**	0.0496
t4-OH-Pro	3.17	(-2.48 - 9.15)	0.2767	0.7238
Non-standard amino acids	-4.41	(-10.48 - 2.07)	0.1772	0.5476
Breakdown pathways				
Glucogenic amino acids[5]	0.82	(-5.00 - 6.99)	0.7881	0.8989
Ketogenic amino acids[6]	-2.21	(-7.82 - 3.75)	0.4591	0.8411
Ureagenic amino acids[7]	-1.03	(-6.80 - 5.10)	0.7355	0.8989
Amino acid ratios				
Citrulline:arginine	-7.81	(-13.68 - -1.54)	0.0155	0.1051
Citrulline:ornithine	5.95	(-0.46 - 12.77)	0.0693	0.2619
Fischer ratio[8]	5.93	(-0.03 - 12.24)	0.0511	0.2170
Ornithine:arginine	-13.76	(-19.38 - -7.75)	**2.E-05**	0.0006
Tyrosine:phenylalanine	-8.93	(-14.08 - -3.48)	**0.0016**	0.0279

[1] IGF-I concentrations were corrected for study centre and IGF-I assay batch

[2] Adjusted for centre, metabolite assay batch, age group and BMI

[3] Calculated as $(e^{\beta}-1) \times 100$; β is the regression coefficient from the linear regression

[4] p-values marked in red were conventionally statistically significant ($p < 0.05$). p-values marked in bold were statistically significant after allowing for multiple testing using a false discovery rate-controlling procedure at $\alpha = 0.05$ (Benjamini–Hochberg) and correspond to $p_{adj} < 0.05$ also marked in red

[5] Glucogenic: ile + val + his + met + phe + thr + trp + ala + arg + asn + asp + glu + gln + gly + pro + ser + tyr

[6] Ketogenic: ile + leu + lys + phe + trp + thr + tyr

[7] Ureagenic: arg + cit + orn

[8] Fischer ratio: Branched-chain/(phe + trp + tyr)

9.8 FIGURES

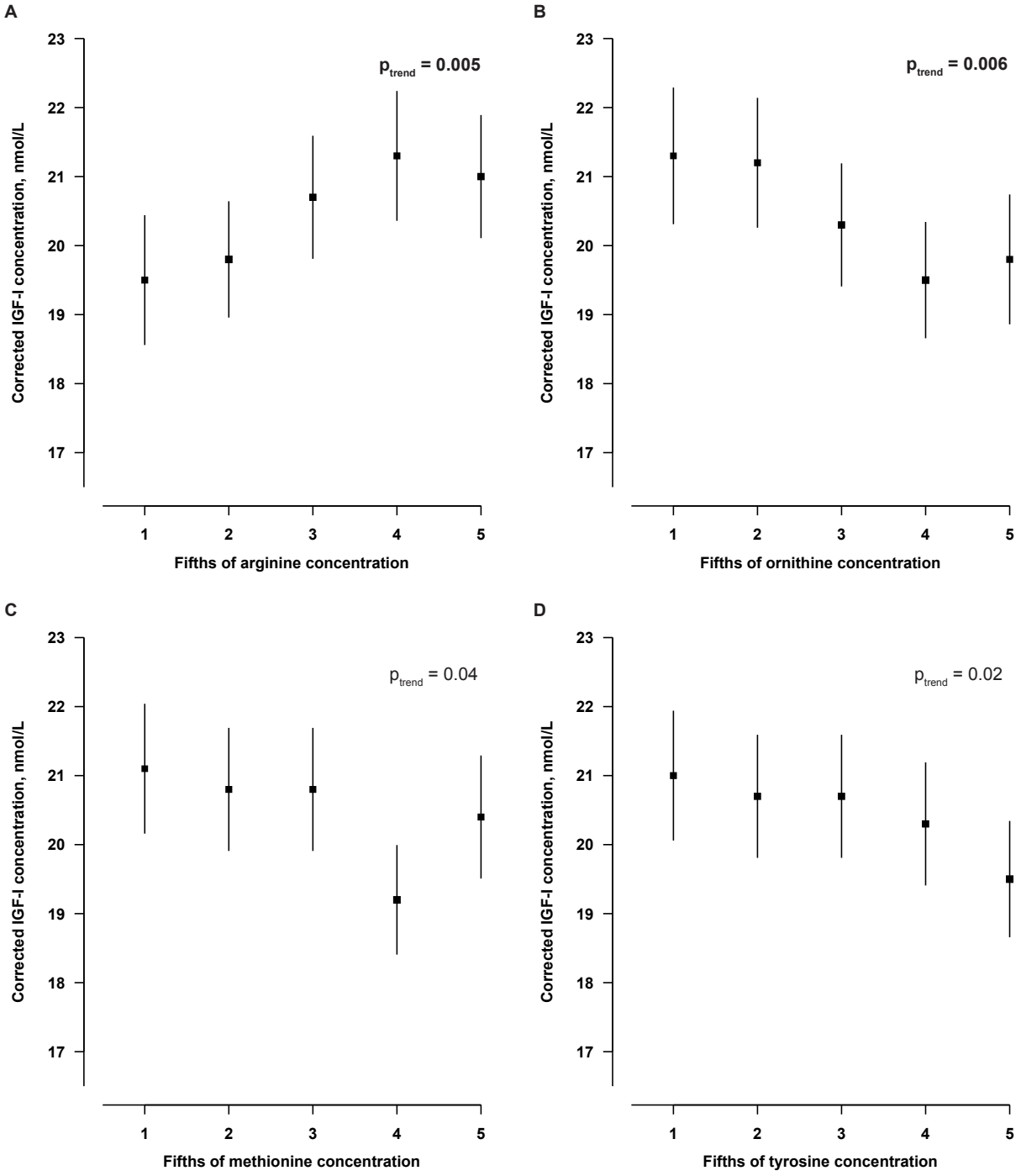


Figure 9.2: Corrected IGF-I concentration by fifths of selected amino acid concentrations in 1,142 men in EPIC

A. Arginine. B. Ornithine. C. Methionine. D. Tyrosine. Amino acids with $p_{trend} < 0.05$ were included in the figure. p_{trend} marked in bold were statistically significant after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). IGF-I was corrected for study centre and IGF-I assay batch, and analyses were adjusted for study centre, metabolite assay batch, age (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years) and BMI (fourths of the distribution, unknown).

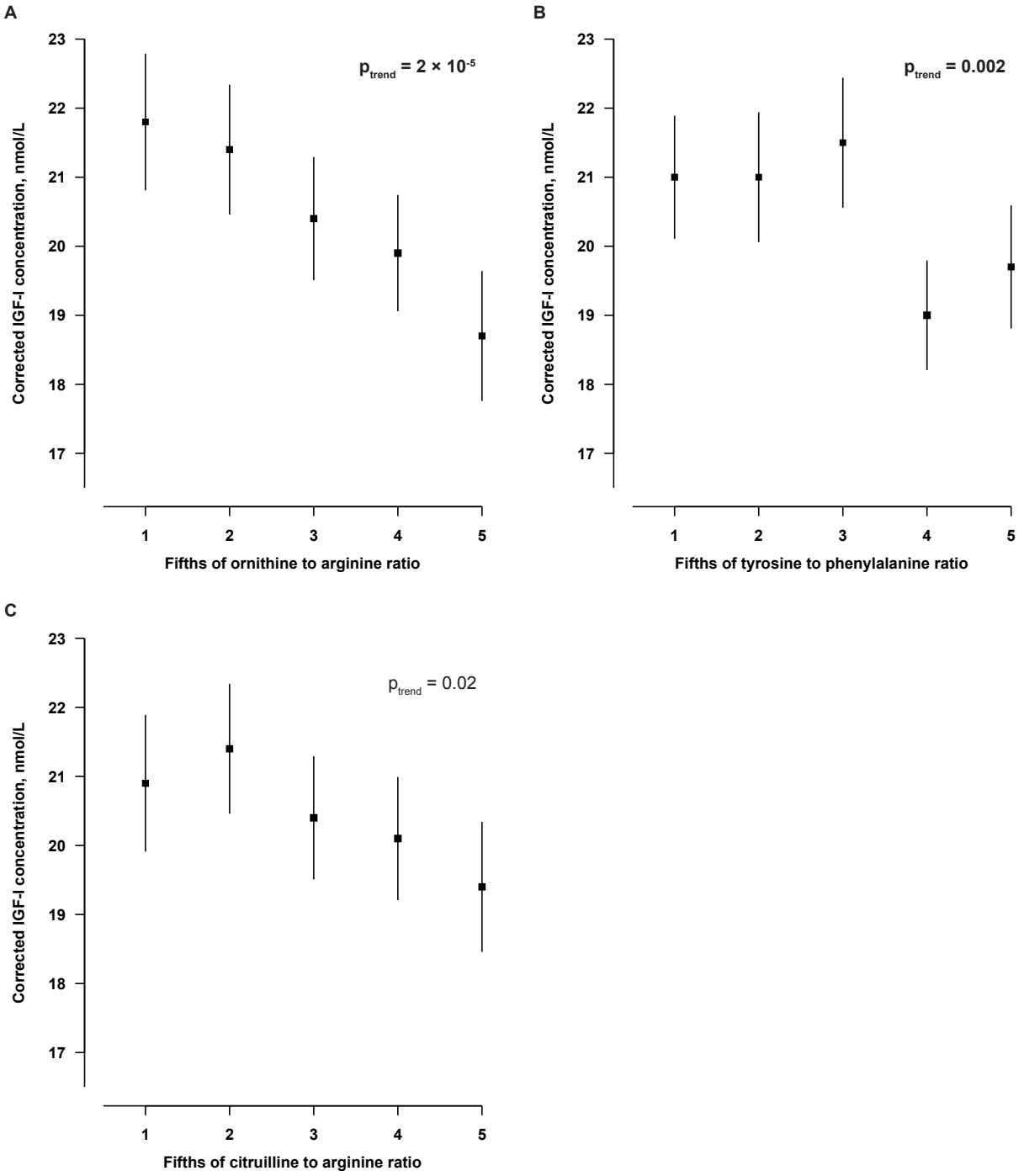


Figure 9.3: Corrected IGF-I concentration by fifths of selected amino acid ratios in 1,142 men in EPIC
A. Ornithine to arginine ratio. B. Tyrosine to phenylalanine ratio. C. Amino acid ratios with $p_{trend} < 0.05$ were included in the figure. Citrulline to arginine ratio. p_{trend} marked in bold were statistically significant after controlling the false discovery rate at $\alpha = 0.05$ (Benjamini–Hochberg). IGF-I was corrected for study centre and IGF-I assay batch, and analyses were adjusted for study centre, metabolite assay batch, age (< 55; $55 - 59$; $60 - 64$; $65 - 69$; ≥ 70 years) and BMI (fourths of the distribution, unknown).

10. Summary and recommendations for future research

10.1 Summary of findings

In this thesis, metabolomics data were analysed to investigate how plasma concentrations of amino acids and other metabolites are related to diet (focusing on protein), circulating insulin-like growth factor-I (IGF-I) and prostate cancer risk, in men from the European Prospective Investigation into Cancer and Nutrition (EPIC).

Cross-sectional analyses including 392 men from EPIC-Oxford showed that amino acid and metabolite profiles were associated with diet group, defined according to intake of animal protein, i.e. meat-eaters, fish-eater, vegetarians and vegans. Concentrations of six amino acids differed across the four diet groups with vegans having lower concentrations of lysine, methionine, tryptophan and tyrosine, and higher concentrations of alanine and glycine, than meat-eaters. Of all 118 metabolites, 79% differed across the four diet groups, with most being lowest in vegans. Especially strong associations were found for glycerophospholipids and sphingolipids, which were also responsible for the separation of vegans from the other diet groups in the analysis of metabolite patterns.

In case-control analyses nested within EPIC of pre-diagnostic metabolite concentrations, including up to 1,593 matched sets, we saw indications of associations between metabolites and overall risk of prostate cancer and stronger associations with risk of high-risk tumour sub-types and prostate cancer death. Risk of overall prostate cancer was positively associated with concentrations of an amino acid trans-4-hydroxyproline (t4-OH-Pro), three glycerophospholipids and a sphingolipid, and was inversely associated with an acylcarnitine and an amino acid (citrulline), but the associations did not remain after correcting for multiple testing. Concentrations of individual glycerophospholipids and sphingolipid, and patterns loading positively on these metabolites were related to lower risk of high grade and especially advanced stage prostate cancer. Analyses of further metabolite patterns showed that acylcarnitines other than C3 (alone or in combination with glutamine, ornithine and taurine) might be inversely associated with advanced prostate cancer, while C3 might be positively associated with death from prostate cancer, and that arginine in combination with hexose might be positively related to high grade prostate cancer. However, analyses by time to diagnosis suggested that some metabolites and patterns might be markers of sub-clinical disease rather than being relevant for aetiology.

Further cross-sectional analyses, in up to 1,697 control participants from the nested case-control study, showed that intake of protein from various sources is related to amino acid concentrations and that certain amino acids are related to circulating IGF-I concentrations. Firstly,

intakes of total and animal protein were positively related to t4-OH-Pro and inversely to glycine. Intake of protein from dairy products was positively associated concentrations of several amino acids, including branched-chain and other essential amino acids. In contrast, intake of plant protein was not strongly associated with plasma amino acids. Secondly, circulating arginine concentration (positively), ornithine, the ratios of ornithine to arginine and tyrosine to phenylalanine (all inversely) were associated with IGF-I. There was little overlap between plasma amino acids associated with protein intake and plasma amino acids associated with IGF-I, respectively, suggesting that amino acid concentrations do not play a major mediating role in the association between protein intake and IGF-I.

10.1.1 Findings across analyses

This thesis has contributed substantially to our knowledge about specific associations of metabolite concentrations with diet, circulating concentrations of IGF-I and prostate cancer risk. However, the implication of the results are relatively limited in terms of informing strategies for prostate cancer prevention because few metabolites were observed in all steps from diet to prostate cancer risk. Nonetheless, intake of total, animal (both positively) and dairy (inversely) protein were associated with t4-OH-Pro, and dairy protein intake was further positively related with proline concentration (Figure 10.1 on the next page). Concentrations of these two amino acids might be related to higher risk of overall, advanced, aggressive and/or death from prostate cancer. Moreover, higher ornithine concentrations were observed in vegans compared to meat-eaters and plant protein intake might be positively associated with ornithine concentration, which in turn was associated inversely with IGF-I concentrations. On its own, ornithine was not associated with risk but it was part of a treelet component which was associated with higher risk of advanced stage prostate cancer. Finally, higher arginine concentration was related to higher IGF-I concentration and, in combination with hexose, to higher risk of high grade prostate cancer. Arginine was however not associated with protein intake or diet group. If the associations with ornithine and arginine are causal, further work that aims to identify modifiable determinants of circulating concentrations of these amino acids might advance our knowledge of modifiable risk factors for prostate cancer.

Vegetarians and especially vegans had lower concentrations of most metabolites, including glycerophospholipids and sphingolipids, than meat-eaters, and most of the associations between metabolites and risk of more aggressive sub-types of prostate cancer were inverse. This might imply that men who do not consume meat might have a higher risk of more aggressive prostate cancer. This may however not be consistent with data from the Adventist Health Study, which reported a lower overall prostate cancer risk in vegetarians than non-vegetarians [240], nor with data from Adventist Health Study-2 (statistically significant) and EPIC-Oxford (statistically

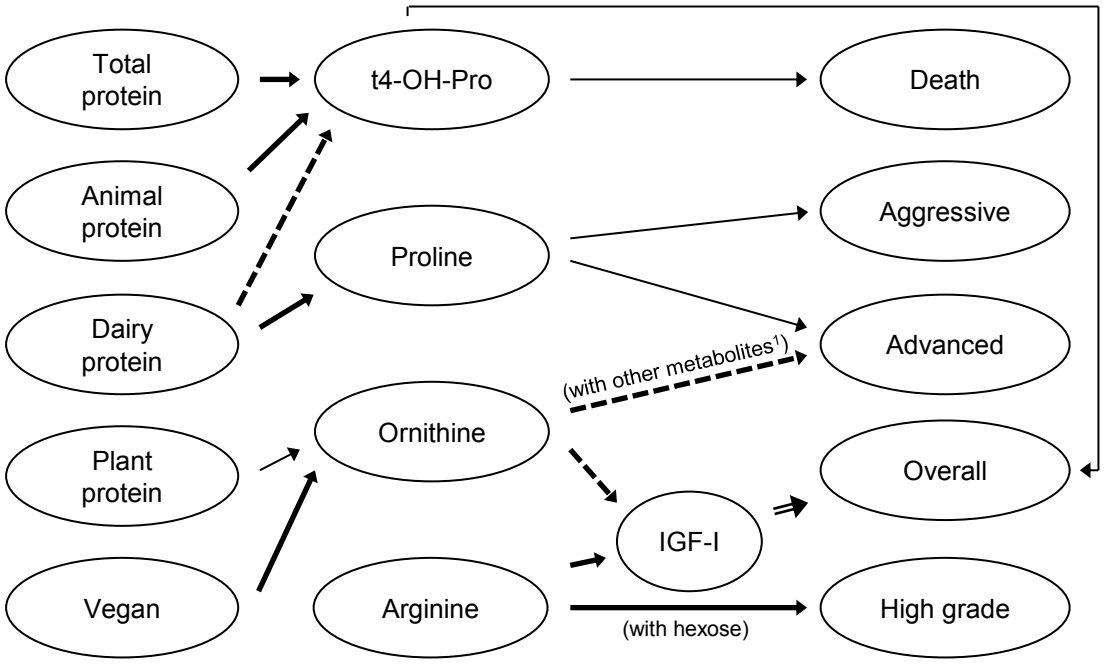


Figure 10.1: Association of selected amino acid concentrations with diet, IGF-I concentration and risk of prostate cancer as observed in this thesis

Thick arrows represent associations remaining statistically significant after correction for multiple testing in univariate analyses or at $p < 0.05$ in multivariate analyses. Thin arrows represent associations at $p < 0.05$ in univariate analyses. Positive and inverse associations are shown using full and dashed arrows, respectively. The double arrow represent an established risk factor for prostate cancer [2], not investigated in the current thesis.

[1] Other metabolites in treelet component 2 besides ornithine were acylcarnitines C16, C18, C18:1, C18:2 and glutamine and taurine.

non-significant) both observing a lower risk of overall prostate cancer in vegans compared to meat-eaters [130, 131]. Although possibly this latter association is due to lower IGF-I concentrations in vegans than in meat-eaters [79, 81]. However, the available evidence on the risk of prostate cancer in vegetarians and vegans is too sparse to draw firm conclusions.

10.1.2 The thesis in context of existing literature

Research into associations between metabolite concentrations and prostate cancer risk is a new and evolving field, and the work presented in this thesis contributes substantially to knowledge in this field. Each of the analyses is one of the very first to investigate their specific topic and all have the largest sample size to date. The large sample sizes in the prospective analyses allowed us to categorise tumour sub-types more finely than previous studies and for the first time to include death from prostate cancer as an outcome.

Although comparison of results from epidemiological studies on metabolomics is not straightforward (owing to differences in biological media, the assay technologies and statistical tools), this thesis and results from the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study [28] suggest that men with higher concentrations of certain lipids might have a lower risk of high-risk prostate tumour sub-types. Such replication of possible associations is essential for interpret-

ation of results. Further work by the research community with regards to harmonising and pooling data (as described in Section 10.3.5 on page 350) may improve comparison and thus interpretation of results.

10.2 Methodological considerations

While the main strengths of the research presented in this thesis are novelty and large size of the study populations from a well-characterised cohort, the work has some methodological limitations making it difficult to rule out the possibility that chance, bias and confounding might have influenced the results to some extent.

10.2.1 The role of chance

The ability to measure a large number of metabolites requires consideration of how to deal with the high dimensionality of the data in the statistical analyses.

To control the excess risk of type I error purely due to chance, we corrected for multiple testing and/or reduced the number of tests using novel and innovative as well as classical methods.

Firstly, to correct for multiple testing, we controlled the family-wise error rate (FWER) using the Bonferroni method in the analysis of amino acids by diet group (Chapter 4). Simple and widely used, this is a very conservative method [140] and true associations might have been missed in Chapter 4. In all other analyses, the false discovery rate (FDR) was controlled using the Benjamini–Hochberg method, which is less conservative than controlling the FWER. There is no consensus in the field about which method to use when correcting for multiple testing; controlling the FDR instead of the FWER was chosen for the latter analyses due to their exploratory nature. However, the FDR-controlling procedure might still be a relatively conservative method, as it does not account for correlations between metabolites. Thus, associations significant at a conventional significance level ($p < 0.05$) have also been reported in this thesis. Secondly, to reduce the number of tests, patterns were identified in men's metabolite profile using the classic principal component analysis and using the more recently developed method of treelet transform.

Although all the statistical analyses included more participants than the previously published papers on the same topics, the statistical power was still limited for analyses of prostate tumour sub-types. Insufficient statistical power makes it difficult to identify significant associations that may truly exist and difficult to rule out the play of chance when results suggest a possible association. Thus modest differences in metabolite concentrations and pattern scores between cases and controls might have been missed and spurious findings cannot be ruled out. More data from EPIC and other cohorts are needed to facilitate more robust analyses, especially of risk of specific tumour sub-types and of prostate cancer death.

10.2.2 The role of bias

10.2.2.1 Selection of participants

The study participants were mainly middle-aged to older men from UK and other European countries. Participants were not representative of the general population but this does not bias the comparisons within the study, e.g. of metabolite concentrations in cases and controls. However, the selection of participants may limit the generalisability of our results to men with roughly similar characteristics to those of our participants. The results on prostate cancer risk might thus, for example, not apply to non-Caucasian men, who may have metabolite profiles [165] and prostate cancer risk [32, 33] different than Caucasian men, and results from the cross-sectional analyses might not be generalisable to women.

10.2.2.2 Measurement of metabolites

Assay platform: The metabolomics assay platform was chosen based on the *a priori* interest in circulating amino acid concentrations and on cost. The Absolute*IDQ*® p180 Kit quantified concentrations of 19 of the 20 standard amino acids and three non-standard amino acids for the vast majority of our study participants. Naturally, results based on a targeted assay are limited to the metabolites included by the platform. While untargeted assays are able to measure a much wider range of metabolites, known as well as unknown, no single platform can capture the entire metabolome. Furthermore, difficulties in annotating and quantifying unknown metabolites are as yet key limitations of studies using untargeted assays [11].

Reproducibility: Only a single blood sample was available for each participant. This might have attenuated the results of the prospective analyses of prostate cancer risk if a single measure does not represent long-term exposure. However, methodological studies have reported relatively high reproducibility (median intra-class correlation coefficient 0.54 to 0.70 over 0.3 to 2.3 years) of metabolites studied in the current thesis and have suggested that a single measurement of metabolite concentrations might be adequate for most metabolites [126, 146, 193], although the reproducibility was lower for some acylcarnitines and a few amino acids [126, 193].

Pre-analytical conditions: Many pre-analytical factors might impact on measurement of metabolite concentrations by introducing extra variation; these include fasting status, time of day at blood collection, the use of anticoagulant in plasma samples, process delay at various temperatures, and freeze-thaw cycles of samples for prior assays. However, the work presented here and elsewhere suggests a limited impact on most metabolite concentrations of fasting status [147, 161, 171, 172], time of day at blood collection [144, 172], process delay and freeze-thaw cycles [144, 145]. Certain amino acids and phosphatidylcholines are among the metabolites per-

forming less well for fasting status [143] and process delay at room temperature [144]. Moreover, use of anticoagulants in plasma samples can affect measurements of metabolites, for citrate especially amino acids, glycerophospholipid and sphingomyelins may be affected [195, 194].

We have limited the impact of the additional variation in metabolite concentrations attributed to pre-analytical factors in the design, assay and analysis phases of the work presented here. Citrate was used in all plasma samples. Fasting status and time of day at blood collection were accounted for by adjusting or matching for these variables, while centre was used as a proxy for sample handling in matching and statistical adjustments. Despite this thorough effort, we cannot rule out some attenuation of our results due to pre-analytical factors.

Analytical considerations: Analytical variation, quantified by the coefficient of variation (CV), might add additional random measurement error to the metabolite data. The mean overall CVs for all measured metabolites ranged from 8 to 11% in the three assays performed for this thesis. The mean CVs for amino acids were in a similar range (9 to 12%). Our mean CVs were slightly higher than the previously reported 7% for amino acids [126, 146], but are in line with or slightly better than the 11% reported across all metabolites measured using a similar assay platform to ours [126]. While reasonably good analytic reproducibility of metabolite concentrations was achieved, some attenuation of study results is likely to have occurred due to analytic variation.

Despite extensive calibration and quality checking in the laboratory, slight differences in metabolite concentrations between batches may have occured. To avoid this potentially impacting our results, samples from the four diet groups were distributed randomly among batches, matched case-control pairs were always assayed in the same batch, and batch was adjusted for in the cross-sectional analyses including control participants only.

10.2.2.3 Measurement of diet

When assessing habitual dietary intake, measurement errors are inevitable. This is likely to lead to an attenuation of most of the results, however it might have inflated specific results.

The associations between intake of protein types and circulating amino acid concentrations (Chapter 8) are likely to be underestimated due to random measurement error for protein intakes. Although the validation studies of the dietary measurement tools used in EPIC suggested a modest to good validity for protein intake, a certain degree of measurement error in intake estimated from the questionnaires is likely. To account for some of this measurement error, nutrient densities (protein intake expressed as percentage of energy) instead of absolute protein intake (g/d) were used and analyses were further adjusted for total energy intake, as food

frequency questionnaires (FFQ) are better at assessing the composition of the diet than absolute amounts [142].

In Chapters 4 and 5, the main dietary exposure variable was not intake as such but diet group, i.e. meat-eaters, fish-eaters, vegetarians and vegans, defined based on four simple yes/no-questions regarding intake of meat, fish, dairy products and egg. The simplicity of the question compared to the full FFQ might have provided a relatively valid categorisation of men into the four diet groups. Thus the analyses of metabolite concentrations by diet group might have been less affected by measurement error in the dietary variable than the analyses based on protein intakes. In contrast, the results on intakes of amino acids by diet group, showing higher intakes of all amino acids in meat-eaters than in the other diet groups, especially vegans (Section 4.4.3.1 on page 40), may have been inflated, as foods replacing animal products in vegan diets might not have been well captured by the FFQ.

10.2.2.4 Follow-up and categorisation of prostate cancer

Loss to follow-up is likely to be limited as cancer registration and vital status are collected in EPIC from registries or through extensive and validated active follow-up. Moreover, any loss to follow-up is unlikely to be related to metabolite concentrations and is thus unlikely to have introduced bias.

Variation in the histological grading and stage classification between pathologists and over time could result in some misclassification of tumours by stage and grade category [196, 197, 198, 199]. Moreover, heterogeneity exists within tumour sub-types [200], although we have categorised these more finely than previous prospective studies [28, 29, 82] (except one small study of sub-types of stage [83]). Both misclassifications and heterogeneity might have led to attenuation of risk estimates in analyses of prostate tumour sub-types. In contrast, high accuracy of death certificates for identification of prostate cancer death has been reported [241, 242].

10.2.2.5 Reverse causality

The time from prostate cancer initiation to diagnosis might take several years [25], potentially leading to reverse causality in our analyses of risk, despite the prospective design. In the analyses of metabolites and overall prostate cancer risk, we stratified by time to diagnosis at five years, and found that some associations with some metabolites might be affected by reverse causality, e.g. for citrulline. Although not relevant for aetiology, such findings might prove interesting with regards to early detection.

10.2.3 The role of confounding

Metabolite concentrations are not only related to diet but also to lifestyle and genetic factors, some of which might be related to dietary habits and/or prostate cancer risk. Therefore, it is essential to consider confounding, and this was facilitated by the extensive collection of data in EPIC on for example diet, lifestyle and factors related to blood handling. Factors observed to be related to metabolite concentrations were age, body mass index, energy intake, alcohol intake, study centre, time since last meal at blood collection, process delay and assay batch, while smoking and physical activity were only associated with a few metabolites (Tables 5.5 on page 76 and 6.5 on page 149). Based on these results and *a priori* considerations of potential confounding, factors to include as covariates in the analyses were selected. For each analysis, several models were considered to assess which variables were included in the final models – often with little difference in results between different models. Additionally, matching of cases and controls was used in the prospective analyses as a means of controlling by design for the effects of key confounding factors, such as age and study centre. In spite of the rigorous approach to confounder control, residual confounding from crudely measured and/or categorised variables, and unknown confounding variables cannot be ruled out.

10.2.4 Summary

While this thesis adds considerably to the current knowledge of plasma metabolite concentrations, diet, IGF-I and prostate cancer risk by providing the largest and some of the first analyses of these associations, it also had some limitations. The use of a single blood sample, factors related to measurement of metabolite concentrations and of dietary intake, and variation in the categorisation of tumour sub-types have most likely attenuated most of our results to some extent. Therefore, true associations might have been missed and associations might be stronger than those observed. Limited power and relatively conservative procedures for dealing with multiple testing might further have contributed to associations being overlooked. Moreover, despite comprehensive confounder control, potential residual confounding cannot be ruled out.

10.3 Future research

10.3.1 Other metabolomics platforms

Using other metabolites platforms, e.g. untargeted or targeted platforms covering other metabolites, may advance our understanding of the metabolome in relation to prostate cancer risk. To date, a key trade-off is between study sample size and therefore statistical power, and the number of metabolites measured at a given cost. In the future, if more comprehensive assays

become available, extending our range of metabolites might prove valuable.

10.3.2 Temporality of the association between metabolites and prostate cancer risk

A key question requiring further investigation is whether the associations of metabolite concentrations and patterns with prostate cancer sub-types, observed in this thesis and in other prospective studies, might be affected by pre-existing tumours. While this has been investigated in some previous analyses [28, 29, 82, 83], more data are needed to provide sufficient statistical power for such analyses to be informative. When data from more prostate cancer cases and controls in EPIC becomes available in due course, we plan to conduct analyses simultaneously stratified by follow-up time and tumour sub-types. A Mendelian randomisation study, as described below (Section 10.3.3), may also provide additional insights.

10.3.3 Genomics

Combining metabolomics data with other '-omics' data, such as genomics data, might advance our understanding of prostate cancer aetiology further.

Mendelian randomisation studies, which use genetic markers of metabolite concentration, are not affected by reverse causation and are less prone to confounding than traditional epidemiological studies because genetic variants are allocated randomly at conception [243]. Such studies have been successful in identifying metabolites that might be causally related to risk of other outcomes. For example, a higher risk of type 2 diabetes in individuals with high concentrations of branched-chain amino acids has been reported [244].

Investigating relationships between metabolite concentrations and common genetic variants associated with prostate cancer risk [35] might also provide some insights into the mechanisms underlying these latter associations.

10.3.4 Tumour tissue

More detailed tumour characterisation than that achieved currently, using information on tumour stage and grade, might decrease the risk of misclassifications and heterogeneity within sub-types that may be aetiologically distinct. Molecular profiling of tumour tissue is a promising tool for categorisation of prostate cancer tumours [200, 245].

Currently, a pilot study assessing the feasibility of retrieving, processing and assaying tumour tissue from EPIC-Oxford prostate cancer cases is taking place. Collected tumour tissue samples are being stained for IGF-I, its receptor and key tumour-related proteins (e.g. p53 and PTEN). Furthermore, DNA and RNA sequencing is being performed focusing on somatic mutations in

key prostate cancer regions. If this work is feasible and scalable, it might prove invaluable in the search for risk factors for high-risk prostate tumour sub-types.

10.3.5 Collaborative work

Some of the above mentioned work requires very large datasets, which will only be feasible if the resources of several cohorts are combined.

The COnsortium of METabolomics Studies (COMETS), formed in 2014, comprises 34 prospective studies (including EPIC) and two consortia [246, 247]. COMETS aims to provide a platform for discussion, collaboration and knowledge exchange in order to improve our understanding of the metabolome in relation to disease, including cancer.

The Endogenous Hormones, Nutritional Biomarkers and Prostate Cancer Collaborative Group (EHNBPCCG), led from the Cancer Epidemiology Unit, University of Oxford, has pooled data from 40 studies to investigate the role of hormones and nutritional biomarkers in the aetiology of prostate cancer [248]. Currently, work is underway to centralise metabolomics data for pooled analyses of prostate cancer risk.

Such consortia should provide important contributions to the field of metabolomics and prostate cancer risk by providing large-scale datasets and facilitating the development and standardisation of methods.

10.4 Conclusion

The work presented in this thesis adds to the novel field of research on metabolites and prostate cancer risk by providing large-scale studies on plasma metabolite concentrations in relation to dietary intake focused on protein intake, circulating concentrations of IGF-I, and risk of prostate cancer overall, by tumour sub-types and risk of prostate cancer death. The main finding was that pre-diagnostic plasma metabolite concentrations and patterns were associated with risk of more aggressive prostate cancer sub-types and prostate cancer death. However, some metabolites might be markers of sub-clinical disease rather than being relevant for aetiology. We also found that men's consumption of protein from various foods is reflected in their plasma concentrations of amino acids and other metabolites, and evidence of associations between concentrations of specific amino acids and circulating IGF-I concentration.

This field of research could be advanced further by collaborative efforts, combining datasets to increase statistical power, by including data generated through the use of other metabolite platforms and '-omics' technologies, and by more detailed characterisation of prostate cancer sub-types.

References

[1] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Available at: `http://globocan.iarc.fr`. Accessed 03 Feb 2014.

[2] Travis RC, Appleby PN, Martin RM, Holly JM, Albanes D, Black A, et al. A Meta-analysis of Individual Participant Data Reveals an Association between Circulating Levels of IGF-I and Prostate Cancer Risk. Cancer Res. 2016;76(8):2288–300.

[3] Pollak M. The insulin and insulin-like growth factor receptor family in neoplasia: an update. Nat Rev Cancer. 2012;12(3):159–69.

[4] Fontana L, Adelaiye RM, Rastelli AL, Miles KM, Ciamporcero E, Longo VD, et al. Dietary protein restriction inhibits tumor growth in human xenograft models of prostate and breast cancer. Oncotarget. 2013;4(12):2451–61.

[5] Liu KA, Lashinger LM, Rasmussen AJ, Hursting SD. Leucine supplementation differentially enhances pancreatic cancer growth in lean and overweight mice. Cancer Metab. 2014;2(1):6.

[6] Crowe FL, Key TJ, Allen NE, Appleby PN, Roddam A, Overvad K, et al. The association between diet and serum concentrations of IGF-I, IGFBP-1, IGFBP-2, and IGFBP-3 in the European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiol Biomarkers Prev. 2009;18(5):1333–40.

[7] Young NJ, Metcalfe C, Gunnell D, Rowlands MA, Lane JA, Gilbert R, et al. A cross-sectional analysis of the association between diet and insulin-like growth factor (IGF)-I, IGF-II, IGF-binding protein (IGFBP)-2, and IGFBP-3 in men in the United Kingdom. Cancer Causes Control. 2012;23(6):907–17.

[8] Allen NE, Key TJ, Appleby PN, Travis RC, Roddam AW, Tjonneland A, et al. Animal foods, protein, calcium and prostate cancer risk: the European Prospective Investigation into Cancer and Nutrition. Br J Cancer. 2008;98(9):1574–81.

[9] World Cancer Research Fund International/American Institute for Cancer Research Continuous Update Project Report. Diet, Nutrition, Physical Activity, and Prostate Cancer; 2014. Available at: `www.wcrf.org/sites/default/files/Prostate-Cancer-2014-Report.pdf`.

[10] Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. Am J Clin Nutr. 2014;99(6):1286–1308.

[11] Kelly RS, Vander Heiden MG, Giovannucci E, Mucci LA. Metabolomic Biomarkers of Prostate Cancer: Prediction, Diagnosis, Progression, Prognosis, and Recurrence. Cancer Epidemiol Biomarkers Prev. 2016;25(6):887–906.

[12] Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant molecular biology. 2002;48:155–171.

[13] Dunn WB, Goodacre R, Broadhurst DI, Atherton HJ, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chemical Society Reviews. 2011;40(1):387–426.

[14] Scalbert A, Brennan L, Fiehn O, Hankemeier T, Kristal BS, van Ommen B, et al. Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. Metabolomics. 2009;5(4):435–58.

[15] Allen NE, Key TJ, Appleby PN, Travis RC, Roddam AW, Rinaldi S, et al. Serum insulin-like growth factor (IGF)-I and IGF-binding protein-3 concentrations and prostate cancer risk: results from the European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiol Biomarkers Prev. 2007;16(6):1121–7.

[16] Price AJ, Allen NE, Appleby PN, Crowe FL, Travis RC, Tipper SJ, et al. Insulin-like growth factor-I concentration and risk of prostate cancer: results from the European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiol Biomarkers Prev. 2012;21(9):1531–41.

[17] Schmidt JA, Rinaldi S, Scalbert A, Ferrari P, Achaintre D, Gunter MJ, et al. Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: a cross-sectional analysis in the EPIC-Oxford cohort. Eur J Clin Nutr. 2016;70(3):306–12.

[18] Schmidt JA, Rinaldi S, Ferrari P, Carayol M, Achaintre D, Scalbert A, et al. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. Am J Clin Nutr. 2015;102(6):1518–26.

[19] Schmidt JA, Fensom GK, Rinaldi S, Scalbert A, Appleby PN, Achaintre D, et al. Pre-diagnostic metabolite concentrations and prostate cancer risk in 1077 cases and 1077 matched controls in the European Prospective Investigation into Cancer and Nutrition. BMC Med. 2017;15(1):122.

[20] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65(2):87–108.

[21] Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, et al. International variation in prostate cancer incidence and mortality rates. Eur Urol. 2012;61(6):1079–1092.

[22] American Cancer Society. Cancer Facts & Figures 2010. American Cancer Society; 2010. Available at: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2010/cancer-facts-and-figures-2010.pdf.

[23] Cuzick J, Thorat MA, Andriole G, Brawley OW, Brown PH, Culig Z, et al. Prevention and early detection of prostate cancer. Lancet Oncol. 2014;15(11):e484–92.

[24] Vickers AJ. Four flawed arguments against prostate-specific antigen screening (and 1 good one). Urology. 2015;85(3):491–4.

[25] Giovannucci E. Commentary: Serum lycopene and prostate cancer progression: a reconsideration of findings from the prostate cancer prevention trial. Cancer Causes Control. 2011;22(7):1055–9.

[26] Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst. 2009;101(6):374–83.

[27] Littlejohns TJ, Travis RC, Key TJ, Allen NE. Lifestyle factors and prostate-specific antigen (PSA) testing in UK Biobank: Implications for epidemiological research. Cancer Epidemiol. 2016;45:40–46.

[28] Mondul AM, Moore SC, Weinstein SJ, Karoly ED, Sampson JN, Albanes D. Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocolpherol, beta-carotene cancer prevention (ATBC) study. Int J Cancer. 2015;137(9):2124–32.

[29] Huang J, Mondul AM, Weinstein SJ, Koutros S, Derkach A, Karoly E, et al. Serum metabolomic profiling of prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer screening trial. Br J Cancer. 2016;115(9):1087–95.

[30] Williams N, Hughes LJ, Turner EL, Donovan JL, Hamdy FC, Neal DE, et al. Prostate-specific antigen testing rates remain low in UK general practice: a cross-sectional study in six English cities. BJU Int. 2011;108(9):1402–8.

[31] Byrne KS, Castano JM, Chirlaque MD, Lilja H, Agudo A, Ardanaz E, et al. Vasectomy and Prostate Cancer Risk in the European Prospective Investigation Into Cancer and Nutrition (EPIC). J Clin Oncol. 2017;35(12):1297–1303.

[32] Rebbeck TR, Haas GP. Temporal trends and racial disparities in global prostate cancer prevalence. Can J Urol. 2014;21(5):7496–506.

[33] Kimura T. East meets West: ethnic differences in prostate cancer epidemiology between East Asians and Caucasians. Chin J Cancer. 2012;31(9):421–9.

[34] Kicinski M, Vangronsveld J, Nawrot TS. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. PLoS One. 2011;6(10):e27130.

[35] Amin Al Olama A, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet. 2014;46(10):1103–1109.

[36] Pietro GD, Chornokur G, Kumar NB, Davis C, Park JY. Racial Differences in the Diagnosis and Treatment of Prostate Cancer. Int Neurourol J. 2016;20(Suppl 2):S112–9.

[37] Eeles R, Goh C, Castro E, Bancroft E, Guy M, Al Olama AA, et al. The genetic epidemiology of prostate cancer and its clinical implications. Nat Rev Urol. 2014;11(1):18–31.

[38] Hoffmann TJ, Van Den Eeden SK, Sakoda LC, Jorgenson E, Habel LA, Graff RE, et al. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. Cancer Discov. 2015;5(8):878–91.

[39] Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013;45(4):385–91.

[40] Amin Al Olama A, Kote-Jarai Z, Schumacher FR, Wiklund F, Berndt SI, Benlloch S, et al. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. Hum Mol Genet. 2013;22(2):408–15.

[41] Bonilla C, Lewis SJ, Rowlands MA, Gaunt TR, Davey Smith G, Gunnell D, et al. Assessing the role of insulin-like growth factors and binding proteins in prostate cancer using Mendelian randomization: Genetic variants as instruments for circulating levels. Int J Cancer. 2016;139(7):1520–33.

[42] Cao Y, Lindstrom S, Schumacher F, Stevens VL, Albanes D, Berndt S, et al. Insulin-like growth factor pathway genetic polymorphisms, circulating IGF1 and IGFBP3, and prostate cancer survival. J Natl Cancer Inst. 2014;106(6):dju085.

[43] Grimberg A. Mechanisms by which IGF-I may promote cancer. Cancer Biol Ther. 2003;2(6):630–5.

[44] Lima GA, Correa LL, Gabrich R, Miranda LC, Gadelha MR. IGF-I, insulin and prostate cancer. Arq Bras Endocrinol Metabol. 2009;53(8):969–75.

[45] Harrison S, Lennon R, Holly J, Higgins JP, Gardner M, Perks C, et al. Does milk intake promote prostate cancer initiation or progression via effects on insulin-like growth factors (IGFs)? A systematic review and meta-analysis. Cancer Causes Control. 2017;28(6):497–528.

[46] Norat T, Vieira AR, Chan D, Aune D, Abar L, Navarro D, et al. World Cancer Research Fund International Systematic Literature Review, The Associations between Food, Nutrition and Physical Activity and the Risk of Prostate Cancer. World Cancer Research Fund, Continuous Update Project; 2014. Available at: `http://www.wcrf.org/sites/default/files/Prostate-Cancer-SLR-2014.pdf`.

[47] Perez-Cornago A, Appleby PN, Pischon T, Tsilidis KK, Tjonneland A, Olsen A, et al. Tall height and obesity are associated with an increased risk of aggressive prostate cancer: results from the EPIC cohort study. BMC Med. 2017;15(1):115.

[48] Davies NM, Gaunt TR, Lewis SJ, Holly J, Donovan JL, Hamdy FC, et al. The effects of height and BMI on prostate cancer incidence and mortality: a Mendelian randomization study in 20,848 cases and 20,214 controls from the PRACTICAL consortium. Cancer Causes Control. 2015;26(11):1603–16.

[49] Khankari NK, Shu XO, Wen W, Kraft P, Lindstrom S, Peters U, et al. Association between Adult Height and Risk of Colorectal, Lung, and Prostate Cancer: Results from Meta-analyses of Prospective Studies and Mendelian Randomization Analyses. PLoS Med. 2016;13(9):e1002118.

[50] Chen Q, Chen T, Shi W, Zhang T, Zhang W, Jin Z, et al. Adult weight gain and risk of prostate cancer: A dose-response meta-analysis of observational studies. Int J Cancer. 2016;138(4):866–74.

[51] Gao C, Patel CJ, Michailidou K, Peters U, Gong J, Schildkraut J, et al. Mendelian randomization study of adiposity-related traits and risk of breast, ovarian, prostate, lung and colorectal cancer. Int J Epidemiol. 2016;45(3):896–908.

[52] Allott EH, Masko EM, Freedland SJ. Obesity and prostate cancer: weighing the evidence. Eur Urol. 2013;63(5):800–9.

[53] Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JP. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ. 2015;350:g7607.

[54] Jian Gang P, Mo L, Lu Y, Runqi L, Xing Z. Diabetes mellitus and the risk of prostate cancer: an update and cumulative meta-analysis. Endocr Res. 2015;40(1):54–61.

[55] Tsilidis KK, Allen NE, Appleby PN, Rohrmann S, Nothlings U, Arriola L, et al. Diabetes mellitus and risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition. Int J Cancer. 2015;136(2):372–81.

[56] Key TJ, Appleby PN, Travis RC, Albanes D, Alberg AJ, Barricarte A, et al. Carotenoids, retinol, tocopherols, and prostate cancer risk: pooled analysis of 15 studies. Am J Clin Nutr. 2015;102(5):1142–57.

[57] Heinonen OP, Albanes D, Virtamo J, Taylor PR, Huttunen JK, Hartman AM, et al. Prostate cancer and supplementation with alpha-tocopherol and beta-carotene: incidence and mortality in a controlled trial. J Natl Cancer Inst. 1998;90(6):440–6.

[58] Virtamo J, Pietinen P, Huttunen JK, Korhonen P, Malila N, Virtanen MJ, et al. Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: a postintervention follow-up. JAMA. 2003;290(4):476–85.

[59] Gaziano JM, Glynn RJ, Christen WG, Kurth T, Belanger C, MacFadyen J, et al. Vitamins E and C in the prevention of prostate and total cancer in men: the Physicians' Health Study II randomized controlled trial. JAMA. 2009;301(1):52–62.

[60] Klein EA, Thompson J I M, Tangen CM, Crowley JJ, Lucia MS, Goodman PJ, et al. Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA. 2011;306(14):1549–56.

[61] Cui Z, Liu D, Liu C, Liu G. Serum selenium levels and prostate cancer risk: A MOOSE-compliant meta-analysis. Medicine. 2017;96(5):e5944.

[62] Allen NE, Travis RC, Appleby PN, Albanes D, Barnett MJ, Black A, et al. Selenium and Prostate Cancer: Analysis of Individual Participant Data From Fifteen Prospective Studies. J Natl Cancer Inst. 2016;108(11).

[63] Algotar AM, Stratton MS, Ahmann FR, Ranger-Moore J, Nagle RB, Thompson PA, et al. Phase 3 clinical trial investigating the effect of selenium supplementation in men at high-risk for prostate cancer. Prostate. 2013;73(3):328–35.

[64] Marshall JR, Tangen CM, Sakr WA, Wood J D P, Berry DL, Klein EA, et al. Phase III trial of selenium to prevent prostate cancer in men with high-grade prostatic intraepithelial neoplasia: SWOG S9917. Cancer Prev Res. 2011;4(11):1761–9.

[65] Duffield-Lillico AJ, Dalkin BL, Reid ME, Turnbull BW, Slate EH, Jacobs ET, et al. Selenium supplementation, baseline plasma selenium status and incidence of prostate cancer: an analysis of the complete treatment period of the Nutritional Prevention of Cancer Trial. BJU Int. 2003;91(7):608–12.

[66] Kristal AR, Darke AK, Morris JS, Tangen CM, Goodman PJ, Thompson IM, et al. Baseline Selenium Status and Effects of Selenium and Vitamin E Supplementation on Prostate Cancer Risk. J Natl Cancer Inst. 2014;106(3):djt456.

[67] Price AJ, Travis RC, Appleby PN, Albanes D, Barricarte Gurrea A, Bjorge T, et al. Circulating Folate and Vitamin B12 and Risk of Prostate Cancer: A Collaborative Analysis of Individual Participant Data from Six Cohorts Including 6875 Cases and 8104 Controls. Eur Urol. 2016;70(6):941–51.

[68] Wang R, Zheng Y, Huang JY, Zhang AQ, Zhou YH, Wang JN. Folate intake, serum folate levels, and prostate cancer risk: a meta-analysis of prospective studies. BMC Public Health. 2014;14:1326.

[69] Tio M, Andrici J, Cox MR, Eslick GD. Folate intake and the risk of prostate cancer: a systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2014;17(3):213–9.

[70] Wang Y, Cui R, Xiao Y, Fang J, Xu Q. Effect of Carotene and Lycopene on the Risk of Prostate Cancer: A Systematic Review and Dose-Response Meta-Analysis of Observational Studies. PLoS One. 2015;10(9):e0137427.

[71] Chen P, Zhang W, Wang X, Zhao K, Negi DS, Zhuo L, et al. Lycopene and Risk of Prostate Cancer: A Systematic Review and Meta-Analysis. Medicine. 2015;94(33):e1260.

[72] Lane JA, Oliver SE, Appleby PN, Lentjes MA, Emmett P, Kuh D, et al. Prostate cancer risk related to foods, food groups, macronutrients and micronutrients derived from the UK Dietary Cohort Consortium food diaries. Eur J Clin Nutr. 2017;71(2):274–83.

[73] Roddam AW, Allen NE, Appleby P, Key TJ. Endogenous sex hormones and prostate cancer: a collaborative analysis of 18 prospective studies. J Natl Cancer Inst. 2008;100(3):170–83.

[74] Aune D, Rosenblatt DAN, Chan DS, Vieira AR, Vieira R, Greenwood DC, et al. Dairy products, calcium, and prostate cancer risk: a systematic review and meta-analysis of cohort studies. Am J Clin Nutr. 2015;101(1):87–117.

[75] Tsilidis KK, Travis RC, Appleby PN, Allen NE, Lindstrom S, Albanes D, et al. Insulin-like growth factor pathway genes and blood concentrations, dietary protein and risk of

prostate cancer in the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). Int J Cancer. 2013;133(2):495–504.

[76] Bradbury KE, Balkwill A, Tipper SJ, Crowe FL, Reeves GK, Green J, et al. The association of plasma IGF-I with dietary, lifestyle, anthropometric, and early life factors in postmenopausal women. Growth Horm IGF Res. 2015;25(2):90–5.

[77] Anderson GH, Blendis LM. Plasma neutral amino acid ratios in normal man and in patients with hepatic encephalopathy: correlations with self-selected protein and energy consumption. Am J Clin Nutr. 1981;34(3):377–85.

[78] Nagata C, Wada K, Tsuji M, Hayashi M, Takeda N, Yasuda K. Plasma amino acid profiles are associated with biomarkers of breast cancer risk in premenopausal Japanese women. Cancer Causes Control. 2014;25(2):143–9.

[79] Allen NE, Appleby PN, Davey GK, Kaaks R, Rinaldi S, Key TJ. The associations of diet with serum insulin-like growth factor I and its main binding proteins in 292 women meat-eaters, vegetarians, and vegans. Cancer Epidemiol Biomarkers Prev. 2002;11(11):1441–8.

[80] Krajcovicova-Kudlackova M, Simoncic R, Bederova A, Babinska K, Beder I. Correlation of carnitine levels to methionine and lysine intake. Physiol Res. 2000;49(3):399–402.

[81] Allen NE, Appleby PN, Davey GK, Key TJ. Hormones and diet: low insulin-like growth factor-I but normal bioavailable androgens in vegan men. Br J Cancer. 2000;83(1):95–7.

[82] Mondul AM, Moore SC, Weinstein SJ, Mannisto S, Sampson JN, Albanes D. 1-stearoylglycerol is associated with risk of prostate cancer: results from serum metabolomic profiling. Metabolomics. 2014;10(5):1036–41.

[83] Huang J, Mondul AM, Weinstein SJ, Karoly ED, Sampson JN, Albanes D. Prospective serum metabolomic profile of prostate cancer by size and extent of primary tumor. Oncotarget. 2017;8(28):45190–9.

[84] Kuhn T, Floegel A, Sookthai D, Johnson T, Rolle-Kampczyk U, Otto W, et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. BMC Med. 2016;14:13.

[85] Lima AR, Bastos Mde L, Carvalho M, Guedes de Pinho P. Biomarker Discovery in Human Prostate Cancer: an Update in Metabolomics Studies. Transl Oncol. 2016;9(4):357–70.

[86] Takenaka A, Oki N, Takahashi SI, Noguchi T. Dietary restriction of single essential amino acids reduces plasma insulin-like growth factor-I (IGF-I) but does not affect plasma IGF-binding protein-1 in rats. J Nutr. 2000;130(12):2910–4.

[87] Clemmons DR, Seek MM, Underwood LE. Supplemental essential amino acids augment the somatomedin-C/insulin-like growth factor I response to refeeding after fasting. Metabolism. 1985;34(4):391–5.

[88] Dillon EL, Sheffield-Moore M, Paddon-Jones D, Gilkison C, Sanford AP, Casperson SL, et al. Amino acid supplementation increases lean body mass, basal muscle protein synthesis, and insulin-like growth factor-I expression in older women. J Clin Endocrinol Metab. 2009;94(5):1630–7.

[89] Melnik BC, John SM, Carrera-Bastos P, Cordain L. The impact of cow's milk-mediated mTORC1-signaling in the initiation and progression of prostate cancer. Nutr Metab. 2012;9(1):74.

[90] Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. Metabolite profiling for plant functional genomics. Nat Biotechnol. 2000;18(11):1157–61.

[91] Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181–9.

[92] HMDB: the Human Metabolome Database. HMDB: the Human Metabolome Database. Available at: `http://www.hmdb.ca/`. Accessed 14 Sep 2017.

[93] Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0 – The Human Metabolome Database in 2013. Nucleic Acids Res. 2013;41(D1):D801–7.

[94] Liesenfeld DB, Habermann N, Owen RW, Scalbert A, Ulrich CM. Review of mass spectrometry-based metabolomics in cancer research. Cancer Epidemiol Biomarkers Prev. 2013;22(12):2182–201.

[95] Wilson ID. Chromatographic and Electrophoretic Separations Combined with Mass Spectrometry for Metabonomics. In: Lindon JC, Nicholson JK, Holmes E, Lindon JC, editors. The Handbook of Metabonomics and Metabolomics. Amsterdam and London: Elsevier; 2007. p. 149–69.

[96] Trock BJ. Application of metabolomics to prostate cancer. Urol Oncol. 2011;29(5):572–81.

[97] Senn H, Schlotterbeck G. NMR Spectroscopy Techniques for Application to Metabonomics. In: Lindon JC, Nicholson JK, Holmes E, Lindon JC, editors. The Handbook of Metabonomics and Metabolomics. Amsterdam and London: Elsevier; 2007. p. 55–112.

[98] Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. Int J Epidemio. 1997;26(Suppl. 1):S6–14.

[99] Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002;5(6B):1113–24.

[100] Davey GK, Spencer EA, Appleby PN, Allen NE, Knox KH, Key TJ. EPIC-Oxford: lifestyle characteristics and nutrient intakes in a cohort of 33 883 meat-eaters and 31 546 non meat-eaters in the UK. Public Health Nutr. 2003;6(3):259–68.

[101] Margetts BM, Pietinen P. European Prospective Investigation into Cancer and Nutrition: validity studies on dietary assessment methods. Int J Epidemiol. 1997;26(Suppl. 1):S1–5.

[102] EPIC group of Spain. Relative validity and reproducibility of a diet history questionnaire in Spain. III. Biochemical markers. Int J Epidemio. 1997;26(Suppl. 1):S110–7.

[103] Katsouyanni K, Rimm EB, Gnardellis C, Trichopoulos D, Polychronopoulos E, Trichopoulou A. Reproducibility and relative validity of an extensive semi-quantitative food frequency questionnaire using dietary records and biochemical markers among Greek schoolteachers. Int J Epidemio. 1997;26(Suppl. 1):S118–27.

[104] Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. Int J Epidemio. 1997;26(Suppl. 1):S152–60.

[105] Ocke MC, Bueno-de Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemio. 1997;26(Suppl. 1):S49–58.

[106] Bohlscheid-Thomas S, Hoting I, Boeing H, Wahrendorf J. Reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. Int J Epidemio. 1997;26(Suppl. 1):S71–S81.

[107] Bingham SA, Cassidy A, Cole TJ, Welch A, Runswick SA, Black AE, et al. Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. Br J Nutr. 1995;73(4):531–50.

[108] Bingham SA, Gill C, Welch A, Day K, Cassidy A, Khaw KT, et al. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. Br J Nutr. 1994;72(4):619–43.

[109] Ministry of Agriculture, Fisheries and Food. Food Portion Sizes. vol. 2. London: Her Majesty's Stationery Office; 1993.

[110] Holland B, Welch AA, Unwin ID, Buss DH, Paul AA, Southgate DAT. McCance and Widdowson's The Composition of Foods. 5th ed. Cambridge: The Royal Society of Chemistry; 1991.

[111] Holland B, Unwin ID, Buss DH. Cereals and cereal products. Third supplement to McCance and Widdowson's The composition of foods, 4th ed. Cambridge: Royal Society of Chemistry; 1988.

[112] Holland B, Unwin ID, Buss DH. Milk products and eggs. Fourth supplement to McCance and Widdowson's The composition of foods, 4th ed. Cambridge: Royal Society of Chemistry; 1989.

[113] Holland B, Unwin ID, Buss DH. Vegetables, herbs and spices. Fifth supplement to McCance and Widdowson's The composition of foods, 4th ed. Cambridge: Royal Society of Chemistry; 1991.

[114] Holland B, Unwin ID, Buss DH. Fruit and nuts. First supplement to McCance and Widdowson's The composition of foods, 5th ed. Cambridge: Royal Society of Chemistry; 1992.

[115] Holland B, Welch A, Buss DH. Vegetable dishes. Second supplement to McCance and Widdowson's The composition of foods, 5th ed. Cambridge: Royal Society of Chemistry; 1992.

[116] Holland B, Brown J, Buss DH. Fish and fish products. Third supplement to McCance and Widdowson's The composition of foods, 5th ed. Cambridge: Royal Society of Chemistry; 1993.

[117] Chan W, Brown J, Buss DH. Miscellaneous foods. Fourth supplement to McCance and Widdowson's The composition of foods, 5th ed. Cambridge: Royal Society of Chemistry; 1994.

[118] Chan W, Brown J, Lee SM, Buss DH. Meat, poultry and game. Fifth supplement to McCance and Widdowson's The composition of foods, 5th ed. Cambridge: Royal Society of Chemistry; 1995.

[119] Chan W, Brown J, Church SM, Buss DH. Meat products and dishes. Sixth supplement to McCance and Widdowson's The composition of foods, 5th ed. Cambridge: Royal Society of Chemistry; 1996.

[120] US Department of Agriculture, Agricultural Research Service. USDA National Nutrient Database for Standard Reference, Release 26. Available at: `http://www.ars.usda.gov/ba/bhnrc/ndl`. Accessed 10 Apr to 6 May 2014.

[121] Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr. 2003;6(4):407–13.

[122] Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. Public Health Nutr. 2002;5(4):561–5.

[123] Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. Mod Pathol. 2004;17(3):292–306.

[124] Biocrates Life Sciences. The Standard in Targeted Metabolomics, Absolute*IDQ* p180 Kit; 2017. Available at: `http://www.biocrates.com/images/p180_Folder_2017-09-05.pdf`.

[125] Siskos AP, Jain P, Romisch-Margl W, Bennett M, Achaintre D, Asad Y, et al. Interlaboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma. Anal Chem. 2017;89(1):656–65.

[126] Floegel A, Drogan D, Wang-Sattler R, Prehn C, Illig T, Adamski J, et al. Reliability of serum metabolite concentrations over a 4-month period using a targeted metabolomic approach. PLoS One. 2011;6(6):e21103.

[127] Bender DA, Millward DJ. Protein metabolism and requirements. In: Geissler C, Powers H, editors. Human Nutrition. 11th ed. Edinburgh, London, New York, Oxford, Philidelphia, St. Louis, Sidney and Toronto: Elsevier Limited; 2005. p. 143–64.

[128] Boye J, Wijesinha-Bettoni R, Burlingame B. Protein quality evaluation twenty years after the introduction of the protein digestibility corrected amino acid score method. Br J Nutr. 2012;108 Suppl 2:S183–211.

[129] Cynober LA. Plasma amino acid levels with a note on membrane transport: characteristics, regulation, and metabolic significance. Nutrition. 2002;18(9):761–6.

[130] Key TJ, Appleby PN, Crowe FL, Bradbury KE, Schmidt JA, Travis RC. Cancer in British vegetarians: updated analyses of 4998 incident cancers in a cohort of 32,491 meat eaters, 8612 fish eaters, 18,298 vegetarians, and 2246 vegans. Am J Clin Nutr. 2014;100 Suppl 1:378s–85s.

[131] Tantamango-Bartley Y, Knutsen SF, Knutsen R, Jacobsen BK, Fan J, Beeson WL, et al. Are strict vegetarians protected against prostate cancer? Am J Clin Nutr. 2016;103(1):153–60.

[132] Abdulla M, Andersson I, Asp NG, Berthelsen K, Birkhed D, Dencker I, et al. Nutrient intake and health status of vegans. Chemical analyses of diets using the duplicate portion sampling technique. Am J Clin Nutr. 1981;34(11):2464–77.

[133] Abdulla M, Aly KO, Andersson I, Asp NG, Birkhed D, Denker I, et al. Nutrient intake and health status of lactovegetarians: chemical analyses of diets using the duplicate portion sampling technique. Am J Clin Nutr. 1984;40(2):325–38.

[134] Rana SK, Sanders TA. Taurine concentrations in the diet, plasma, urine and breast milk of vegans compared with omnivores. Br J Nutr. 1986;56(1):17–27.

[135] Laidlaw SA, Shultz TD, Cecchino JT, Kopple JD. Plasma and urine taurine levels in vegans. Am J Clin Nutr. 1988;47(4):660–3.

[136] Paul AA, Southgate DAT. McCance and Widdowson's The Composition of Foods. 4th ed. London: Her Majesty's Stationery Office; 1978.

[137] McKee T, McKee J. Biochemistry – the molecular basis of life. 3rd ed. New York: McGraw-Hill; 2003.

[138] Lenders CM, Liu S, Wilmore DW, Sampson L, Dougherty LW, Spiegelman D, et al. Evaluation of a novel food composition database that includes glutamine and other amino acids derived from gene sequencing data. Eur J Clin Nutr. 2009;63(12):1433–9.

[139] Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. Malden, MA, USA; Oxford, UK; and Victoria, Australia: Blackwell Science Ltd; 2003.

[140] Chadeau-Hyam M, Campanella G, Jombart T, Bottolo L, Portengen L, Vineis P, et al. Deciphering the complex: methodological overview of statistical models to derive OMICS-based biomarkers. Environ Mol Mutagen. 2013;54(7):542–57.

[141] Metges CC. Contribution of microbial amino acids to amino acid homeostasis of the host. J Nutr. 2000;130(7):1857S–64S.

[142] Willett W. Nutritional Epidemiology. 2nd ed. Monographs in Epidemiology and Biostatistics. New York and Oxford: Oxford University Press; 1998.

[143] Gurdeniz G, Hansen L, Rasmussen MA, Acar E, Olsen A, Christensen J, et al. Patterns of time since last meal revealed by sparse PCA in an observational LC-MS based metabolomics study. Metabolomics. 2013;9(5):1073–81.

[144] Breier M, Wahl S, Prehn C, Fugmann M, Ferrari U, Weise M, et al. Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples. PLoS One. 2014;9(2):e89728.

[145] Anton G, Wilson R, Yu ZH, Prehn C, Zukunft S, Adamski J, et al. Pre-analytical sample quality: metabolite ratios as an intrinsic marker for prolonged room temperature exposure of serum samples. PLoS One. 2015;10(3):e0121495.

[146] Townsend MK, Clish CB, Kraft P, Wu C, Souza AL, Deik AA, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin Chem. 2013;59(11):1657–67.

[147] Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes. 2013;62(2):639–48.

[148] Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448–53.

[149] Wu W, Kang S, Zhang D. Association of vitamin B6, vitamin B12 and methionine with risk of breast cancer: a dose-response meta-analysis. Br J Cancer. 2013;109(7):1926–44.

[150] Larsson SC, Giovannucci E, Wolk A. Methionine and vitamin B6 intake and risk of pancreatic cancer: a prospective study of Swedish women and men. Gastroenterology. 2007;132(1):113–8.

[151] Xu J, Yang S, Cai S, Dong J, Li X, Chen Z. Identification of biochemical changes in lactovegetarian urine using ^{1}H NMR spectroscopy and pattern recognition. Anal Bioanal Chem. 2010;396(4):1451–63.

[152] Wu GD, Compher C, Chen EZ, Smith SA, Shah RD, Bittinger K, et al. Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. Gut. 2016;65(1):63–72.

[153] Bradbury KE, Crowe FL, Appleby PN, Schmidt JA, Travis RC, Key TJ. Serum concentrations of cholesterol, apolipoprotein A-I and apolipoprotein B in a total of 1694 meat-eaters, fish-eaters, vegetarians and vegans. Eur J Clin Nutr. 2014;68(2):178–83.

[154] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol. 1995;57(1):289–300.

[155] Cole LK, Vance JE, Vance DE. Phosphatidylcholine biosynthesis and lipoprotein metabolism. Biochim Biophys Acta. 2012;1821(5):754–61.

[156] Jolliffe IT. Principal component analysis. 2nd ed. Springer series in statistics. New York: Springer; 2002.

[157] Suhr D. Principal Component Analysis vs. Exploratory Factor Analysis. Available at: http://www.johnverostek.com/wp-content/uploads/2014/01/203-30.pdf.

[158] Trygg J, Lundsted T. Chemometrics Techniques for Metabonomics. In: Lindon JC, Nicholson JK, Holmes E, Lindon JC, editors. The Handbook of Metabonomics and Metabolomics. Amsterdam and London: Elsevier; 2007. p. 171–99.

[159] Moskal A, Pisa PT, Ferrari P, Byrnes G, Freisling H, Boutron-Ruault MC, et al. Nutrient patterns and their food sources in an International Study Setting: report from the EPIC study. PLoS One. 2014;9(6):e98647.

[160] Gorst-Rasmussen A, Dahm CC, Dethlefsen C, Scheike T, Overvad K. Exploring dietary patterns by using the treelet transform. Am J Epidemiol. 2011;173(10):1097–104.

[161] Fages A, Ferrari P, Monni S, Dossus L, Floegel A, Mode N, et al. Investigating sources of variability in metabolomic data in the EPIC study: the Principal Component Partial R-square (PC-PR2) method. Metabolomics. 2014;10(6):1074–1083.

[162] Zeisel SH, Mar MH, Howe JC, Holden JM. Concentrations of choline-containing compounds and betaine in common foods. J Nutr. 2003;133(5):1302–7.

[163] US Department of Agriculture, Agricultural Research Service. USDA Database for the Choline Content of Common Foods, Release 2. Available at: http://www.ars.usda.gov/Services/docs.htm?docid=6232. Accessed 20 Oct 2014.

[164] Gault CR, Obeid LM, Hannun YA. An overview of sphingolipid metabolism: from synthesis to breakdown. Adv Exp Med Biol. 2010;688:1–23.

[165] Holmes E, Loo RL, Stamler J, Bictash M, Yap IK, Chan Q, et al. Human metabolic phenotype diversity and its association with diet and blood pressure. Nature. 2008;453(7193):396–400.

[166] Yu Z, Kastenmuller G, He Y, Belcredi P, Moller G, Prehn C, et al. Differences between human plasma and serum metabolite profiles. PLoS One. 2011;6(7):e21230.

[167] Guertin KA, Moore SC, Sampson JN, Huang WY, Xiao Q, Stolzenberg-Solomon RZ, et al. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. Am J Clin Nutr. 2014;100(1):208–17.

[168] O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. Am J Clin Nutr. 2011;93(2):314–21.

[169] Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, et al. Susceptibility of human metabolic phenotypes to dietary modulation. J Proteome Res. 2006;5(10):2780–8.

[170] Cross AJ, Major JM, Sinha R. Urinary biomarkers of meat consumption. Cancer Epidemiol Biomarkers Prev. 2011;20(6):1107–11.

[171] Sampson JN, Boca SM, Shu XO, Stolzenberg-Solomon RZ, Matthews CE, Hsing AW, et al. Metabolomics in epidemiology: sources of variability in metabolite measurements and implications. Cancer Epidemiol Biomarkers Prev. 2013;22(4):631–40.

[172] Townsend MK, Bao Y, Poole EM, Bertrand KA, Kraft P, Wolpin BM, et al. Impact of Pre-analytic Blood Sample Collection Factors on Metabolomics. Cancer Epidemiol Biomarkers Prev. 2016;25(5):823–9.

[173] Noble WS. How does multiple testing correction work? Nat Biotechnol. 2009;27(12):1135–7.

[174] Crowe FL, Appleby PN, Travis RC, Key TJ. Risk of hospitalization or death from ischemic heart disease among British vegetarians and nonvegetarians: results from the EPIC-Oxford cohort study. Am J Clin Nutr. 2013;97(3):597–603.

[175] Fernandez C, Sandin M, Sampaio JL, Almgren P, Narkiewicz K, Hoffmann M, et al. Plasma lipid composition and risk of developing cardiovascular disease. PLoS One. 2013;8(8):e71846.

[176] Stegemann C, Pechlaner R, Willeit P, Langley SR, Mangino M, Mayr U, et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. Circulation. 2014;129(18):1821–31.

[177] Ridgway ND. The role of phosphatidylcholine and choline metabolites to cell proliferation and survival. Crit Rev Biochem Mol Biol. 2013;48(1):20–38.

[178] Taniguchi M, Okazaki T. The role of sphingomyelin and sphingomyelin synthases in cell death, proliferation and migration-from cell and animal models to human disorders. Biochim Biophys Acta. 2014;1841(5):692–703.

[179] Mitrunen K, Pettersson K, Piironen T, Bjork T, Lilja H, Lovgren T. Dual-label one-step immunoassay for simultaneous measurement of free and total prostate-specific antigen concentrations and ratios in serum. Clin Chem. 1995;41(8):1115–20.

[180] Breuillard C, Cynober L, Moinard C. Citrulline and nitrogen homeostasis: an overview. Amino Acids. 2015;47(4):685–91.

[181] Paschos A, Pandya R, Duivenvoorden WC, Pinthus JH. Oxidative stress in prostate cancer: changing research concepts towards a novel paradigm for prevention and therapeutics. Prostate Cancer Prostatic Dis. 2013;16(3):217–25.

[182] Fernandez-Peralbo M, Gomez-Gomez E, Calderon-Santiago M, Carrasco-Valiente J, Ruiz-Garcia J, Requena-Tapia M, et al. Prostate Cancer Patients–Negative Biopsy Controls Discrimination by Untargeted Metabolomics Analysis of Urine by LC-QTOF: Upstream Information on Other Omics. Scientific Reports. 2016;6:38243.

[183] Heller W, Harzmann R, Bichler KH, Schmidt K. Urinary hydroxyproline in healthy patients and in prostate patients with and without bone metastases. Curr Probl Clin Biochem. 1979;9:249–56.

[184] Mooppan MM, Wax SH, Kim H, Wang JC, Tobin MS. Urinary hydroxyproline excretion as a marker of osseous metastasis in carcinoma of the prostate. J Urol. 1980;123(5):694–6.

[185] Phang JM, Donald SP, Pandhare J, Liu Y. The metabolism of proline, a stress substrate, modulates carcinogenic pathways. Amino Acids. 2008;35(4):681–90.

[186] Ross AB, Svelander C, Undeland I, Pinto R, Sandberg AS. Herring and Beef Meals Lead to Differences in Plasma 2-Aminoadipic Acid, beta-Alanine, 4-Hydroxyproline, Cetoleic Acid, and Docosahexaenoic Acid Concentrations in Overweight Men. J Nutr. 2015;145(11):2456–63.

[187] Pallister T, Jennings A, Mohney RP, Yarand D, Mangino M, Cassidy A, et al. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. PLoS One. 2016;11(6):e0158568.

[188] Wu K, Spiegelman D, Hou T, Albanes D, Allen NE, Berndt SI, et al. Associations between unprocessed red and processed meat, poultry, seafood and egg intake and the risk of prostate cancer: A pooled analysis of 15 prospective cohort studies. Int J Cancer. 2016;138(10):2368–82.

[189] Narayan P, Dahiya R. Alterations in sphingomyelin and fatty acids in human benign prostatic hyperplasia and prostatic cancer. Biomed Biochim Acta. 1991;50(9):1099–108.

[190] Zhou X, Mao J, Ai J, Deng Y, Roth MR, Pound C, et al. Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics. PLoS One. 2012;7(11):e48889.

[191] Beger RD. A review of applications of metabolomics in cancer. Metabolites. 2013;3(3):552–74.

[192] Baenke F, Peck B, Miess H, Schulze A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. Dis Model Mech. 2013;6(6):1353–63.

[193] Carayol M, Licaj I, Achaintre D, Sacerdote C, Vineis P, Key TJ, et al. Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC. PLoS One. 2015;10(8):e0135437.

[194] Gonzalez-Covarrubias V, Dane A, Hankemeier T, Vreeken RJ. The influence of citrate, EDTA, and heparin anticoagulants to human plasma LC-MS lipidomic profiling. Metabolomics. 2013;9(2):337–48.

[195] Barri T, Dragsted LO. UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics: effect of experimental artefacts and anticoagulant. Anal Chim Acta. 2013;768:118–28.

[196] Allsbrook J W C, Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver Reproducibility of Gleason Grading of Prostatic Carcinoma: General Pathologists. Hum Pathol. 2001;32(1):81–8.

[197] Gordetsky J, Epstein J. Grading of prostatic adenocarcinoma: current state and prognostic implications. Diagn Pathol. 2016;11:25.

[198] AJCC – American Joint Committee on Cancer. Cancer Staging Manual. Available at: `https://cancerstaging.org/references-tools/deskreferences/Pages/default.aspx`. Accessed 13 Mar 2017.

[199] Persson J, Wilderang U, Jiborn T, Wiklund PN, Damber JE, Hugosson J, et al. Interobserver variability in the pathological assessment of radical prostatectomy specimens: findings of the Laparoscopic Prostatectomy Robot Open (LAPPRO) study. Scand J Urol. 2014;48(2):160–7.

[200] Fraser M, Berlin A, Bristow RG, van der Kwast T. Genomic, pathological, and clinical heterogeneity as drivers of personalized medicine in prostate cancer. Urol Oncol. 2015;33(2):85–94.

[201] Foerster J, Hyotylainen T, Oresic M, Nygren H, Boeing H. Serum Lipid and Serum Metabolite Components in relation to anthropometric parameters in EPIC-Potsdam participants. Metabolism. 2015;64(10):1348–58.

[202] Gorst-Rasmussen A. tt: Treelet transform with Stata. Stata Journal. 2012;12(1):130–46.

[203] Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev. 2004;62(5):177–203.

[204] Dahm CC, Gorst-Rasmussen A, Crowe FL, Roswall N, Tjonneland A, Drogan D, et al. Fatty acid patterns and risk of prostate cancer in a case-control study nested within the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr. 2012;96(6):1354–61.

[205] Schmidt JA, Gorst-Rasmussen A, Nystrom PW, Christensen JH, Schmidt EB, Dethlefsen C, et al. Baseline patterns of adipose tissue fatty acids and long-term risk of breast cancer: a case-cohort study in the Danish cohort Diet, Cancer and Health. Eur J Clin Nutr. 2014;68(10):1088–94.

[206] Assi N, Moskal A, Slimani N, Viallon V, Chajes V, Freisling H, et al. A treelet transform analysis to relate nutrient patterns to the risk of hormonal receptor-defined breast cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). Public Health Nutr. 2016;19(2):242–54.

[207] Schoenaker DA, Dobson AJ, Soedamah-Muthu SS, Mishra GD. Factor Analysis is More Appropriate to Identify Overall Dietary Patterns Associated with Diabetes When Compared with Treelet Transform Analysis. J Nutr. 2013;143(3):392–8.

[208] Reuter SE, Evans AM. Carnitine and Acylcarnitines. Clin Pharmacokinet. 2012;51(9):553–72.

[209] Roy MJ, Dionne S, Marx G, Qureshi I, Sarma D, Levy E, et al. In vitro studies on the inhibition of colon cancer by butyrate and carnitine. Nutrition. 2009;25(11-12):1193–201.

[210] Dionne S, Elimrani I, Roy MJ, Qureshi IA, Sarma DR, Levy E, et al. Studies on the chemopreventive effect of carnitine on tumorigenesis in vivo, using two experimental murine models of colon cancer. Nutr Cancer. 2012;64(8):1279–87.

[211] Roscilli G, Marra E, Mori F, Di Napoli A, Mancini R, Serlupi-Crescenzi O, et al. Carnitines slow down tumor development of colon cancer in the DMH-chemical carcinogenesis mouse model. J Cell Biochem. 2013;114(7):1665–73.

[212] Ni J, Xu L, Li W, Wu L. Simultaneous determination of thirteen kinds of amino acid and eight kinds of acylcarnitine in human serum by LC-MS/MS and its application to measure the serum concentration of lung cancer patients. Biomed Chromatogr. 2016;30(11):1796–1806.

[213] Jin X, Yun SJ, Jeong P, Kim IY, Kim WJ, Park S. Diagnosis of bladder cancer and prediction of survival by urinary metabolomics. Oncotarget. 2014;5(6):1635–45.

[214] Miyagi Y, Higashiyama M, Gochi A, Akaike M, Ishikawa T, Miura T, et al. Plasma free amino acid profiling of five types of cancer patients and its application for early detection. PLoS One. 2011;6(9):e24143.

[215] Derezinski P, Klupczynska A, Sawicki W, Palka JA, Kokot ZJ. Amino Acid Profiles of Serum and Urine in Search for Prostate Cancer Biomarkers: a Pilot Study. Int J Med Sci. 2017;14(1):1–12.

[216] McNeal CJ, Meininger CJ, Reddy D, Wilborn CD, Wu G. Safety and Effectiveness of Arginine in Adults. J Nutr. 2016;146(12):2587s–93s.

[217] Eidelman E, Twum-Ampofo J, Ansari J, Siddiqui MM. The Metabolic Phenotype of Prostate Cancer. Front Oncol. 2017;7:131.

[218] Giovannucci E, Rimm EB, Wolk A, Ascherio A, Stampfer MJ, Colditz GA, et al. Calcium and fructose intake in relation to risk of prostate cancer. Cancer Res. 1998;58(3):442–7.

[219] Van Hemelrijck M, Garmo H, Holmberg L, Walldius G, Jungner I, Hammar N, et al. Prostate cancer risk in the Swedish AMORIS study: the interplay among triglycerides, total cholesterol, and glucose. Cancer. 2011;117(10):2086–95.

[220] Haggstrom C, Stocks T, Ulmert D, Bjorge T, Ulmer H, Hallmans G, et al. Prospective study on metabolic factors and risk of prostate cancer. Cancer. 2012;118(24):6199–206.

[221] Cutruzzola F, Giardina G, Marani M, Macone A, Paiardini A, Rinaldo S, et al. Glucose Metabolism in the Progression of Prostate Cancer. Front Physiol. 2017;8:97.

[222] Kumar RG, Juengst SB, Wang Z, Dams-O'Connor K, Dikmen SS, O'Neil-Pirozzi TM, et al. Epidemiology of Comorbid Conditions Among Adults 50 Years and Older With Traumatic Brain Injury. J Head Trauma Rehabil. 2017;Epub ahead of print.

[223] Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun. 2016;7:11122.

[224] Giovannucci E, Pollak M, Liu Y, Platz EA, Majeed N, Rimm EB, et al. Nutritional predictors of insulin-like growth factor I and their relationships to cancer in men. Cancer Epidemiol Biomarkers Prev. 2003;12(2):84–9.

[225] Nelson DL, Cox MM. Lehninger Principles of Biochemistry. 6th ed. Basingstoke: Macmillan Higher Education; 2013.

[226] Curis E, Nicolis I, Moinard C, Osowska S, Zerrouk N, Benazeth S, et al. Almost all about citrulline in mammals. Amino Acids. 2005;29(3):177–205.

[227] Nikolic ZZ, Pavicevic D, Romac SP, Brajuskovic GN. Genetic variants within endothelial nitric oxide synthase gene and prostate cancer: a meta-analysis. Clin Transl Sci. 2015;8(1):23–31.

[228] Janakiram NB, Mohammed A, Madka V, Kumar G, Rao CV. Prevention and treatment of cancers by immune modulating nutrients. Mol Nutr Food Res. 2016;60(6):1275–94.

[229] Gannon PO, Godin-Ethier J, Hassler M, Delvoye N, Aversa M, Poisson AO, et al. Androgen-regulated expression of arginase 1, arginase 2 and interleukin-8 in human prostate cancer. PLoS One. 2010;5(8):e12107.

[230] Dejong CH, van de Poll MC, Soeters PB, Jalan R, Olde Damink SW. Aromatic amino acid metabolism during liver failure. J Nutr. 2007;137(6 Suppl 1):1579S–85S.

[231] Benoit K. Linear Regression Models with Logarithmic Transformations. Methodology Institute London School of Economics; 2011. Available at: `http://kenbenoit.net/assets/courses/ME104/logmodels2.pdf`.

[232] Wu G, Bazer FW, Burghardt RC, Johnson GA, Kim SW, Knabe DA, et al. Proline and hydroxyproline metabolism: implications for animal and human nutrition. Amino Acids. 2011;40(4):1053–63.

[233] US Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory. USDA National Nutrient Database for Standard Reference, Release 28. Available at: `http://www.ars.usda.gov/ba/bhnrc/ndl`. Accessed 18 Jul 2017.

[234] Bates CJ, Nelson M, Ulijaszek SJ. Nutritional assessment methods. In: Geissler CA, Powers HJ, editors. Human Nutrition. 11th ed. Edinburgh, London, New York, Oxford, Philidelphia, St. Louis, Sidney and Toronto: Churchill Livingstone; 2005. p. 573–95.

[235] Key TJ, Appleby PN, Reeves GK, Roddam AW. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. Lancet Oncol. 2010;11(6):530–42.

[236] Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet. 2004;363(9418):1346–53.

[237] Lukanova A, Lundin E, Zeleniuch-Jacquotte A, Muti P, Mure A, Rinaldi S, et al. Body mass index, circulating levels of sex-steroid hormones, IGF-I and IGF-binding protein-3: a cross-sectional study in healthy women. Eur J Endocrinol. 2004;150(2):161–71.

[238] Kanaley JA. Growth hormone, arginine and exercise. Curr Opin Clin Nutr Metab Care. 2008;11(1):50–4.

[239] Hsueh EC, Knebel SM, Lo WH, Leung YC, Cheng PN, Hsueh CT. Deprivation of arginine by recombinant human arginase in prostate cancer cells. J Hematol Oncol. 2012;5:17.

[240] Fraser GE. Associations between diet and cancer, ischemic heart disease, and all-cause mortality in non-Hispanic white California Seventh-day Adventists. Am J Clin Nutr. 1999;70(3 Suppl):532s–8s.

[241] Turner EL, Metcalfe C, Donovan JL, Noble S, Sterne JA, Lane JA, et al. Contemporary accuracy of death certificates for coding prostate cancer as a cause of death: Is reliance on death certification good enough? A comparison with blinded review by an independent cause of death evaluation committee. Br J Cancer. 2016;115(1):90–4.

[242] Perez-Gomez B, Aragones N, Pollan M, Suarez B, Lope V, Llacer A, et al. Accuracy of cancer death certificates in Spain: a summary of available information. Gac Sanit. 2006;20 Suppl 3:42–51.

[243] Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1–22.

[244] Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, et al. Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. PLoS Med. 2016;13(11):e1002179.

[245] The Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163(4):1011–25.

[246] Stolzenberg-Solomon RZ, Moore S, Ulrich C, Poole E, Temprosa M, Verma M, et al. Abstract 4238: COnsortium for METabolomics Studies (COMETS): leveraging resources to accelerate scientific discovery. Cancer Res. 2017;77(13 Supplement):Abstract nr 4238.

[247] National Cancer Institute, Division of Cancer Control & Population Sciences, Epidemiology and Genomics Research Program. COnsortium of METabolomics Studies. Available at: `https://epi.grants.cancer.gov/comets/`. Accessed 04 Aug 2017.

[248] Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford. Endogenous Hormones, Nutritional Biomarkers and Prostate Cancer. Available at: https://www.ceu.ox.ac.uk/research/endogenous-hormones-nutritional-biomarkers-and-prostate-cancer. Accessed 03 Aug 2017.