

Untangling the links between obesity and cancer using genomic and phenomic approaches

Dr Benjamin H L Harris

MBBCh (Hons) MSc MRCP (UK) Fellow, HEA FAcadMEd OA

Cancer Research UK Clinical Training Research Fellow

Submission for consideration of Doctor of Philosophy

Primary Supervisor: Professor Francesca Buffa

Other Supervisors: Dr Simon Lord, Professor Lisa Heather

SARS-CoV-2 Pandemic Impact Statement

During my DPhil, the SARS-CoV-2 pandemic disrupted all aspects of our lives. My research and therefore this thesis were not exempt, particularly as I attempted to help as a clinician and as an academic. This thesis therefore took more of a bioinformatic bent than first anticipated, and even incorporates some research into SARS-CoV-2 (*Chapter 5*).

Statement of originality

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own. This is also reflected in the authorship of the related submitted papers (overleaf).

With all research there is collaboration and I must sincerely thank those who have helped me in my research journey (see acknowledgments).

Summary of first author works:

Harris B.H.L.*, Di Giovannantonio M.*, Zhang P., Kitchen-Smith I., Xiong L., Sahgal N., Stracquadanio G., Wallace M., Blagden S., Lord S., Harris D.A., Harris A.L, Buffa F.M., Bond G. Heritable genetic variants in key cancer genes link cancer risk with anthropometric traits *Journal of Medical Genetics* 2021 58(6): 392-399.

Role: study design, data cleaning, bioinformatics analysis, paper drafting, iteration and submission. This work is related to *Chapter 3*.

Harris B.H.L.*, Zuhair M.*, Di Giovannantonio M., Rosadas C., Khan M., Short C., Thaventhiran T., Quinlan R., Taylor A, Calvez R., Taylor G.P, Tedder R.S, McClure M.O, Fertleman M. Asymptomatic Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection in a Rehabilitation Facility: Evolution of the Presence of Nasopharyngeal SARS-CoV-2 and Serological Antibody Responses
The Journal of Infectious Diseases 2021 223(2): 192-196.

Role: study design, patient consent, sample collection, statistical analysis, paper drafting, iteration and submission. This work is related to *Chapter 5*.

Harris B.H.L.*, Pereira C.*, Di Giovannantonio M., Rosadas C., Short C., Quinlan R., Sureda-Vives M., Fernandez N., Day-Weber I., Khan M., Marchesin F., Katsanovskaja K., Parker E., Taylor G.P, Tedder R.S, McClure M.O, Dani M., Fertleman M. The Association Between Antibody Response to Severe Acute Respiratory Syndrome Coronavirus 2 Infection and Post-COVID-19 Syndrome in Healthcare Workers
The Journal of Infectious Diseases 2021 223(10): 1671-1676.

Role: study design, sample collection, statistical analysis, paper drafting, iteration and submission. This work is related to *Chapter 5*.

Harris B.H.L.*, Di Giovannantonio M.*, Zhang P., Harris D.A, Karpe F., Lord S., Allen N.E., Maughan T.S., Bryant R. J., Harris A.L., Bond G., Buffa F.M. Integrated phenotypic and genetic analyses highlight the important role of fat-free mass in cancer risk
Nature Communications [in review]

Role: study design, data cleaning, bioinformatics analysis, paper drafting, iteration and submission. This work is related to *Chapters 2 & 4*.

Harris B.H.L., Macaulay V., Harris D.A, Karpe F., Lord S., Harris A.L, Buffa F.M. Obesity: a perfect storm for carcinogenesis
Nature Reviews Cancer [in review]

Role: article design, literature review, paper drafting, iteration and submission. This work is related to Chapter 1.

*joint first author

Other published works during my DPhil:

Skwarski M., McGowan D.R., Belcher E., Di Chiara F., Stavroulias D., McCole M., Derham J.L., Chu K.Y., Teoh E., Chauhan J., O'Reilly D., **Harris B.H.L.**, Macklin P.S., Bull J.A., Green M., Rodriguez-Berriguete G., Prevo R., Folkes L.K., Campo L., Ferencz P., Croal P.L., Flight H., Qi C., Holmes J., O'Connor J.P.B., Gleeson F.V., McKenna W.G., Harris A.L., Bulte D., Buffa F.M., Macpherson R.E., Higgins G.S.

Mitochondrial Inhibitor Atovaquone Increases Tumor Oxygenation and Inhibits Hypoxic Gene Expression in Patients with Non-Small Cell Lung Cancer
Clinical Cancer Research 2021 27(9): 2459-2469

Role: sample preparation, RNA extraction, paper drafting and iteration.

List of Abbreviations/Acronyms 1
Abstract 8

Chapter 1: Obesity, a perfect storm for carcinogenesis 11
Chapter summary 12
1.1 Introduction 13
1.1.1 Biochemistry and physiology in obesity 16
1.1.2 Hallmarks of cancer 18
1.2 Sustaining Proliferative Signalling 26
1.3 Evading Growth Suppressors 31
1.4 Resisting Cell Death 36
1.5 Enabling Replicative Immortality 40
1.6 Inducing Angiogenesis 42
1.7 Activating Invasion and Metastasis 45
1.8 An Enabling Characteristic: Genome Instability and Mutation 52
1.9 An Enabling Characteristic: Tumor-Promoting Inflammation 55
1.10 An Emerging Hallmark: Avoiding Immune Destruction 59
1.11 An Emerging Hallmark: Reprogramming Energy Metabolism 64
1.11.1 Glycolysis & mitochondrial metabolism 64
1.11.2 Free fatty acids 66
1.11.3 mTOR & AMPK 67
1.11.4 TP53 68
1.11.5 Adipocyte-derived exosomes 69
1.12 Obesity and cancer presentation and treatment 71
1.13 Conclusion 76

Chapter 2: Untangling cancer, body mass index and obesity: anthropometric risk factors in breast, colorectal and prostate cancer 79
Chapter Summary 80
2.1 Introduction 82
2.2 Methods 86
2.2.1 The UK Biobank 87
2.2.2 Cancer data 87
2.2.3 Anthropometry and body composition 88
2.2.4 Study population 89
2.2.5 Epidemiological Analysis 90
2.3 Results 93
2.3.1 Post-menopausal breast cancer 93
2.3.2 Prostate 98
2.3.3 Colorectal 103
2.4 Discussion 113
2.4.1 BMI, WBFM and WBFFM across three cancer types 113
2.4.2 Interpretation of estimates of WBFFM and WBFM 117
2.4.3 Other interesting findings: pharmacological interventions 118

2.4.4 Conclusion 119

Chapter 3: Cancer risk and anthropometric traits - heritable genetic variants in tumour suppressor genes and proto-oncogenes 122

Chapter Summary 123

3.1 Introduction 124

3.2 Methods 127

3.2.1 The UK Biobank 127

3.2.2 Anthropometric data 127

3.2.3 Cancer data 127

3.2.4 Genetic data 128

3.2.5 Sample quality control 128

3.2.6 Identification of functional cancer gene SNPs 128

3.2.7 SNP quality control 130

3.2.8 SNP-wise analysis 131

3.2.9 Linkage and lead SNPs 131

3.3 Results 133

3.4 Discussion 142

3.4.1 MAP3K1 142

3.4.2 FANCA 145

3.4.3 TP53 148

3.5 Conclusion 156

Chapter 4: Genetic pleiotropy, obesity and anthropometric traits 157

Chapter Summary 157

4.1 Introduction 160

4.2 Methods 163

4.2.1 The UK Biobank 163

4.2.2 Genomic Association Analyses 163

4.2.3 Pleiotropic Associations 163

4.2.4 Expression quantitative trait loci 164

4.3 Results 166

4.3.1 GWAS of BMI, WBFFM and WBFM 166

4.3.1.1 BMI 166

4.3.1.2 WBFFM 169

4.3.1.3 WBFM 172

4.3.2 GWAS of Colorectal, Prostate and Post-Menopausal Cancers 174

4.3.2.1 Colorectal 174

4.3.2.2 Prostate 177

4.3.2.3 Post-menopausal breast cancer 179

4.3.3 Simultaneous anthropogenetic and oncogenic pleiotropy 181

4.4 Discussion 187

4.4.1 CUPID hits the spot 189

4.5 Conclusion 192

Chapter 5: An unexpected ending - Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) 194
Preface 195
Chapter Summary 196
5.1 Introduction 198
5.2 Methods 201
5.2.1 Asymptomatic patients 201
5.2.2 Healthcare workers 202
5.2.3 RT-PCR 202
5.2.4 Detection of anti- SARS-CoV-2 antibodies 202
5.2.5 Statistical analysis 203
5.2.6 Pleiotropic association analysis 203
5.3 Results 204
5.3.1 Asymptomatic patients: evolution of the presence of nasopharyngeal SARS-CoV-2 and serological antibody responses 204
5.3.2 Nasopharyngeal SARS-CoV-2 levels and serological antibody responses in healthcare workers: relation to the development of Post-COVID-19 syndrome 207
5.3.3 SNPs, SARS-CoV-2 and genetic pleiotropy 211
5.4 Discussion 213
5.4.1 Asymptomatic SARS-CoV-2 patients & individuals with Post-COVID-19 Syndrome 213
5.4.2 SNPs, SARS-CoV-2 and genetic pleiotropy 215

Chapter 6: Final Thoughts 217

Additional investigations and achievements 222

Acknowledgements 225

Appendices 228

References 271

List of Abbreviations/Acronyms

AdipoR 1,2	Adiponectin Receptor 1,2
AEBP	Adipocyte Enhancer Binding Protein
AF	Atrial Fibrillation
A-FABP	Adipose-derived Fatty Acid Binding Protein
AhR	Aryl hydrocarbon Receptor
Akt	Protein Kinase B
AKTIP	AKT Interacting Protein
ALT	Alternative Lengthening of Telomeres pathway
AMPK	AMP-dependent Kinase
ANGPTLs	Angiopoietin-Like proteins
AoGV	Adoxophyes orana GranuloVirus
APPL1	Adaptor Protein, Phosphotyrosine interacting with PH domain and Leucine zipper 1
ARH1	ADP-ribose(arginine)protein hydrolase 1
ASCs	Adipose-derived Stromal Cells
ASCT2	Sodium-Linked aminoacid transporter
ATR	Ataxia-Telangiectasia and Rad3-related protein
ATRIP	ATR-interacting protein
BAME	Black And Minority Ethnic
Bax	Bcl-2 Associated X-protein
B-raf	B-Raf Proto-Oncogene, Serine/Threonine Kinase
BMI	Body Mass Index
CCLE	Cancer Cell Line Encyclopedia
CCND	Cyclin D
ChK1	Checkpoint kinase 1

cMET (HGFR)	Mesenchymal Epithelial Transition factor (HGF receptor)
COSMIC	Catalogue of Somatic Mutations In Cancer
COX-2	Cyclooxygenase 2
CRC	Colorectal cancer
Ct	Cycle Threshold (in quantitive PCR)
CTPA	Computerised Tomography Pulmonary Angiogram
CUPID1	CCND1 Upstream Intergenic DNA repair 1, non-coding
DABA	Double Antigen Binding Assay
DAMPs	Damage-Associated Molecular Patterns
DCC	'Deleted in Colorectal Cancer' tumour suppressor
EBV	Epstein-Barr Virus
EGFR	Epidermal Growth Factor Receptor
EIF	Eukaryotic (protein synthesis) Initiation Factor
ELISA	Enzyme Linked ImmunoAssay
EMT	Epithelial to Mesenchymal Transition
eQTL	expression Quantitative Trait Loci
ER+/-	Oestrogen receptor positive/negative
FAK (PTK2)	Focal adhesion kinase (Protein tyrosine kinase 2)
FANCA	Fanconi Anaemia Complementation group A
FAO	Fatty Acid Oxidation
FasL	*F*S-7-*a*ssociated *s*urface antigen ligand
FFMI	Fat-Free Mass Index
FGFR	Fibroblast Growth Factor Receptor
FMI	Fat-Mass Index
FMI	Fat Mass Index
FTO	Fat Mass and Obesity associated gene
GAPdH	GlycerAldehyde 3-Phosphate dehydrogenase
GH	Growth Hormone
GIANT	Genetic Investigation of ANthropometric Traits

Glut 1,2 …	Members of the Glucose transporter family
GPX4	Glutathione Peroxidase 4
GWAS	Genome Wide Association Studies
HDL/VLDL	High Density Lipoprotein/Very Low Density Lipoprotein.
HER3	Human Epidermal Growth Factor Receptor 3
HGF	Hepatocyte Growth Factor
HHT	Hereditary Hemorrhagic Telangiectasia
HIF	Hypoxia Inducible Factor
HRT	Hormone Replacement Therapy
hTERT	Human Telomerase Reverse Transcriptase
ICAM-1	InterCellular Adhesion Molecule-1
ICD	International Classification of Diseases
IDO	InDoleamine-pyrrole 2,3-diOxygenase
IFN	Interferon
IGF	Insulin-like Growth Factor
IGFBP	IGF Binding Protein
IgG/IgM	Immunoglobulin G/M
IL-1,2,3 ..	Interleukins 1,2,3 ….
iNKT cells	Invariant Natural Killer T-Cells
IQR	Interquartile Range
IR	Insulin Receptor
IRS-1	Insulin Receptor Substrate 1
JAK	Janus Kinase
LAP	Latency Associated Peptide
LD	Linkage Disequilibrium
LDH	Lactate DeHydrogenase
LKB1	Liver Kinase B1
lncRNA	Long Non-Coding RNA
LPS	LipoPolySaccharide

LRG1	Leucine Rich Alpha-2-Glycoprotein 1
LTBP	Latent TGF-β Binding Proteins
MAF	Minor Allele Frequency
MAPK ERK	Mitogen Activated Protein Kinase
MAPKK MAP2K MEK	Mitogen-Activated Protein Kinase Kinase
MAPKKK MAP3K	Mitogen-Activated Protein Kinase Kinase Kinase
MCP-1 CCL2	Monocyte Chemoattractant Protein 1
MCT 1,2 ..	Monocarboxylate Transporter 1,2 ……
MDSC	Myeloid-Derived Suppressor Cells
MET	Metabolic Equivalents
miRNA	microRNA
MMP	Matrix metalloprotease
MRI	Magnetic Resonance Imaging
mTOR	Mechanistic Target Of Rapamycin
NETs	Neutrophil Extracellular Traps
NHEJ	Non-Homologous End-Joining
NICE	National Institute for Health and Care Excellence
NF-κB	Nuclear Factor κ-light-chain-enhancer of activated B cells
NOS	Nitric Oxide Synthase
NPCS	Non-Post COVID-19 Syndrome (SARS-CoV-2 sufferers without PCS)
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
NSCLC	Non-Small Cell Lung Cancer
OD	Optical Density (absorbance)
OMIM	Online Mendelian Inheritance in Man
OR	Odds Ratio
orf	Open reading frame

OTR	Oxidative phosphorylation enzymes Transcriptional Response
PAI	Plasminogen Activator Inhibitor protein
PCA	Principal Component Analysis
PCS	Post-COVID-19 Syndrome (“Long COVID”)
PD-1	Programmed cell Death protein 1
PD-L1	Programmed cell Death protein 1 Ligand
PFK	PhosphoFructoKinase
PGE2	Prostaglandin E2
PI	Principal Investigator
PI3K	Phosphatidyl inositide 3-kinase
PIP3	Phosphatidyl Inositol 3,4,5 trisPhosphate
PKA	Protein Kinase A (cAMP dependent)
PKM2	Pyruvate Kinase (isozyme M2)
PLC	PhosphoLipase C
PlGF	Placental Growth Factor
PLOD2	Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase 2
PP2A	Protein Phosphatase 2a
PMBC	Post-menopausal breast cancer
PPAR	Peroxisome proliferator-activated receptor
PSA	Prostate Specific Antigen
PTEN	protein Phosphatase and TENsin homolog
PUMA	p53 Upregulated Modulator of Apoptosis
QC	Quality Control
RAGE	Receptor Advanced Glycation End product
RB	RetinoBlastoma associated
RBD	Receptor Binding Domain
RNA-Seq	RNA sequencing
ROS	Reactive Oxygen Species
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction

RYGB	Roux-en-Y Gastric Bypass surgery
S/CO	Signal/Cut Off
SARS	Severe Acute Respiratory Syndrome
SCoV2	SARS-causing Corona Virus 2
SE	Standard Error
SGB	Simson-Golabi-Behmel syndrome
SHOX	Short Stature Homeobox gene
SNPs	Single Nucleotide Polymorphisms
snRNP	small nuclear RiboNucleoProtein
SphK	Sphingosine Kinase
SREBP	Steroid Response Element Binding Protein
STAT	Signal Transducer and Activator of Transcription
T2DM	Type 2 Diabetes Mellitus
TBX2	T-Box Transcription Factor 2
TCA	TriCarboxylicAcid
TCGA	The Cancer Genome Atlas
TGFβ	Transforming Growth Factor β
TH	Tyrosine Hydroxylase
TIGAR	TP53-inducible glycolysis and apoptosis regulator
TLR	Toll-like Receptor
TNM	Tumour, Node and Metastasis (staging)
TNBC	Triple Negative Breast Cancer
TNF α	Tumor Necrosis Factor α
TRAF	TNF Receptor Associated Factor
TSC	Tuberous Sclerosis Complex
US	United States of America
UTR	UnTranslated Region
VEGF	Vascular Endothelial Growth Factor

WAT	White Adipose Tissue
WBFFM	Whole Body Fat-Free Mass
WBFM	Whole Body Fat Mass

Abstract

The pandemics of obesity, cancer and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) are major global healthcare challenges of this generation. This thesis investigates the interplay between anthropometric measurements, malignancy and, unanticipatedly, SARS-COV-2 infection.

Obesity-related cancers account for 40% of the cancer cases observed in the US, and obesity is set to overtake smoking as the most widespread modifiable risk factor for carcinogenesis. My initial literature review uses as a framework the *Hallmarks of Cancer* as defined by Hanahan and Weinberg. Each *Hallmark* is shown to be impacted by the physiological state of obesity, particularly through immunomodulation and facilitation of various elements of the metastatic pathway. Most work in this area was carried out in model organisms and cell lines. There are few studies on large prospective human cohorts, but large datasets are now becoming available and one is explored in depth in this thesis.

Previous human studies have shown a reliance on body mass index (BMI) as a surrogate for obesity. Studies in other fields have highlighted how overreliance on BMI can lead to misinterpretation. Now, the opportunity exists to dissect the factors behind BMI links with cancer, looking independently at whole body fat mass (WBFM) and whole body fat-free mass (WBFFM) more directly using bioelectrical impedance in a cohort of 500,000 individuals (the UK Biobank). Further, genomics approaches are used to extract signal from noise in large datasets, but as yet these are not commonly used in epidemiological studies. One such approach, principal component analysis (PCA), is employed here to tease apart the signal coming from the weight component of BMI and its association with cancer.

This work focusses on three cancers, breast (post-menopausal), prostate and colorectal cancers, all of which have been previously shown to associate with obesity. Increasing quintiles of BMI, WBFM and WBFFM are all associated with an increased risk of

post-menopausal breast cancer, with the highest quintile of WBFFM giving the highest odds ratio of the three (OR=1.60). PCA suggests that both increased WBFM and WBFFM may both confer increased risk of post menopausal breast cancer. By contrast, obesity is associated with a reduced risk of prostate cancer and the protective effect probably originates from WBFM. In colorectal cancer in men, increased BMI, WBFM and WBFFM all carry increased risk, although there is no obvious association in women. The merit of taking measures of fat mass and fat-free mass when investigating cancer risk is highlighted, raising interesting questions around the roles of adiposity and fat-free mass in relation to cancer. Perhaps genetic pleiotropy might help explain these results.

Anthropometric measures are consistently found to associate with differential cancer risk. Inherited genetic variants in tumour suppressors and proto-oncogenes also increase cancer risk. A further focus of this thesis is to investigate whether/how these two associations might be functionally linked. Through a novel approach, integrating inherited and somatic cancer genetic data, single nucleotide polymorphisms were identified which associate with different cancer types and differential gene expression in at least one tissue type (eSNPs). Potential associations of these SNPs with anthropometric traits and cancers were then explored.

A set of eSNPs related to three important cancer genes, FANCA, MAP3K1 and TP53, were identified as associating with both anthropometric measures and cancer risk. Notably, a previously unrecognised strong association was identified between the rs78378222[C] SNP in the 3'-UTR of TP53 and increased risk of developing non-melanomatous skin cancer (OR=1.36), brain malignancy (OR=3.12) together with increased standing height, fat-free mass and basal metabolic rate. This offers a novel genetic link between these anthropometric traits and cancer risk.

This eSNP analysis focused on a small subset of the genome leaving a large proportion unexplored. Therefore, this was extended using genome-wide association studies (GWAS). These identified additional pleiotropic haplotypes showing differential association with BMI, WBFM and WBFFM and the three cancers explored in the epidemiological study (post-menopausal breast cancer, prostate cancer and colorectal cancer). Twenty-two haplotypes associated with prostate cancer and anthropometric traits. Five haplotypes exhibited positive effects on both (i.e. increased prostate cancer risk and increased birth weight). eQTL analysis yielded SNPs within these haplotypes which influence expression of CD81, C11orf21, ASCL2, TH and IGF2. Eighteen haplotypes associated with both post-menopausal breast cancer and anthropometric measures. Three of these haplotypes were positively associated with post-menopausal breast cancer risk and anthropometric traits, and two of these also positively associated with measures of WBFFM. One SNP within these haplotypes, rs537626[C], which falls within the lncRNA CUPID1, has been validated as associated with breast cancer elsewhere. CUPID1 is a regulator of DNA damage repair and rs537626[C] may influence its function.

The SARS-COV-2 pandemic influenced the needs of the world and indeed this work. Redirection of resources and safety of the nation became paramount. I was amongst the first researchers to describe longitudinal immunological responses to SARS-COV-2 infections in two important patient groups (asymptomatic patients and those with Post-COVID-19 Syndrome). More directly related to the body of this thesis, as cancer and obesity both emerged as important risk factors for SARS-COV-2 outcome, common genetic roots became an area of curiosity. Interestingly, two SNPs associated with poor SARS-CoV-2 outcome and WBFFM (rs1265097[A] and rs3130279[A).

This thesis uses a variety of bioinformatics techniques to investigate links between anthropometric measures and cancer, adding several genetic loci to the literature worthy of further investigation.

Chapter 1: Obesity, a perfect storm for carcinogenesis

Chapter summary

Obesity-related cancers already account for 40% of the cancer cases observed in the US and obesity is overtaking smoking as the most widespread modifiable risk factor for carcinogenesis. In this review, we use as a framework the Hallmarks of Cancer, as defined by Hanahan and Weinberg, to delineate how obesity might influence the carcinogenic hallmarks in somatic cells. Topics discussed include the effects of obesity on (a) sustaining proliferative signalling; (b) evading growth suppressors; (c) resisting cell death; (d) enabling replicative immortality; (e) inducing angiogenesis; (f) activating invasion and metastasis; (g) reprogramming energy metabolism; (h) avoiding immune destruction, together with its effects on genome instability and tumour-promoting inflammation. We are presenting the current understanding in this evolving field and highlight some areas in need of further cross-disciplinary focus. For instance, the relative importance of the many potentially causative obesity-related factors are unclear for each type of malignancy. Even within a single tumour type, it is currently unknown whether one obesity-related factor consistently plays a predominant role, or if this varies between patients or, even in a single patient with time. Clarifying how the Hallmarks are affected by obesity may lead to novel treatments and prevention strategies for the increasing obese population. Indeed, differences in obese patients may provide preventative or treatment opportunities. The *obesity paradox* has been observed, for example, in melanoma and breast cancer patients treated with immunotherapy, where obese patients have a higher survival rate. So this raises the possibility of perhaps harnessing this inflammatory state for patient benefit and raises the question of the added value of patient stratification for emerging treatments.

1.1 Introduction

The pandemics of obesity and cancer are major worldwide healthcare challenges. By 2030 it is projected that up to 50% of the adult population will be obese[1]. In parallel, the incidence of cancer is rising, with the lifetime incidence now estimated at approximately 1 in 2[2] and obesity is set to be the main modifiable cause of cancer, overtaking smoking[3]. Obesity-related cancers already account for 40% of the cancer cases observed in the US[4].

Obesity is a largely preventable, multifactorial and complex disease[5]. The World Health Organisation (WHO) and the International Agency for Research on Cancer define obesity as 'abnormal or excessive fat accumulation that presents a risk to health'[6]. A simple measure commonly used to define obesity is body mass index, BMI, which attempts to normalise weight for an individual's height. It is defined as weight (kg) / [height (m)]^2 with values $\geq$30 kg/m^2 indicative of obesity and between 25-29.9 kg/m^2 indicating an 'overweight' state[7]. Waist circumference can also be used, with 94 cm in males or 80 cm in females indicating obesity. The cut off points given here, although commonly used, have been subject of much debate and various adaptations have been proposed, to adjust for ethnicity or other factors.

With barely any restrictions on food availability in the West, combined with an increasingly sedentary lifestyle, the prevalence of obesity is increasing rapidly. Obesity increases the risk for a number of chronic diseases including type 2 diabetes, ischaemic heart disease, stroke, sleep apnoea, cirrhosis, osteoarthritis, and markedly increases morbidity as well as mortality. And, although it tends to be defined through BMI (which does not describe fat distribution per se), how fat is deposited spatially has been noted to affect the adverse effects of obesity. Increased upper body fat has been associated with an increased risk of multiple conditions, including type 2 diabetes[8] and hypertension[9], whereas others have shown that increasing amounts of lower body fat are associated with a reduced risk of metabolic complications[10]. More detailed measures of fat mass have now been

devised to look at segmental fat mass, including bioelectrical impedance and dual energy X-ray absorptiometry. However, BMI as a surrogate for total fat mass is still the most widely used parameter in epidemiological studies.

Increasing BMI and obesity has been associated with an increased risk of cancer in a wide range of tissues[11] and absence of body fatness lowers risk of several cancers including: breast (post-menopausal), bowel, endometrium, oesophagus, ovary, liver, gastric cardia, gallbladder, pancreas, kidney, meningioma, multiple myeloma and thyroid[6]. (Fig. 1.1). Whilst this epidemiological data merely demonstrates an association, different cancers appear differentially affected, with causality in some cases backed by evidence from model systems. Supporting a causal relationship, we see that reduction of body mass following bariatric surgery is associated with a reduced cancer risk, particularly in obesity-related cancers[12]. And by 2035, it is estimated that ~40% of endometrial, >25% oesophageal, >20% renal and ~20% liver cancers globally will be attributable to having a high body mass index (≥25 kg/m^2)[13]. Further, obesity in general is associated with worse prognosis in patients with cancer, although the jury is still out in certain tumour types[14].

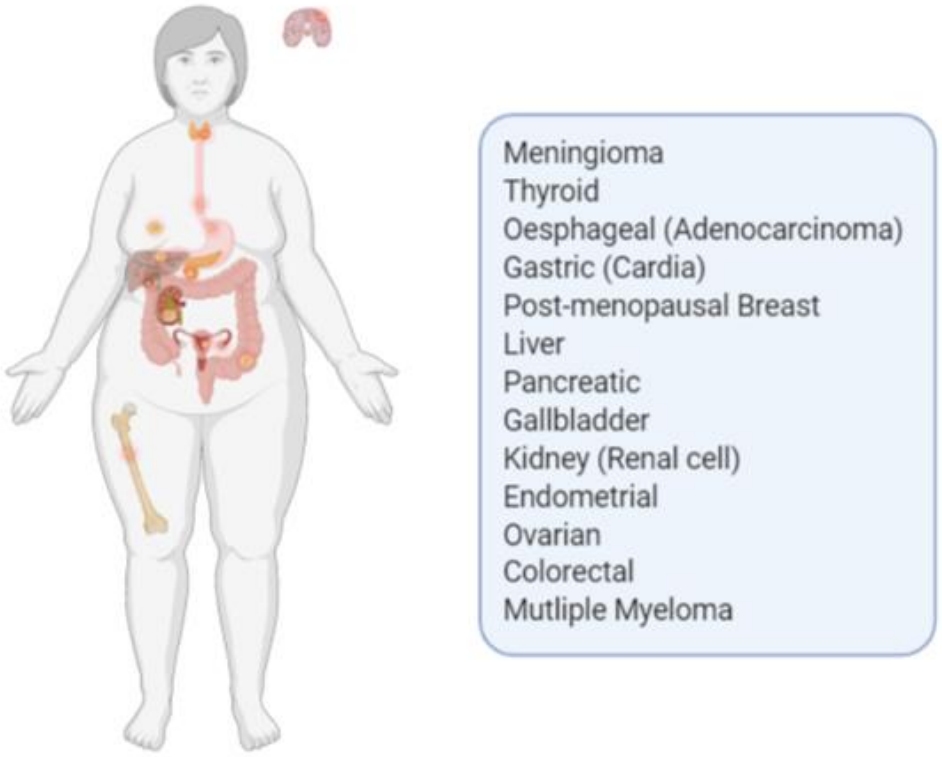


Figure 1.1: *Cancers associated with the obese state.*

1.1.1 Biochemistry and physiology in obesity

There is no doubt that obesity does change the body in a number of different ways, some of which can be linked to carcinogenesis. Some of the first discovered were the different hormonal profiles observed in obese patients. The levels of a number of growth factors are raised in obesity. Perhaps the most obvious of these is leptin, which (aside from its effect on appetite) activates several mitogenic pathways in a number of cell types (e.g. MAPK, JAK-STAT, PI3K/Akt pathways)[15,16]. Leptin is released from adipose tissue and hence could provide a direct link between obesity and increased susceptibility to cancer. In this it is opposed by the anti-mitogenic adipokine, adiponectin (also a product of adipose tissue)which acts through the cAMP signalling system[17].

Adipose tissue is a major source of oestrogen, a potent growth factor, and has been linked to the development of breast cancer[18]. Oestrogen classically binds to intracellular steroid receptor proteins and affects transcription, but it may also activate other growth promoting pathways such as the MAPK system[19].

Adipose tissue (and in particular adipose tissue resident macrophages and T cells) also typically releases proinflammatory cytokines, such as IL-6[18,20], which circulate and may alter the nature/activity of the cells of the immune system. This seems to be a continuous process in obesity, which is sometimes said to represent a state of chronic inflammation[21], which in turn can be associated with cancer development.

There are a number of other endocrine and paracrine factors originating outside the adipose tissue that have been suggested to promote abnormal levels of cell growth in obesity. Perhaps the most widely invoked are insulin and insulin-like growth factor (IGF-1)[22–24]. Both bind to tyrosine kinase receptors on the surface of a range of tissues, and trigger the MAPK and PI13K signalling cascades. Insulin levels do tend to rise in obesity, probably

due to a compensatory response to insulin intolerance – but its role in carcinogenesis in the obese state is debatable[25]. Conversely, IGF-1 does have mitogenic effects in vitro[23,26], but its levels do not consistently rise in obesity[27]. Workers who have proposed a role of IGF-1 in carcinogenesis typically attribute its effects not to (measurable) changes in total IGF-1 but to (putative) changes in free IGF-1 due to perhaps changes in levels of an IGF-1 binding proteins[28].

Other suggested mediators of carcinogenesis include growth factors such as TGFβ[29] and TNFα[30], whose levels may change in obesity, circulating proteases, metabolites such as cholesterol, fatty acids and kynurenine, some putatively produced by the gut microbiome, and gene modulators such as circulating miRNAs[31]. Finally, it has been suggested that obesity leads to epigenetic DNA modification (particularly cytosine methylation) and hence modulation of gene expression[32]. This can account for stability of changes in gene expression e.g. by a mother's diet affecting offspring, or infant obesity leading to adult obesity, but studies to date have been of relatively low resolution making it difficult to confirm epigenetic links to cancer in the context of obesity.

These various factors altered by obesity are summarised in Table 1.1. In the remainder of this chapter, emphasis will be shifted from consideration of how these compounds may affect the carcinogenesis through their impacts on the Hallmarks of Cancer as defined by *Hanahan and Weinberg*[33,34].

1.1.2 Hallmarks of cancer

In a seminal paper *Hallmarks of Cancer, by D. Hanahan & R.A. Weinberg*, defined the 'hallmarks of cancer', i.e. the properties that cells require for the development and maintenance of tumour growth[33]. These were listed as

1. Sustaining Proliferative Signalling
2. Evading Growth Suppressors
3. Resisting Cell Death
4. Enabling Replicative Immortality
5. Inducing Angiogenesis
6. Activating Invasion and Metastasis

These (particularly the first four) are likely to be interrelated – normal cell growth represents a balance between growth and antigrowth signals for example – but tumours often start by this balance being upset, for example, when a growth factor is produced in abnormal amounts.

In a subsequent paper by D. Hanahan & R.A. Weinberg, the authors added four more elements[34] (Fig. 1.2). These included the ability to avoid cell death due to recognition and destruction by the body's immune system, an area now widely exploited with the advent of PD-1/PDL-1 antagonists[35]. They noted the ability of inflammation to promote tumour growth, which they termed an enabling characteristic of cancers, and this area is of particular interest here since obesity is associated with a chronic inflammatory state. Two additional hallmarks they noted were the ability of tumour cells to reprogram their metabolism (typically towards a more glycolytic, rather than oxidative, direction) and a propensity for genetic instability and mutation.

This chapter is structured around individual ‘hallmarks’ and we find particular effectors (such as leptin, IL-6 and some hormones) affect multiple hallmarks and therefore appear in several sections. However, which obesity-induced factors predominate in tumorigenesis at different sites, or how local the effects of each factor are, is not well-defined. Thus I cover salient concepts and their basis in the literature and ouch on new therapeutic observations obtained when treating obese patients and the influence of obesity on prognosis. However, such is the broad nature of the subject area, it is beyond this chapter’s scope to discuss in depth every mechanism, experimental protocol and model system in all studies: these are available in the referenced material. This work aims to highlight general themes from what might be specific studies. Indeed, this field has its controversies, unanswered questions and new emerging hypotheses and it is not possible to debate all of these within this chapter but I aim to signpost the reader towards key areas to stimulate further work.

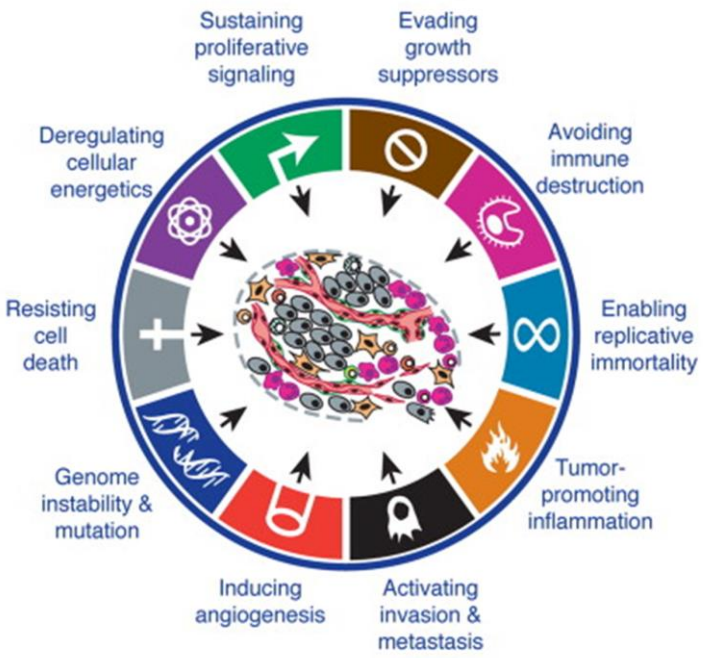


Figure 1.2: *Hallmarks of Cancer.*

Schematic showing enabling characteristics and emerging hallmarks of cancerous cells/ tissues. Diagram reproduced from *Hallmarks of cancer: the Next Generation* (Hanahan & Weinberg, 2011)

Factor	Direction in Obesity	Associated Hallmarks	Organism	Tissue	Exemplar Reference(s)
Adipokines					
Adiponectin	DOWN	II, V, VI	Human	Plasma	[36]
Leptin	UP	I, III, IV, V, IX, X	Human	Serum	[37]
Resistin	UP	IV, VI	Human	Serum	[38]
Dietary factors					
27-hydroxycholesterol	Postulated to be UP	IX	Human	Serum	[39]
Chymotrypsin	UP	II	Rat	Pancreas	[40]
Fatty acids (profiles need further elucidation)	UP	VII, X	Human	Serum	[41]
Glucose	UP	I, VII, X	Human	Plasma	[42]
Kynurenine	UP	I, III, V, IX	Human	Plasma	[43]
PCSK9	UP	VI	Human	Serum	[44]
Oncometabolites					
Lipid peroxides (e.g. measured by malondialdehyde)	UP	VII	Human	Serum	[45]
ROS	UP	VII, VIII, X	Rat	Colon/Kidney	[46]
Cytokines and cytokine-like proteins					
IFNγ	UP	VIII	Human	Serum	[47]
IL-1β	UP	VI, VII, VIII, IX	Human	Adipose	[48]
IL-6	UP	I, III, VI, VII, VIII, IX	Human	Plasma	[20]

IL-8	UP	VI	Human	Serum	[49]
IL-10	DOWN	IX	Human	Serum	[50]
IL-12	UP	VIII	Human	Serum	[47]
IL-18	UP	VII, VIII, IX	Human	Plasma/Adipose	[51]
MCP-1	UP	IX	Human	Serum	[52]
TNFα	UP	I, II, V, VI,VIII, IX	Human	Serum	[53]
TGFβ1	UP	II, III, V, VI, IX	Human	Plasma	[54]
Immune cells, markers and their products					
iNKT cells	DOWN	IX	Human	Omentum	[55]
B cells (exhausted)	UP	IX	Ex vivo	Adipose	[56]
CD8+ T cells	UP	VIII, IX	Mouse Models	Adipose	[57]
Eosinophils	DOWN	VIII, IX	Mouse Models	Adipose	[58]
Pro-inflammatory macrophages	UP	VIII	Mouse Models	Adipose	[59]
NETs	UP	VI	Human	Serum	[60]
PD1	UP	IX	Human	Serum	[61]
Th1 cells	UP	VIII	Mouse Models	Adipose	[62]
Th2 cells	DOWN	VIII	Mouse Models	Adipose	[63]
T_{reg} cells	DOWN	VIII, IX	Mouse Models	Adipose	[64]

Proteins of the fibrinolytic system					
PAI-1	UP	VI	Human	Adipose	[65]
Tissue factor	UP	V	Mouse Models	Brain, lung, kidney, heart, liver, and adipose tissues	[66]
Hormones					
Cholecystokinin	UP	I	Mouse Models	Pancreas	[67]
Insulin	UP	I, VI, VII, VIII, X	Human	Plasma	[68]
Oestrogen	UP	I, VII	Human	Plasma	[69]
Growth factors and their binding proteins					
FGF19	DOWN	Potentially X	Human	Serum	[70]
FGF21	UP	X	Human	Serum	[71]
FGF23	UP	I	Human	Serum	[72]
HGF	UP	V	Human	Serum	[73]
IGF-1	Controversial	I, III, VI			[27,74–76]
IGF-2	UP	I	Human	Serum	[77]
IGFBP-2	DOWN	I, III, VI	Human	Serum	[28]
IGFBP-3	Controversial	I			[78,79]
PlGF	UP	V, VIII	Human	Serum	[80]
VEGF-A	UP	III, V	Human	Serum	[81]

Extra-cellular matrix components and remodeling					
Collagen VI	UP	III, VI, VIII	Mouse Models	Adipose	[82]
Endotrophin	UP	III, VIII	Mouse Models	Adipose	[82]
MMPs E.g. MMP-2	UP	VI	Cell Lines		[83]
Transcriptional modulators, enzymes and other salient species					
AEBP1	UP	II	Mouse Models	White adipose tissue	[84]
A-FABP	UP	I, X	Human	Plasma	[85]
ANGPTL2	UP	V, VI	Humans/ Mouse Models	Serum	[86]
ATR	DOWN	II, VII	Mouse Models	Colon	[87]
Ceruloplasmin	UP	V	Human	Plasma & Adipose	[88]
Fas/FasL	UP	III	Human	Adipose	[89]
HIF1α	UP	III, V, X	Mouse Models	Adipose	[90]
PGE2	UP	III	Human	Adipose	[91]
PPARγ	DOWN	VIII, X	Mouse Models	Adipose	[92]
Certain exosomal miRNAs	UP	VI, X	Mouse Models		[93]

E.g. miR-122					
Certain free miRNAs E.g. miR-27a	UP	VI	Cell Lines		[94]
Telomerase	UP	IV	Cell Lines		[95]
TP53	Controversial	II, III, IV, VII, VIII, X			[96–99]

Table 1.1: *Influence of the obese state on potential carcinogenic factors.*
The particular hallmarks and enabling characteristics of cancer affected are indicated by Roman numerals. Each factor's potential local/systemic impact is shown. I = Sustaining Proliferative Signalling; II = Evading Growth Suppressors; III = Resisting Cell Death; IV = Enabling Replicative Immortality; V = Inducing Angiogenesis; VI = Activating Invasion and Metastasis; VII = Genome Instability and Mutation; VIII = Tumour-Promoting Inflammation; IX = Avoiding Immune Destruction; X = Reprogramming Energy Metabolism.

1.2 Sustaining Proliferative Signalling

A fundamental property of cancer cells is their ability to divide and proliferate uncontrollably. Normal cells are typically controlled via regulated release of growth signals, either produced within the same tissue or provided via the blood. These control the cell cycle and ensure that chromosomal replication and segregation, cell division, secretion of tissue matrix material and replication of organelles such as mitochondria are integrated to maintain normal tissue structure and function. These controls typically operate via receptor (membrane) kinases, such as the tyrosine kinase receptors for growth factors, and intracellular kinases such as mTOR and AMPK, which allow the cell to switch between the anabolic state (promoting cell growth) and the catabolic state[100,101].

In the obese state, blood leptin levels rise, as a result of the increased secretion from adipose tissue[37]. Circulating leptin helps to regulate appetite, insulin sensitivity and hormones such as thyroid stimulating hormone but also has potentially pro-oncogenic actions. Leptin is a growth factor, binding to a receptor tyrosine kinase (Ob-R/LepR), and activating the JAK/STAT, ERK and PI3K/Akt signaling pathways[15] (illustrated in Fig. 2). In vitro, leptin promotes cell cycle progression by inducing the production of cyclin D1 which drives cells past the G1/S checkpoint in various cancer cell lines (ZR-75-1, Hep3B, HepG2)[102,103]. However leptin is not the only factor at play in obesity tumour growth, highlighted by recent work in pancreatic adenocarcinoma mouse models showing accelerated tumorigenesis in obese mice deficient in leptin signaling with the possible involvement of cholecystokinin[67].

Another growth promoter, which has been associated with increased leptin production, is oestrogen, whose role in breast cancer is well established. Indeed, oestrogen receptor status (ER) is a determinant of breast treatment and influences prognosis. Oestrogen has also been linked to the development of endometrial and colorectal cancer in humans, although the colorectal relationship appears more complex[104,105]. In post-

menopausal women, hormone replacement therapy (oestrogen combined with progesterone) has been associated with reduced colorectal cancer risk[106]. However, recent work shows oestrogen (7-beta estradiol) alone increased mitosis but the combination of oestrogen and progesterone increases apoptosis in murine colon tumours[107]. Oestrogen also has been shown to influence proliferation in pancreatic cancer cell lines (HTB147, CFPAC, Panc-1), related to the oestrogen receptor subtype ratio[108] . Notably, a major source of oestrogen is adipose tissue, where it is made by the enzyme aromatase, and serum oestrogens are raised in obese humans[69]. And, there have also been claims that the gut microbiome may modulate oestrogen load which may be pertinent in obesity[109].

Obesity is often accompanied by hyperglycaemia[42], providing dividing cells with more fuel and affecting growth factor secretion in several different cancers[110,111]. Hyperglycaemia itself can affect cells in a number of ways related to carcinogenesis and can continue to affect cells, at least in vitro, even after return to euglycemic conditions[112,113]. Firstly, hyperglycaemia allows more fuel for dividing cells. In addition, hyperglycaemia seems to effect growth factor secretion in a number of different cancer cell lines[110,111]. For instance, hyperglycaemia causes epigenetic up-regulation of the Neuregulin 1 (NRG1) gene, a ligand for HER3 (part of the EGFR family) and this in turn can activate the PI3K/ERK1-2 pathways, driving cellular proliferation *in vitro*[113]. Further the obese state can induce mutations in oncogenic drivers of proliferation, such as HRas[114].

Insulin is released in response to increased blood glucose. Indeed, the hormone insulin, and its structural relation, insulin-like growth factor 1 (IGF-1), have been linked to both carcinogenesis and obesity[115]. Insulin is produced in the pancreas in response to raised glucose levels, while IGF-1 is largely produced by the liver in response to growth hormone, although it is made in all tissues, including adipocytes[116]. Insulin levels rise chronically in obesity[68], and type 2 diabetes, which is characterised by insulin resistance and thus high circulating levels of insulin, has been linked to an increased incidence of human obesity-

related cancers including colorectal, hepatic, pancreatic, breast and endometrial[117]. Recent work on the *scrib* gene, which is involved in setting cell polarity, reveals that raised *Drosophila* insulin-like peptides systemically decreases tumor-suppressive cell competition and thus initiating tumorigenesis. Although this work was carried out in *Drosophila*[118], *scrib* has homologs in mice and humans, and appears to be a tumour suppressor in hepatocellular carcinoma in these organisms[119].

IGF-1 has also been linked to carcinogenesis; however, in contrast to insulin, there seems to be no clear correlations between circulating total IGF-1 levels and obesity in humans[27]. However, circulating total IGF-1 does not always reflect IGF-1 tissue levels, which are governed by the levels of IGF binding proteins (e.g. IGFBP-2)[74]. IGFBP-2 has been shown to decrease in obesity and hyperinsulinemia, potentially increasing the levels of free IGF-1[28,75]. The majority of studies looking at IGF binding proteins have had relatively small sample sizes and indeed not always consistent results for the proteins IGFBP-1 and IGFBP-3, with some showing no significant[120] change in obesity and others suggesting an increase in IGFBP-3[121]. Larger prospective studies are needed here. Interestingly, there is cross talk between IGF-1 and oestrogen, for instance the IGF-1 promotor contains an oestrogen responsive element and IGF-1R is upregulated by oestrogen[122,123]. In breast tissue, IGF-1 might operate as a paracrine signal, being produced in the stromal cells but promoting tumour growth in the epithelium as demonstrated in *ex vivo* models[23]. IGF-1's close relative, IGF-2, may have a role in fat distribution and insulin resistance. IGF-2 has been reported to be raised in circulation in human obesity and reduced after weight loss[77] and has been shown to increase cellular proliferation through the insulin receptor at least in 3T3-like fibroblasts generated from mouse embryos nullizygous[124] for IGF-1 and IGF-1R. Further, upregulation of IGF-2 occurs in a number of human malignancies, including breast, oesophageal and colorectal cancer[125] . The involvement of both insulin-like growth factors in the link between obesity and cancer remains to be clarified.

High blood levels of the cytokine IL-6 have been reported in obese patients[20] and in patients with hepatocarcinoma[126]. IL-6 is released by macrophages associated with adipose tissue, and stimulates cell growth in hepatocellular carcinoma mouse models through STAT3, ERK, and JNK activation and alterations in AKT and mTOR[127]. Another putative growth factor is TNFα which can cause phosphorylation of JNK and p38. TNFα production, in adipocytes and immune cells, is stimulated in obesity by leptin in rodent models and humans[128]. Both are also important in the development of inflammation (see *Avoiding Immune Destruction* and *Tumour-Promoting Inflammation*).

Indeed other, less conventional, modulators can act as growth stimulators in obese individuals. These include adipose-derived fatty acid binding protein (A-FABP). A-FABP promotes the migration and proliferation of breast cancer cell lines in vitro (MCF-7, E0771) and thus may be a novel factor driving proliferation in obesity-induced tumours. A-FABP is elevated in the serum of obese humans and has been shown to promote proliferation and invasiveness by enhancing the stemness of mammary tumours[129]**.** Members of the fibroblast growth factor family, fibroblast growth factors (FGF) 19, 21 and 23 have been shown to be significantly changed in overweight and obese humans[70–72]. FGF23 is found in higher levels in overweight and obese individuals, has been shown to enhance proliferation and invasion in prostate cancer cell lines (LNCaP, PC3) and has been reported to be elevated in the serum of various cancer patients, including multiple myeloma[130]. The other members of the FGF family are discussed in *Reprogramming Energy Metabolism*. Another modulator is kynurenine, a tryptophan metabolite associated with excessive food intake. Kynurenine is important in the synthesis of nicotinamide and NAD^+ which is used to power proliferation by being a substrate used in glycolysis, the citric acid cycle, and oxidative phosphorylation[131] and kynurenine has been linked to both breast and colorectal cancers[132,133]. Kynurenine acts as a major endogenous activator of the aryl hydrocarbon receptor (AhR)[134] (which can increase adipocyte proliferation, and tumour progression). It is also noteworthy that AhR can

be activated by lipid peroxidation products which may be elevated in obesity (see *Genome Instability and Mutation)*.

Obesity can increase proliferation through a number of pathways, with perhaps leptin taking a central role. Further work is required to clarify the roles of other major protagonists, particularly looking at the nuances of role of oestrogen in different tumour types, thinking laterally about how to measure the true effects of IGF-1 in obesity and measuring markers of insulin signalling over time in human obesogenic tumour sites. It should also be noted obesity can influence proliferation of immune cells (see *Tumour-Promoting Inflammation*), and proliferation with associated tumour growth affects local tissue invasion (not completely independent of *Activating Invasion and Metastasis*).

1.3 Evading Growth Suppressors

In normal cells, growth represents a balance between growth promoting compounds (growth factors) and growth suppressors. Excessive proliferation, as in cancer, can represent a hyperactivation of growth factor receptors - either by excessive production of the growth factor itself, or mutation of a regulatory domain of the receptor - and/or inhibition of growth suppressors.

Growth factors tend to be compounds produced external to the tissue and acting (initially) on growth factor receptors, often at the cell membrane. Growth suppressors, such as 'tumour suppressor proteins', tend to be produced within the cells of the tissue and include TP53, AMP-dependent protein kinase, and retinoblastoma associated (RB) proteins. They may control the decision of cells to proliferate directly, at the level of the cell cycle, or indeed initiate cell destruction (apoptosis, senescence) if the potential damage of allowing growth is deemed too great.

Adiponectin is generally believed to be a growth suppressor released by adipocytes, with both endocrine and paracrine actions. Adiponectin is known to increase insulin sensitivity, helping to regulate peripheral glucose and fatty acid metabolism[135]. Besides being a metabolic regulator, adiponectin has anti-inflammatory and anti-oxidant activity. Normally, adiponectin is present in the blood at relatively high levels (10 ug/ml), 1000x higher than leptin and 10^6x higher than IL-6 or TNFα[53,136]. In human obesity, circulating adiponectin falls by up to 10-fold as visceral fat area increases[137] (the relationship between these is illustrated in Fig. 1.3). The reason for this is uncertain – it seems paradoxical that adiponectin, which is produced by adipocytes, should fall when the quantity of adipose tissue increases – but it might be due to suppression by the paracrine actions of TNFα[138] (whose local levels increase in obesity) and/or decreasing surface area/volume ratio in the larger adipocytes typical of obesity. Low levels of circulating adiponectin are observed in humans with endometrial cancer and are associated with increased breast cancer risk after adjustment for

BMI and age[139,140]. Adiponectin exerts its main actions through 7-transmembrane helix receptors (AdipoR1, AdipoR2)[141], ultimately inactivating the pro-proliferative mTOR[101] and can suppress cell proliferation in oestrogen receptor positive (MCF7), negative (MDA-MB 231) breast cancer cells as well as hepatocellular (HepG2, Huh7) and endometrial (HEC-1-A, RL95-2) cancer cell lines[142–145] and its drop in obesity may contribute to cancer cells evading growth suppression. Also adiponectin can suppress aromatase activity in adipocytes, lowering oestrogen production[143,146]. Perhaps unsurprisingly, leptin actively opposes the growth suppression activities of adiponectin not only by its actions at the cellular level but also at the molecular level. Whereas adiponectin causes the activation of AMPK via phosphorylation of thr172 by LKB1, leptin inactivates AMPK by phosphorylation at ser491 (via p70S6 kinase)[147]. However controversies exist here, with other work showing pro-proliferative effects of adiponectin in MCF-7 cells[148]. With most work carried out *in vitro,* translational relevance to humans is not yet fully understood and perhaps further laboratory work should be undertaken to explore competing claims.

Several tumour suppressor proteins are reduced in obesity. Normally, TP53 is unstable in the cell, but it can be stabilised by phosphorylation by AMPK, which is activated by adiponectin[149]. Thus it might be expected that TP53 concentration falls as adiponectin falls in obesity, and indeed this fall appears to be exacerbated by leptin in cellular models (Swan-71)[97]. In contrast, some authors also claim that TP53 levels increase in obesity, exacerbating the release of inflammatory cytokines[96]. Whichever effect is dominant, a potential influence of TP53 on tumorigenesis in obesity cannot be ignored. Also, polymorphisms of TP53 have been linked to obesity and lean body mass in mice and humans[98,99]. Another tumour suppressor, PTEN, attenuates the PI3K-Akt pathway[150] , the most commonly activated pathway in human cancers[151]. This tumour suppressing phosphatase causes the breakdown of PIP_3, and switches off the Akt pathway. Adipocyte enhancer-binding protein 1 (AEBP1) rises in obesity, causing a drop in PTEN activity and

consequent stimulation of adipocyte proliferation in mice[84] and this increase in adipocyte proliferation may lead to further cross-talk between adipocytes and tumour cells, for instance by leptin. Knockdown of AEBP1 reduces proliferation in various cell lines including, colorectal, gastric and glioblastoma cell lines (HT-29, MGC803, XN0422, U87MG, U138MG) and high expression of AEBP1 has been linked with poor prognosis in colorectal and other malignancies[152].

Members of the TGFβ family are complex players in tumour development[153]. These are produced in macrophages and other lymphoid cells and, their levels, particularly TGFβ1, rise in obese adipose tissue in mouse models as well as in the circulation in human obesity[154,54]. TGFβ can act as a tumour suppressor in early cancer development[155]. However, TGFβ can also promote stemness and metastasis, and this effect might be more dominant in obese individuals. Notably, leptin has been shown to increase protein levels of TGFβ1 in ER+ and ER- breast cancer cells (MCF7, MCF10AT1) and a neutralising antibody against TGFβ1 blocks leptin-mediated actions[156]. Further work here is needed to clarify the role of TGFβ in the different stages of cancer in obesity.

There may also be dietary effects on tumour suppression. High-fat diets seem to cause inhibition of the checkpoint protein through lysine homocysteinylation of ataxia-telangiectasia and Rad3-related protein (ATR) in mouse colorectal cancer models[87]. This provides a mechanism by which cells can avoid cell-cycle arrest, apoptosis and exhibit accelerated proliferation. High protein diets lead to overproduction of gut serine proteases (such as chymotrypsin). These may proteolyse another group of tumour suppressor proteins, (DCC, neogenin) which interact with extracellular matrix proteins, the netrins. This has been demonstrated in neuroblastoma and colorectal cell lines (SH-SY5Y, CaCo-2)[157] but how far this applies to sites outside the gut in humans is unclear. Another serine protease, proprotein convertase subtilisin/kexin type 9 (PCSK9), has been shown to be raised in obese human

serum[44] and is also linked to melanoma metastasis in mice through its role in maintaining high circulating cholesterol levels and likely activation of the TNFα pathway[158].

The drop in adiponectin might be the key factor in evading growth suppression, but this needs more work on a wider range of tumours as it appears not to be universal. The role of the TGFβ family most certainly needs further clarification and perhaps further work on the other tumour suppressor polymorphisms, like those related to TP53, might yield new pathways linking this hallmark with obesity. It should be noted that it is not always easy to separate the positive effects on growth suppression with negative effects on proliferation but I have taken classically defined growth suppressors, such as PTEN, as exemplars.

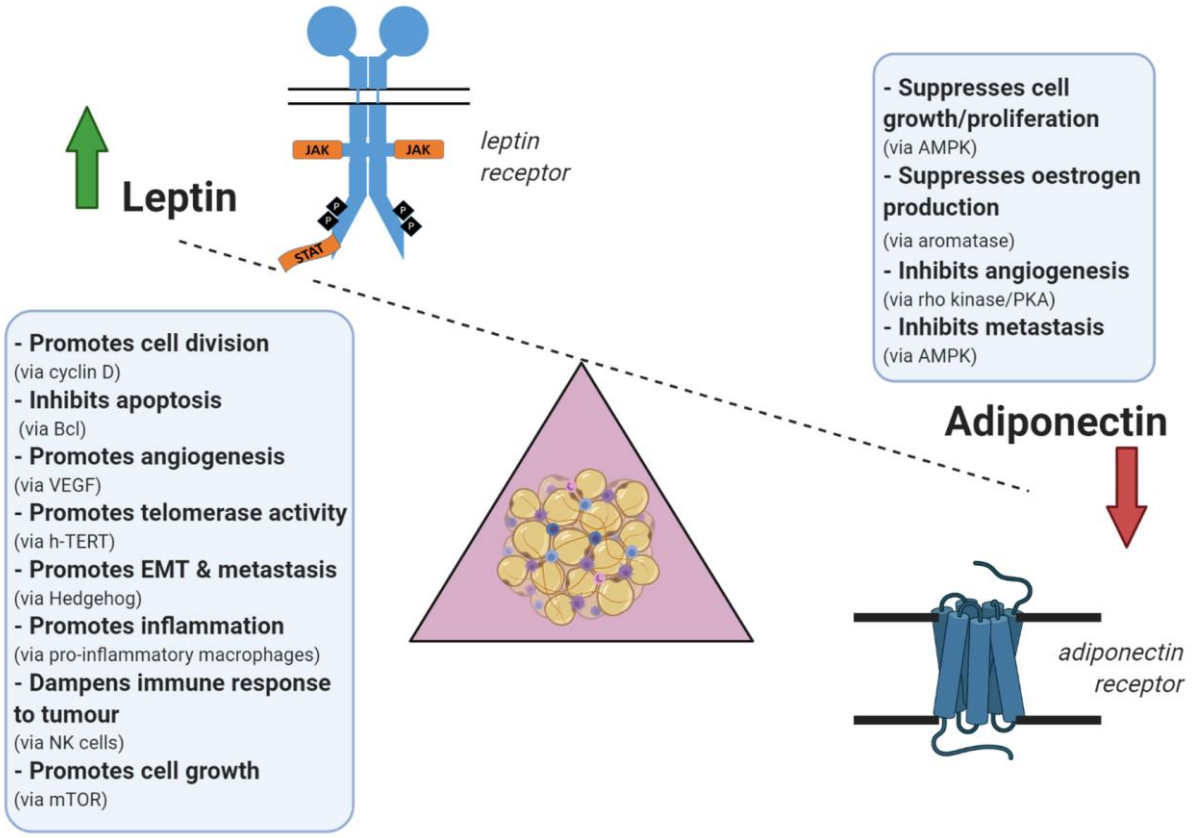


Figure 1.3: *Leptin/adiponectin balance in obesity.*

Leptin and adiponectin are two adipokines produced by adipose tissue. Leptin levels rise in obesity, while adiponectin falls. Leptin binds to a growth factor receptor at the cell surface, and promotes cell proliferation/ opposes cell death by a variety of mechanisms. These include promoting cell growth (via stimulation of mTOR), angiogenesis and metastasis, while at the same time inhibiting anti tumour mechanisms such as apoptosis and immune surveillance. Overall, therefore, the actions of leptin are pro-tumorigenic. Adiponectin, in contrast, binds to a G-protein coupled receptor at the cell surface, and broadly opposes these actions of leptin, by activating kinases such as AMP kinase (which decreases mTOR activity) and protein kinase A (which decreases angiogenesis). It therefore opposes tumorigenesis. Normally the actions of these two hormones are balanced, to maintain cell proliferation at an appropriate rate, but in obesity leptin concentrations rise, and adiponectin falls, promoting a tumorigenic environment.

1.4 Resisting Cell Death

Eukaryotic cells are programmed to die, either after a particular number of divisions or, sacrificially, to avoid damaging the rest of the organism. Avoiding death is a key feature of cancer cells, whether this be by avoiding programmed cell death pathways (telomere shortening, apoptosis, autophagy and pyroptosis), surviving cellular injury and premature cell death (necrosis), avoiding immune attack from neutrophil extracellular traps (NETosis) and enduring lipid peroxidation and oxidative stress (iron-dependent) which can trigger reactive oxygen species (ROS)-reliant cell death (ferroptosis).

The hormone most associated with obesity, leptin, has an inhibitory effect on apoptosis, typically by increasing the expression of the anti-apoptotic proteins Bcl2 (B-Cell CLL/Lymphoma 2) and survivin, reducing production of pro-apoptotic proteins such as Bax and TRAF and possibly reducing caspase activation. This phenomenon has been widely studied *in vitro*, including in breast cancer[159,160,161], ovarian cancer[162], prostate cancer[163], hepatocarcinoma[103] and neuroblastoma[164]. It has been reported that IGF-1, another growth promoter potentially raised in obesity, also opposes apoptosis in a similar manner, by up regulating expression of antiapoptotic proteins through the Bcl system[164,165,166] in cellular models, possibly by decreasing (proapoptotic) TP53[96]. TP53 is typically kept low by ubiquitinylation/proteolysis, but it can be stabilised by phosphorylation, particularly by AMPK. However, inflammatory factors, such as PGE2 - which is at least raised in adipose tissue in obesity[91] - inhibit AMPK, potentially leading to the destruction of TP53 and evasion of apoptosis[96]. Obesity also can influence apoptosis by upregulating collagen VI, which is produced during adipose tissue fibrosis as its cleavage product, endotrophin, has an anti-apoptotic effect in murine mammary tumours[167].

Although resisting cell death is important for the tumour itself, cell death outside the tumour itself may promote resistance to death amongst tumour cells. In obesity, adipocyte hypertrophy is common and adipocytes more frequently undergo apoptosis[89] or necrosis[168] .

This process attracts macrophages, and the dying adipocytes and their surrounding crown-like structures are typical of various fat deposits in human obesity, including subcutaneous and visceral fat[168169] . These dying cells release cytokines, which can drive cell division and further inflammation. IL-6 in particular has anti-apoptotic actions, through up-regulation of Mcl-1 (a member of the Bcl-2 family) in a number of cancer cell lines[170,171] (e.g. C33A, AGS), including prevention of apoptosis in hepatoma cells under the influence of TGFβ[172]. Furthermore, in obesity hypoxia is said to occur in adipose tissue secondary to a number of mechanisms, including cardiac output not keeping pace with adipose expansion and aberrantly reduced postprandial blood flow[173] . Hypoxia can drives tumorigenesis in number of ways[174],[175] ,including via hypoxia inducible factor 1α (HIF1α) and leading to the production of VEGF-A[176] which has anti-apoptotic actions in adenocarcinoma models[177]. Hypoxia can have local and systemic effects and thus can directly influence hypoxic cells or others further afield. Such effects are further discussed in *Inducing Angiogenesis* and *Reprogramming Energy Metabolism*.

Autophagy is a cellular process which removes misfolded or aggregated proteins and damaged organelles. The main role of autophagy is to remove misfolded or aggregated proteins and damaged organelles, it can recycle raw materials in extreme stress, and it is involved in eliminating intracellular pathogens. It is dependent on activation of AMPK in low energy conditions and will result in programmed cell death[178]. Autophagy is thought to be suppressed in obesity and has been demonstrated to be reduced in the livers of obese murine models[179,180, 181,182] and this too may reduce cell death in tumours. Further, adiponectin has been shown to induce cytotoxic autophagy in breast cancer cell lines[183]. However, the situation is not unequivocal as other work has shown that autophagosomes can accumulate in obesity[184–186]. The interplay between obesity, autophagy and cancer is not well understood - but inhibition of autophagy via AMPK, like inhibition of apoptosis in obesity, may reduce cell death and allow cancer cells to proliferate.

Another route to programmed cell death is pyroptosis, which differs from apoptosis in (a) involvement of the inflammasome (b) dependence on caspase 1 activity (rather than caspases 3,6,8 which are activated in apoptosis) and (c) being a pro-inflammatory process, releasing the cytokines IL1β and IL18[187]. Pyroptosis is normally involved in defence against invading bacteria: the cell membrane develops pores, the cytoplasm swells and the cell eventually ruptures, removing intracellular niches for growth of these bacteria. Pyroptosis has been shown to be upregulated in adipose tissue and is a source of adipocyte death in murine obesity[188] and this may be linked to the formation of the crown-like structures, and consequent chronic inflammation. Interestingly, the anti-diabetic drug metformin has been shown to induce pyroptosis in oesophageal cancer cell lines (KYSE510, KYSE140) and so might be of benefit in these obesity-related cancers in the clinic[189].

Ferroptosis is an iron-dependent form of cell death, where iron-containing lipoxygenase enzymes oxidise membrane polyunsaturated fatty acids, forming reactive oxygen species which are then broken down by glutathione peroxidase 4 (GPX4). Leptin has been shown to cause downregulation of glutathione peroxidase 4 (GPX4) and, in combination with iron-overload (which is associated with T2DM common in obesity) this can cause ferroptosis of podocytes which could be a key factor in diabetic nephropathy[190]. However, it has been shown that some cancer cells can survive in the absence of GPX4[191]. This could be due to the increased activity of other endogenous antioxidant pathways or low levels of oxidisable polyunsaturated fatty acids, but the actual cause of resistance is not well-established[192]. It may represent another way for cancer cells to avoid cell death. Currently the induction of ferroptosis, like that of pyroptosis, is being investigated for cancer therapy, and understanding how obesity affects this in the context of cancer would be an interesting area for further work.

Of note, several of the factors influencing apoptosis in obesity may have pro- or anti-apoptotic effects dependent on context. Indeed, a number of competing processes seem to

be involved. For example, kynurenine and its metabolites can induce apoptosis of both NK cells and T- cells, (potentially allowing cancer cells to avoid immune destruction) while inhibiting apoptosis in colorectal cancer cells (HT-29)[193,194,195]. Also, exosomal transfer of small non-coding RNAs from adipocytes may regulate biological processes through post-transcriptional gene silencing. miRNAs, particularly miR-21, from cancer-associated adipocytes, can inhibit apoptosis in ovarian cancer cell lines (SKOV3, OVCA432) by targeting apoptotic protease-activating factor 1 (APAF-1)[196].

Overall the effects of obesity on the mechanisms of avoiding cell death are mixed, with some factors dominating in some scenarios, and other effects less clearly understood. Perhaps, pro-death signals may simply be drowned out in some circumstances, when one considers the complex interaction between potentially anti-apoptotic factors (such as VEGF-A and IL-6) and the soup of growth factors circulating in obesity.

1.5 Enabling Replicative Immortality

Whilst normal cells will undergo a finite number of divisions before senescence, cancer cells acquire unlimited replicative potential[34]. Multiple lines of evidence indicate that telomeres protecting the ends of chromosomes are centrally involved in the capability for unlimited proliferation. Telomeric repeats shorten with each division, finally exposing non-telomeric DNA, precipitating end-to-end chromosomal fusions; causing unstable dicentric chromosomes, and limiting the ability to generate viable cells. Reactivation of telomerase has been identified as a major route by which cancer cells prevent telomeric erosion[197] generating longer telomeres and promoting extended proliferation/immortality.

The influence of obesity on this hallmark is contentious, although some adipokines promote telomerase activity in specific cases. Leptin upregulates telomerase activity as well as the mRNA and protein levels of the catalytic subunit of reverse transcriptase of human telomerase (hTERT) in breast cancer cells (MCF-7)[95] . Plasma leptin levels are correlated with tumour hTERT levels in human breast cancer ($r = 0.484$, $p < 0.01$)[198] and leptin mRNA in human hepatocellular carcinoma is also strongly correlated with hTERT ($r = 0.79$, $p < 0.01$)[199]. Resistin, another raised adipokine in obesity[38] , has also been shown to increase the expression of hTERT in gastric cancer cells (AGS)[200]. IGF-1 has also been implicated in telomere lengthening, although conflicts in the literature exist when contrasting cellular models with the human syndrome of acromegaly[201,202]. It has also been claimed that the tumour suppressor TP53, which may be lowered in obesity (see *Evading Growth Suppressors*), downregulates the expression of hTERT in a cervical cancer cell line (SiHa)[203] in addition to its pro-apoptotic effects. However, in the main, telomeres appear to be shorter in obese individuals[204], perhaps partially due to the influence of insulin[205], and shorter telomeres predispose cells to the *Enabling Characteristic: Genome Instability and Mutation*.

It is known that in some cancers there are alternative methods of telomere lengthening operative during the multi-stage carcinogenic process. One cause of alternative telomere lengthening (ALT) is aberrant DNA damage repair, and altered DNA damage repair pathways have been reported in obesity[206]. However at present, little appears to be known about ALT in obesogenic carcinogenesis.

At present this the link between replicative immortality and the obese state is not well-established. Leptin effects appear the most studied but literature is sparse with little published on other adipokines. Alternative telomere lengthening in obesity seems an inviting avenue for further work.

1.6 Inducing Angiogenesis

Tumours, like normal cells, need fuel and building blocks for growth and proliferation. Blood vessels provide an important conduit for these nutrients. Thus, tumours need to have a vascular supply and, as they grow, they must generate neovasculature through angiogenesis. There are a number of angiogenic factors which are raised in obesity, and could contribute to rise in cancer growth in this condition. Adipocytes and their stromal cells alter their adipokine profile in obesity, especially whilst expanding and dividing, producing a plethora of pro-angiogenic factors including leptin, ceruloplasmin, angiopoietin-like proteins, hepatocyte growth factor (HGF), as well as proteins of the vascular endothelial growth factor (VEGF) family, including placental growth factor (PlGF)[207,208,81].

The primary stimulant for angiogenesis is the growth factor VEGF[209], although in some tumours, placental growth factor (PlGF) may be implicated[209,210]. Binding of the VEGF to its receptor activates the intracellular VEGF cellular cascade, ultimately stimulating endothelial cell proliferation, migration and vascular permeability[211]. Many proangiogenic factors increase VEGF secretion, although other mechanisms may also be active. Leptin has been reported to increase secretion of VEGF-A and -C across a range of different tissues, including adipose tissue in mice[212,213,214] . Indeed, elevated serum VEGF-A has been described in obese humans[81]. Leptin also increases receptor expression (VEGFR-2) and its transphosphorylation in cellular models (HUVEC)[215] and in mammary tumours in mice[216].

Placental growth factor (PlGF) may also stimulate angiogenesis in obesity in a similar way to VEGF, although evidence is limited. PlGF is a member of the vascular endothelial growth factor (VEGF) family and stimulates VEGFR1 to increase angiogenesis as well as being linked to influencing the tumour microenvironment and tumour progression[210,217]. PlGF levels have been shown to increase in obesity in children, various cancer mouse models and in cancer patients[80,218] and correlates with poor survival rates in colorectal cancer[219].

Another pro-angiogenic factor, Tissue Factor (TF) is overexpressed in adipose tissue in obese mouse models[66] and plasma TF is reduced by weight loss in obese humans[220]. It has been proposed increased TF expression by endothelial cells enhances angiogenesis, and increased TF expression in tumour cells (Meth-A) and human pancreatic tumours was associated with increased VEGF expression[221,222]. Furthermore, an alternatively spliced isoform of tissue factor promotes angiogenesis by an alternative route[223].

Angiopoietin-like proteins (ANGPTLs) are other angiogenic and metabolism-modifying factors that can be upregulated in obesity. Of particular note is ANGPTL2 which is abundantly expressed by adipose tissue and whose serum levels are positively correlated with body mass index. ANGPTL2 has been linked with obesity-related cancers in humans, for instance colorectal cancer[224] and has been shown to increase angiogenesis by increasing VEGF-A and Ang II in osteosarcoma mouse models[225] . Also ANGPTL2 has been shown to induce vasculogenesis directly in endothelial colony forming cells, activating c-Jun NH2-terminal kinase (JNK) and increasing membrane type 1 matrix metalloproteinase (MT1-MMP)[226] .

TGFβ and TNFα are elevated in the adipose tissue of obese mice and may have paracrine effects on angiogenesis. If participants in the TGFβ signalling pathway increase, clinical conditions with excess angiogenesis can manifest, such as hereditary haemorrhagic telangiectasia (HHT) and Osler-Rendu-Weber syndrome, suggesting that TGFβ has an pro-angiogenic action[227,]. TNFα induces and modulates angiogenesis in a number of ways, including induction of VEGF-D and LRG1[228]. It is unclear, however, how far these effects are dependent on intermediate cytokine production, e.g. IL-8, which is elevated in the serum of obese humans[49].

The reduced adipocyte production of adiponectin in obesity may influence angiogenesis. Adiponectin has been shown to inhibit angiogenesis through several

pathways, including its effects on the Rho kinase/IFN-inducible protein 10, matrix metalloproteinase 9 and the MAPK and cAMP-PKA pathways[229,230], and so its reduction may promote angiogenesis. Adipose tissue also synthesises hepatocyte growth factor (HGF). HGF is elevated in obese individuals[73], activating a number of downstream pathways in carcinogenesis, including angiogenesis and has been linked to hepatocellular carcinoma[231].

A circulating factor of interest in angiogenesis is nitric oxide (NO) which is synthesised from L-arginine by the nitric oxide synthases (nNOS, iNOS eNOS)[232]. NO acts acutely on oxygen supply to tissues by increasing blood supply, by relaxing smooth muscle and hence opening the blood vessels. In the longer term, NO appears to be involved in angiogenesis directly[233] and enzymes involved in its production have been shown to be upregulated in tumour samples[234]. Conversely, inhibition of NOS will restrict tumour growth[235]. VEGF-A can also stimulate NO production[236] leading to a positive feedback arrangement. Whilst it is clear that NO is involved in angiogenesis in tumours, its relation to obesity is equivocal, with some studies claiming to show blood NO concentration positively correlated to fat mass and BMI[53,237] and others showing an opposite relationship[238]. Interestingly, exhaled NO (presumably an indicator of circulating NO) is reduced in obesity[239].

Other potentially unexpected factors might also have a role in promoting angiogenesis. Ceruloplasmin, a serum ferroxidase involved in copper transport is raised in obese individuals[88,240], and induces angiogenesis probably by stimulating VEGF-A production via the HIF1α signalling pathway. Also, the aforementioned AhR can be activated in by kynurenine. Like HIF1α, when activated the AhR moves to the nucleus and activates genes associated with hypoxia, including VEGF-A[241].

Hypoxia itself likely links many of these angiogenic mediators and is a well-established and key regulatory factor in tumour growth[174],[175] through its effects on gene

expression and treatment blunting[175]. Hypoxia promotes the level of hypoxia inducible factor (HIF), particularly HIF1α, which acts as a transcription factor. In obesity, hypoxic areas of adipose tissue are common as growth of adipocytes outstrips blood supply, and so the HIF1α pathway is activated [90,242,243]. This markedly increases the production of adipocytokines[244], and thus in hypoxia, cellular produced factors are directed, under the influence of HIF1α, towards promoting angiogenesis with the aim of increasing cellular oxygen tension. These include many of the factors mentioned above, including leptin, VEGF, HGF (produced in adipocytes), PIGF[245], ceruloplasmin[246], IL-8[247], ANGPTL2 and ANGPTL4[248].

Obesity seems to have a marked impact on angiogenesis, with many factors ultimately acting through VEGF-A. Tissue hypoxia is a major stimulant of angiogenesis and may influence proceedings whether it occurs in subcutaneous deposits of adipose tissue or in organs where ectopic fat deposition occurs (e.g. pancreas). More investigation is needed as to the roles of ANGPTLs and the effects of ceruloplasmin on angiogenesis in human obesity-related tumours.

1.7 Activating Invasion and Metastasis

An important, and pathological feature of cancer cells is their ability to spread into surrounding and distant tissues i.e. to 'metastasise'. Langley and Fidler have described the various steps in this process in the 'seed soil' hypothesis[249]. Characteristics of the cancer cell itself, "the seed", including disruption of cell-cell adhesion proteins and de-differentiation, play important roles. However, other features "escape routes", "avoiding surveillance" and the final metastatic niche, "the soil", all appear to play a role. Within obese individuals, a number of factors alter these key elements, illustrated in Fig. 1.4.

A familiar pair of adipokines, leptin and adiponectin, influence invasion and metastasis. To allow this process, cells tend to undergo a dedifferentiation process which some have termed epithelial-mesenchymal transition (EMT). The Hedgehog developmental gene complex is involved in this process. Leptin activates the Hedgehog developmental complex[250], and EMT in breast cancer cell lines (MCF-7, SK-BR-3)[251]. Leptin also stimulates the production of IL-6 and TGFβ which have paracrine activity reinforcing EMT[252,], whilst promoting secretion of extracellular matrix remodellers, such as matrix metalloprotease 2 (MMP2) in mammary cells (MCF10A)[83]. Reducing leptin availability to implanted ER+ breast cancer cells (MCF7) significantly decreased the number of metastatic lesions in mice[253]. Resistin, also raised in obesity, induced EMT in breast cancer and epithelial cells (MCF7, MDA-MB-231, MCF-10A)[254]. Conversely, adiponectin can reduce leptin-induced metastasis typically by activating AMPK[255]. Thus, the reduction of adiponectin seen in obesity can favour metastasis.

Hyperinsulinemia is commonly associated with obesity. Insulin can increase the migration ability of colorectal cancer cells (HCT-116) via PI3K/Akt and MAPK[256], while downregulation of the insulin receptor reduces lung metastasis in a mouse model with the highly metastatic breast cancer cell line, LCC6[257]. Further, hyperinsulinemia appears to

influence response to cancer treatments. The non-small cell lung cancer cell line (HCC4006) harbours a EGFR mutation normally rendering it sensitive to the tyrosine kinase inhibitor gefitinib. Co-treating cells with gefitinib and insulin decreased this sensitivity, increasing cellular proliferation and survival. This treatment resistance was attributed to insulin activating the PI3K/AKT pathway[258] and this perhaps plays a role in treatment failure in obese patients. It does also suggest insulin provides a PI3K-activating stimulus without having a specific mutation in this important pathway.

Importantly, invasion, migration and the formation of metastases are complex processes involving tumour and other cells. For example, adipocytes in proximity to tumour cells promote invasiveness in a number of different cell lines (e.g. Pt-93, SW480) and mouse models[259,260]. In breast cancer, adipose-derived stromal cells (ASCs) can cross feed tumour cells by releasing fatty acids, proteases which degrade the extracellular matrix, and by releasing cytokines and miRNAs. IL-6, IL-8 and leptin secreted by ASCs promote invasion by breast cancer cells (MCF7, MDA-MB-231)[261,262]. ASCs also show increased expression of IGF-1 in obesity, enhancing local invasion in mice[263] possibly through its effect on anoikis (elaborated later in this section). ASCs influence the extracellular matrix, depositing denser and stiffer matrices in obesity mouse models. This stimulates growth and invasion through local increased mechanosignaling and increased YAP/TAZ transcription factor activity and potentially further afield through increased secretion of stromal cell–derived factor 1 (SDF-1)[264]. Adipocytes also produce a number of other metastatic-enabling compounds including MMPs, β-hydroxybutyrate, plasminogen activator inhibitor-1 (PAI-1), collagen VI, TNFα and IL-1β[48,65,82,265,266]. And, A-FABP from omental adipocytes appears to promote ovarian cancer metastasis in model systems (discussed further in *Reprogramming Energy Metabolism*)[267] whilst CCL7 from periprostatic adipocytes increases migration of tumour cells (PC-3, Du-145) *in vitro.* Thus local cross-talk between tumour cells, adipocytes and their related cells appears important.

Adipose tissue also might have a role at a distance. Adipose tissue is a major source of circulating exosomal miRNAs, which can control gene expression in distant tissues, at least in mice[268]. These have been proposed to drive tumour progression, invasion and metastasis, through inducing fatty acid oxidation in melanocytes, Wnt signalling in mammary tumours, and cell cycle progression in hepatocarcinoma[269–271]. In adipocyte/ hepatic cell cocultures, an increase in the free miRNA miR-27a was linked to tumour development[94].

The invasion route for transformed cells is enhanced through angiogenesis and lymphangiogenesis. ANGPTL2, for example, is linked with promoting metastasis in human lung cancer and in mouse breast to lung metastatic models[272]. As discussed in the previous section, other factors that increase angiogenesis are present at higher levels in obesity. Once detached from the parent tissue, normal cells typically self-destruct (anoikis)[273] but cancer cells can develop anoikis resistance. Triggering of anoikis involves inhibition of mTOR via AMPK. In breast cancer cells (MCF7, ZR-75, EFM-19), anoikis has been shown to be opposed by IGF-1 (binding to IGF1R)[274]. Since AMPK activity is reduced in obesity, at least in human obese visceral fat[275], and IGF-1 may locally rise, obesity may oppose anoikis and promote metastasis. Indeed IGF-1 has been linked to several stages of invasion and metastasis[276].

The escape route for transformed cells is enhanced through angiogenesis and lymphangiogenesis. Several factors that increase angiogenesis are present at higher levels in obesity, including VEGF, angiopoietin-like proteins, placental growth factor, hepatocyte growth factor and TGF-β (discussed in detail above). Further, these pro-angiogenic factors may also influence responsiveness to other pro-migratory stimuli. For example, the overexpression of ANGPTL2 may play a role in encouraging metastasis, particularly in breast cancer models where increased ANGPTL2 expression appears to increase responsiveness to chemokine CXCL12 and encourage metastasis to bone[277]. ANGPTL2 also promotes metastasis in human lung cancer and in mouse breast to lung metastatic

models[272]. TNFα enhances lymphangiogenesis, which is particularly important as the lymphatic system is an major conduit for tumour metastasis[228] possibly acting through the induction of leucine-rich-α2-glycoprotein1 (LRG1)[278] which has been recently implicated in oncogenesis and promotes both EMT and angiogenesis[279]. Insulin also appears to have a positive role on lymphangiogenesis, since inhibition of its receptor IRS1 reduces lymph vessel formation[280].

IL-18, as well as suppressing antitumour activity of CD70 lymphocytes, discussed in *Avoiding Immune Destruction*, is also linked to increased metastatic potential, through the stimulation of CD44 expression on tissues[281]. CD44 is the principal receptor for transmembrane hyaluronate[282] and, by its interaction with the extracellular matrix, has been implicated in invasion/metastasis in several cancers[283]. In addition, IL-18 promotes inflammation and angiogenesis in some settings [284–286]. Other changes in the extracellular matrix in obesity have been linked to the invasion of cancer cells, with stiffer aligned fibres producing high interstitial pressures and more hypoxic areas of the tumour[287].

Despite these various proangiogenic factors, growth of tumours in obese individuals tends to outstrip the blood supply and they become hypoxic. Obesity tends to accelerate tumour progression as shown in a breast cancer mouse model[288]. This leads to the stabilisation of HIF1α, which also activates EMT, through a number of different effects, including the deubiquitination of Snail and the direct transcriptional induction of TWIST[289–293]. It seems likely that the aryl hydrocarbon receptor, which works similarly to HIF1α, promotes metastasis, since its inhibition has been shown to reduce proliferation, migration, and invasion of TNBC cells[294]. In obesity, the activity of this enzyme is increased by increased levels of kynurenine, giving another potential pathway towards anoikis resistance and metastasis.

Obesity can also affect the target ‘soil’ of the metastatic cells. IL-6 has been linked with preparing metastatic niches by inducing CCL5 expression in lymphatic endothelial cells within pre-metastatic sites and creating chemotactic gradients to recruit CCR5-positive cancer cells into the lymph nodes and lungs in model systems of triple-negative breast cancer metastasis[295]. Obesity-associated lung neutrophilia appears to enhance breast cancer to lung metastases in mouse models[296]. Neutrophils also produce neutrophil extracellular traps (NETs), web-like structures of chromatin, which are proinflammatory. Increased plasmatic NETs have been seen in patients with morbid obesity[60]. NETs have been linked with induction of metastasis and promoting tumour growth, particularly in breast cancer models[297,298]. A summary of how obesity can affect invasion and metastasis can be seen in Fig. 3.

It appears the obese state can prepare cells for all aspects of the metastatic process. This includes stimulating cell proliferation and de-differentiation, through release of tumour cells from tissues (including avoidance of anoikis), providing vessels for their dispersal and providing a final niche for their growth. As obesity is related to aggressive disease, metastasis and short survival in a number of cancer types, we suggest prioritisation of efforts to understand metastatic niche preparation in obesity as this will perhaps yield the greatest benefit to patients.

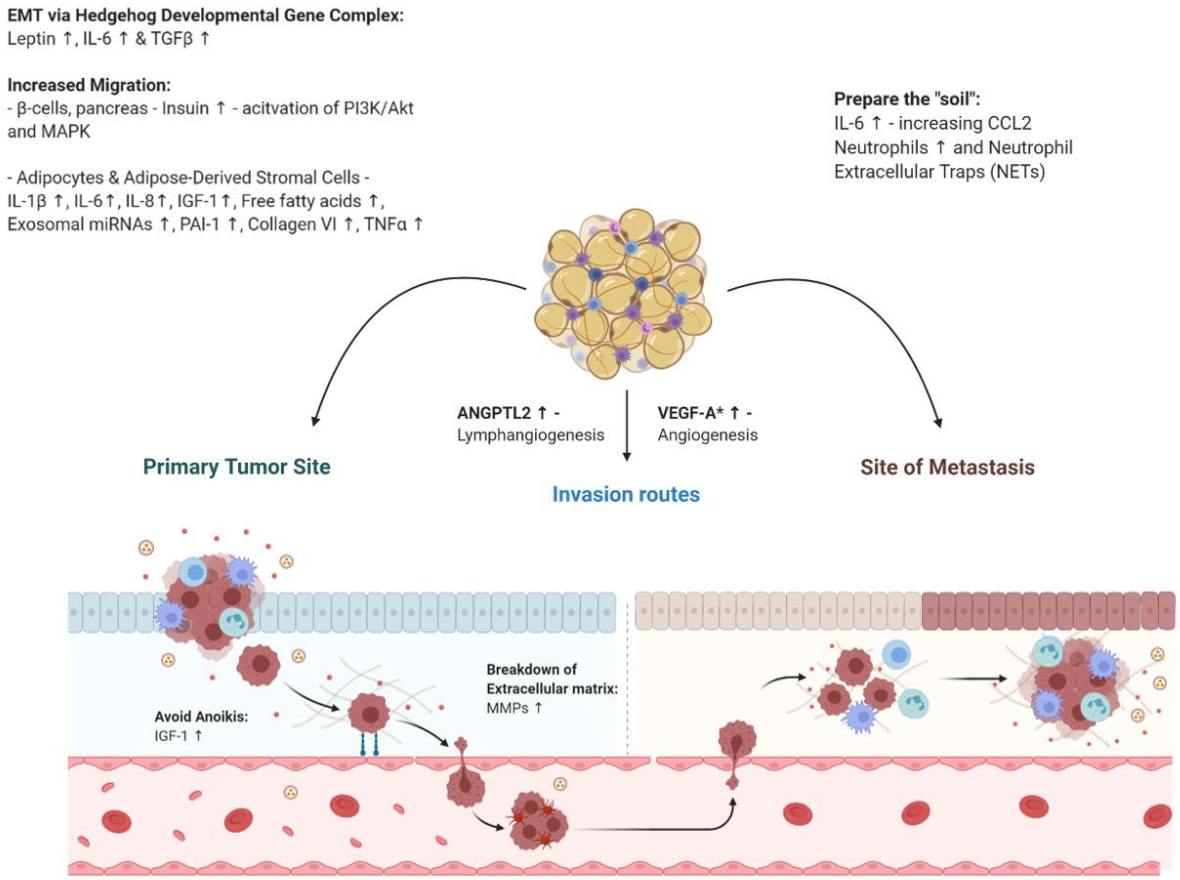


Figure 1.4: *Factors promoting cellular invasion and metastasis in obesity.*
Migration, invasion and metastasis are influenced by a number of different factors in obesity. At the primary tumour site, leptin, IL-6 and TGFβ promote epithelial-mesenchymal transition (EMT). Combining this with systemic hyperinsulinemia, exosomal influences and cross-feeding by nearby adipocytes cellular migration can be encouraged. Cells escape through the blood and lymphatic vessels, increased in number by elevated factors such as VEGF-A and ANGPTL2 (other factors reviewed in the *Inducing Angiogenesis* section). Further, the physiological state of obesity acts to prime distant niches for metastasis through various mechanisms, for instance the induction of CCL5 expression.

1.8 An Enabling Characteristic: Genome Instability and Mutation

Carcinogenesis can be a stepwise process which involves the accumulation of multiple mutations within cells allowing the *hallmarks* to be acquired. These mutations can occur at different rates depending on factors including internal and external stimuli/exposures (for instance radiation), intracellular processes generating genomic damaging materials, and the cellular response to these damaging exposures through pathways such as the DNA damage response and the presence and function of antioxidant enzymes, such as superoxide dismutase.

Excess body fat has been linked with locally increased production of reactive oxygen species (ROS), mainly generated in mitochondria[46,299]. ROS species damage DNA, and mutations often arise from attempts to repair this damage. Oxygen radical formation varies, with circulating fuels (increased glucose, fatty acids) and/or hormones playing a role. Possible mechanisms involve activation of NADPH oxidase (which transfers electrons from NADPH to molecular oxygen) and protein kinase C[300]. For example, abnormally higher levels of circulating insulin, due to insulin resistance in obesity, can cause DNA damage by increasing the translocation of Akt into mitochondria and its activation of NADPH oxidase. Oxidative stress can lead to an increase of lipid peroxides, alkoxyl radicals, aldehydes and ketones, and elevated lipid peroxide levels (estimated by plasma malondialdehyde and 4-hydroxyalkenals levels) have been repeatedly demonstrated to accompany fat accumulation in humans[45,301]. Free radicals cause damage to the hexameric repeats within telomeres, accelerating telomere loss during the cell cycle.While we noted above that longer telomeres led to an increased cancer risk, so too do shorter telomeres and in pre-malignant cells, telomeres can be extensively eroded[302]. Telomere erosion leads to chromosomal abnormalities including fusion, translocations and breakages, all of which contribute to genome instability.

Adipose tissue metabolites may cause damage to DNA independently of ROS, often by aldehyde production. This can lead to genomic instability through glycation which activates receptor advanced glycation end product (RAGE) receptors and their downstream pathways, which include the NFκβ pathway[299]. Indeed, methylglyoxal has been shown to cause DNA modification and strand breaks[303].

Other, less direct,effects on genome stability are secondary to other consequences of obesity. The hormonal effects of obesity - which include increased oestrogen and decreased steroid binding globulin[304], and various other factors reviewed above, drive proliferation and mitosis. Increased division increases the chance of replication error and mutation. In addition, production of oestrogen and its metabolites, which can occur peripherally by cytochrome p450 isozymes (mainly CYP1A1), can cause free radicals and DNA adducts that may precipitate mutation and replication error[305–307]. This may account for the DNA damage, (single strand and double strand breaks) that have been shown to be at higher levels in the lymphocytes of obese indivduals[206].

The ability of a cell to respond to, and repair, DNA damage obviously affects genomic stability. Several of the pathways involved in repairing DNA damage are potentially altered in the obese state[206] and this is borne out by the fact leucocytes from obese patients incur more genomic damage in response to the antitumor agent mitomycin C - which generates double strand breaks in DNA - leaving a greater amount of unrepaired DNA damage in the cells of the obese subjects[308]. Epigenetic mechanisms, including methylation of key proteins, may play a role[206]. Obese individuals are known to have differential methylation patterns on DNA and these can be altered by weight-loss[32,309,310]. DNA methylation is known to down regulate DNA repair enzymes in some cancers (e.g. bladder cancer)[311]. However, although feasible, a direct link between obesity, DNA methylation and impaired DNA repair has not yet been established.

A related finding is that high-fat diets, common in obesity, have been shown to promote colonic lysine homocysteinylation through induction of methionyl-tRNA synthetase. This in turn inactivates ataxia-telangiectasia and Rad3-related protein (ATR), by weakening its association with ATR-interacting protein (ATRIP)[87]. ATR plays an important role in DNA repair, responding to double strand breaks, aberrant replication fork structures, base adducts and DNA crosslinks likely due to the presence of single-stranded DNA structures[312–314]. The deactivation of ATR allows cells that undergo DNA damage stress to accumulate this damage which would otherwise have been repaired. Reliance on different DNA repair pathways may explain varying oncogenic susceptibility between tissues, dependent on their dominant pathway for DNA repair[315,316]. Indeed, ATR appears of particular relevance in breast, ovarian and colorectal cancers[317–319]. And, with increased proliferation, increasing replicative damage accumulates. With obesity being a major factor in colorectal cancer, this could be a plausible mechanistic explanation in how cells may develop genomic instability[309].

Telomere stability may also play a factor. Telomeres and their associated proteins prevent chromosomal fusion and prevent DNA being degraded during cell divisions[320]. Finally, as mentioned above, telomere length is typically reduced in obesity, and this also promotes genomic instability. The “perfect storm” of various factors acting to increase proliferation, shortened telomeres coupled with silencing of key DNA repair pathways, may lead to persistence of DNA damage and ultimately cellular transformation.

1.9 An Enabling Characteristic: Tumor-Promoting Inflammation

Inflammation is, in principle, a protective response mounted by injured tissue. An insult to a tissue (mechanical damage, infection) is met by a series of responses, including the release of various cytokines which attract lymphoid cells (especially macrophages), increased capillary permeability (which allow the macrophages into the tissue itself) and ultimately destruction of the invader/repair of the damage. This is an acute response, mediated by the innate immune system, and is typically over within a relatively short time.

However, obesity is associated with a chronic inflammatory state, in which immune cells accumulate in the adipose tissue and secrete cytokines. Inflammation is widely accepted to be associated with tumour development and progression through a number of mechanisms[321]. Inflammation in tissues reflects a balance between anti-inflammatory and pro-inflammatory factors. In obesity, adipose tissue expands and the population shifts towards Th1 cells, $CD8^+$ T cells and pro-inflammatory macrophages, whilst some T_{reg} and Th2 cell numbers decline[57,59,62–64]. Notably, the macrophage population of adipose tissue can rise to 40-50% of total cells, compared to <10% in lean adipose tissue[322]. Perhaps this is driven in part by monocyte chemotactic protein 1 (MCP-1), which is raised in human obesity[52] and has been shown to cause macrophage proliferation in adipose tissue in mouse models. In addition, mouse models of obesity have shown communication between adipose tissue and bone marrow to increase the production of myeloid cells[323]. The characteristics of adipose tissue macrophages are an area of much debate and is evolving. Currently, it is generally believed that pro-inflammatory macrophages (originally felt to be classical M1, but other data suggests a different pro-inflammatory phenotype[324]) are stimulated by leptin, to produce pro-inflammatory cytokines such as IL-6[325] . Also, high levels of PlGF are associated with obesity in breast and pancreatic cancer patients[209] and PlGF may also help shift macrophages into a pro-inflammatory state. TNFα and IFNγ are produced and the balance of

the lymphoid population alters[47,326]. TNFα promotes the production of cytokines such as IL-6 within adipose tissue[128], through its proteolytic release from the membrane. Eosinophils decrease in number in obesity, at least in mice[58]. These effects may influence the lymphoid populations in organs/tissues that tend to accumulate fat in obesity, such as the pancreas, breast and liver.

Recent work in triple-negative breast cancer patients and mouse models perhaps highlights the importance of a difference in lymphoid and myeloid populations. These studies investigated lean and overweight/obese patients, both with >30% stromal tumour-infiltrating lymphocytes (TILs). They showed that lean patients have a higher rate of a complete pathological response to neoadjuvant chemotherapy (73.1%) than overweight/obese patients with the same total proportion of TILs (44.7%)[327]. This presumably reflects a different make-up of the TILs population. Alternatively, this effect could reflect different levels of T-cell exhaustion as discussed in the next section. As regards myeloid cells, an important player here might be metabolically activated adipose tissue macrophages (MMes). MMes, driven by saturated fatty acids (e.g. palmitate) released by insulin-resistant adipocytes, accumulate in mammary fat of obese mice and humans and promote stemness within the tumour via IL-6[328]. Tumour-infiltrating myeloid cells also show an activated NLRC4 inflammasome in breast cancer mouse models and these can lead to cross-talk with adipocytes to promote angiogenesis and disease progression[329].

A key orchestrator in the induction of inflammation is the intracellular mediator, NF-κB, which, when activated, moves to the nucleus to drive the expression of inflammatory mediators, such as IL-6 and IL-18 (Fig. 4). Its accessory protein, IKK-β, can sense nutrient levels, notably fatty acids[330], leading to the activation of NF-κB and consequent production of inflammatory cytokines in diet-induced obesity. This is parallel to NF-κB activation through Toll like receptors (TLR), which are discussed below. Indeed, active NF-κB has been

described to play a major part in the onset of liver and intestinal carcinogenesis[331]. Opposing inflammation are the inherent anti-inflammatory actions of the Wnt/β-catenin system. In adipocyte hyperplasia, this system is down-regulated, leading to an increase in inflammation. These changes are also associated with changes in the T cell population, away from important subtypes of T_{reg} cells[332].

Further, in obesity, the adipocytes themselves can necrose. The dead cells release lipids, nucleic acids, damage-associated molecular patterns (DAMPs) etc, all of which attract macrophages. These form 'crown-like' structures around the dying adipocytes[333]. Activation of these macrophage pattern receptors activates inflammasomes[334] generating the inflammatory IL-1β, and more macrophages are attracted by endotrophin produced by adipocytes[335].

The microbiome is notably different in obese individuals[336]. This may affect carcinogenesis through local and systemic inflammation, possibly by allowing the leakage of lipopolysaccharide (LPS), which activates TLR4 and stimulates pro-tumorigenic inflammation[331]. Indeed, plasma LPS levels in both humans and rodents with obesity and insulin resistance are approximately doubled compared with those in healthy controls[337,338]. In addition, a healthy microbiome produces short chain fatty acids, such as butyrate, which have anti-inflammatory effects, and it is possible that the altered microbiome in obesity might be less effective in opposing inflammation due to decreased production of short chain fatty acids. Another putative mechanism is an effect of the microbiome in obesity on bile acid metabolism. Deoxycholic acid, a secondary bile acid, increases senescence-associated secretory phenotype (SASP) and thus the pro-inflammatory IL-1β, and leads to hepatocellular carcinoma in mice[339]. At the opposite end of the alimentary tract, the oral microbiome and inflammation of the oral cavity owing to bacteria (periodontitis) has been linked with obesity and also to cancer. Recent work has highlighted a link between periodontal disease, oesophageal adenocarcinoma and gastric carcinoma[340]. Antibodies to

and carriage of various oral pathogens (e.g. *Porphyromonas gingivalis)* are related to an up to two-fold increased risk of pancreatic cancer[341], indicating some systemic involvement. However, prospective and interventional studies involving the microbiome are needed.

Excess body fat and its influence on the immune system is undeniable with both myeloid and lymphoid cells affected. IL-6 appears a key cytokine in the chronic inflammatory state and thus likely in obese tumorigenesis. Controversies still exist surrounding exact macrophage profiles in obesity and the true extent of the influence of microbiota, which single-cell sequencing and interventional studies should help clarify.

1.10 An Emerging Hallmark: Avoiding Immune Destruction

As we have seen, in obesity, adipose tissue shows the characteristics of chronic inflammation, with the accumulation of some types of immune cell, including certain T cells and macrophages which release a variety of cytokine molecules. As such, it might be expected to be primed for immune action against tumours. However, this does not seem to be the case[342] (illustrated in Fig. 1.5). This is shown by poor responses to immune challenges, as evidenced by higher rates of vaccine failure and complications from infection in obese patients[343]. This may result in a reduced immunosurveillance of potential tumours [342,344].

White adipose tissue (WAT), particularly visceral WAT, is readily colonised by immune cells, particularly macrophages, T_{reg} cells, eosinophils and invariant natural killer (iNKT) T cells. This complement is unique to adipose tissue[326]. In lean subjects, the population of immune cells (particularly macrophages and iNKT cells) seem to maintain 'immune surveillance'; they are able to detect and kill potential tumour cells while producing cytokines (e.g. IL-10 and IL-4) which attract and maintain this population. $CD4^+$ and $CD8^+$ cells are also present and involved in killing cancer cells[345]. Since adipose tissue shows an increase in lymphoid content in obesity, one might expect it to be more efficient in its immune surveillance. In fact, this surveillance system appears compromised[346]. There are changes in the immune cell content in obese adipose tissue in humans and mice[347]. T_{reg}, iNKT and eosinophil numbers decline, while the macrophage population increases and shifts to pro-inflammatory state, exhausted memory B cell populations rise and IL-6 becomes the dominant cytokine produced[55,333,348,56].

Obesity has been shown to compromise mechanisms regulating T-cell generation by inducing premature thymic involution in mouse models. Diet-induced obesity significantly increased apoptosis of developing T-cell populations and reduced T-cell diversity[349] . Raised

circulating fatty acid concentration in the tumour microenvironment might also reduce the numbers of tumour-infiltrating $CD8^+$ T cells by inducing iron-dependent cell death (ferroptosis) via CD36[350]. As well as reducing numbers of antitumour cells, it seems that the normal immunosurveillance role of Th1 cells are blunted by the presence of cytokines secreted into the extracellular milieu shifting the balance away from cytotoxic responses towards inflammatory ones[351]. Type 2 diabetes in mice and humans, which is also associated with obesity, has been shown to reduce the multifunctionality of peripheral $CD8^+$ T cells in response to both non-specific and antigen-specific activation[352] . In obesity, T cells become 'exhausted' leading to a lack of function ('inflammaging')[349]. This phenomenon has been shown across multiple species, tissues and tumour models and is driven in part by leptin signalling[61]. There is an increased expression of the programmed death receptor (PD1) on the surface of $CD8^+$ T cells, and this binds to PDL-1 on the surface of tumour cells, preventing the T cells killing their targets[61]. The influence of obesity on the PD1/PDL-1 axis may explain its unexpected benefits on the efficacy of immunotherapy in some cancer patients[353], although other work has not confirmed such benefits and larger more in depth prospective trials are needed clarify any potential benefit[354] (discuss further in *Obesity and cancer presentation and treatment*).

Somewhat paradoxically, the hormone most affected by obesity and linked to tumour growth (above) seems to promote antitumour activities by the immune system[355]. Leptin has been shown to promote the development of NK cells[356], dendritic cells (involved in antigen presentation)[357], Th1 lymphocytes[358], and eosinophils (involved in the adaptive and innate immune responses)[359] while decreasing the level of T_{reg} cells[355]. However, leptin also stimulates the release of IL-6, IL-1β and TNFα (particularly from adipose tissue macrophages)[360] which, together with MCP-1 (also released in inflammation), seem to attract myeloid derived suppressor cells (MDSC) which can suppress important $CD8^+$ T cell-mediated anti-tumour immunity[361,362,363]. So there seems to be a balance between leptin

promoting immunosurveillance, while at the same time causing the accumulation of MDSCs which oppose it, although the majority of the work above was in model systems.

Other factors in the tumour microenvironment have been implicated in immune suppression in obesity. IL-18 is raised in obesity (both in tissues and in the circulation)[51] and, while it directly promotes cell proliferation in stomach cancer and melanoma, it can also drive immune evasion by decreasing the expression of CD70 which normally increases the activity of antitumour NK cells[281]. Further, human and mouse NK cells can become "paralysed" in obesity showing increased lipid accumulation, impaired lytic machinery mobilisation and loss of cytotoxicity against tumour cells .[364].

AhR is expressed at high levels in several tumours, and promotes immune tolerance by increasing the proportion of T_{reg} cells (immunotolerant) at the expense of $CD8^+$ T cells at least in human glioma[134]. AhR is activated by endogenous ligands such as kynurenine, whose levels rise in human obesity[43] and which is itself metabolised into a number of products, particularly 3-HAA, which suppresses the pro-inflammatory mouse Th1 and Th17 cells[365]. Furthermore, macrophages, attracted into the tumour microenvironment in response to MCP-1, produce TGFβ and IL-10 which suppress the cytotoxic activities of Th1 cells[366]. MCP-1, TGFβ and IL-10 have all been described as elevated in the serum of obese individuals[50,52,54]. The production of arginase and of indoleamine dioxygenase are upregulated, leading to depletion of arginine and tryptophan that immune cells need for proliferation. Increasing lactate and adenosine in the tumour microenvironment may also promote immunosuppression[367]. Further, high-fat diets in mice paradoxically appear to starve $CD8^+$ T cells of lipids required for their proliferation. This has been attributed to reprogramming of energy metabolism towards fatty acid catabolism in adjacent tumour cells, through a decrease in proline hydroxylase-3 (well-known for its regulatory impact on HIF1α)[368,369]. $CD8^+$ T cells may also undergo metabolic reprogramming. In breast cancer mouse models, obesity activates the leptin-STAT3 axis causing increased fatty acid

oxidation and a reduction in $CD8^+$ T cell antitumor function[370]. And in sarcoma mouse models, it has been shown that nutrient competition, particularly for glucose, between tumours and T cells is a distinct mechanism, which can lead to T cell hyporesponsiveness[370,371]. Other work in mice has shown that the cholesterol metabolite, 27-hydroxycholesterol, which is thought to be increased in obesity, increases the number of polymorphonuclear neutrophils and γδ-T cells at distant metastatic sites in breast cancer models and this might enhance metastasis by polarising towards an immune-suppressive subtype[372].

Obesity appears to allow tumour cells to evade destruction mainly through its effects on T-cells. Reduction in T-cell production and diversity, coupled with exhaustion appear key, but perhaps provide an “Achilles heel” through PD1 targeting. The role of AhR is intriguing and is worthy of further investigation.

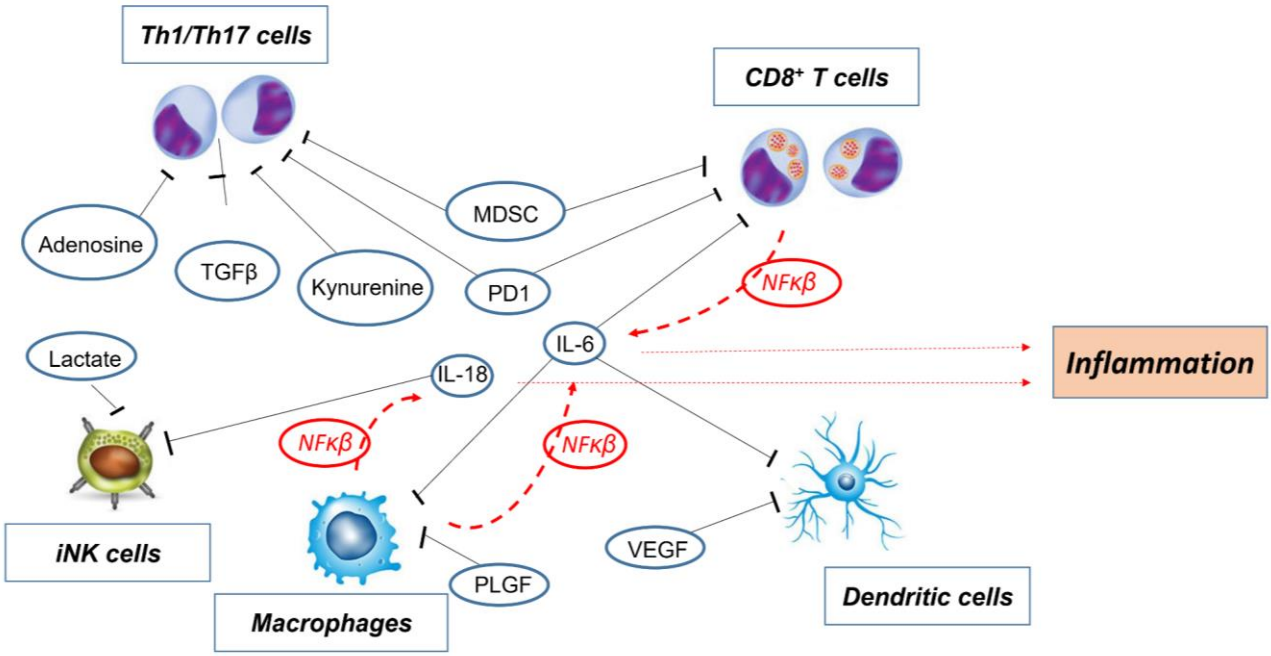


Figure 1.5: *Effects of obesity on immunosuppression and inflammation*.
Cells relating to immune surveillance of cancer, raised in obese adipose tissue, are shown (Th1/Th17 and CD8+ T cells, dendritic cells, pro-inflammatory macrophages and iNK cells). Suppression is shown as black blunted lines. Circulating cytokines IL-6 (acts on dendritic, CD8+ T and macrophages) and IL-18 (acts on iNK cells) are major mediators of immune suppression, but other growth factors/cells suppress immune responses in dendritic cells, (VEGF), macrophages (PLGF), Th1/Th17 cells (TGFβ, MDSC), and CD8+ T cells (MDSC). Small molecule effectors may also be involved - adenosine and kynurenine dampen the immune response in Th1/Th17 cells while lactate suppresses actions of iNK cells.

Simultaneously these cells, notably the pro-inflammatory macrophages and CD8+ T cells, release cytokines IL-6 and IL-18 (red arrows), resulting in a complex feed-back system. They are stimulated to do so by extracellular mediators such as fatty acids (raised in obesity), DAMPS (from dying adipocytes in crown-like structures), IL1β and TNFα, via NF-κB, the predominant intracellular coordinator of inflammation.

1.11 An Emerging Hallmark: Reprogramming Energy Metabolism

Although there has been considerable focus on the genetic basis of cancer, it is clear that there must be considerable metabolic change in tumour cells, to provide both the energy and the raw materials for cell growth and proliferation of cells. While this might be considered simply a consequence of cells adjusting to growth modification in cancer, the discovery of oncogenic metabolites associated with the tricarboxylic acid cycle, and the ability to modify pathways as a route to cancer therapy[373] shows the importance of altered metabolism as a cancer hallmark. Some go so far as to classify cancer as a 'metabolic disease'[374], and certainly the interplay between cell growth and proliferation, and energy metabolism (discussed below) highlight the requirement for tight integration between cell energy metabolism and tumour growth.

1.11.1 Glycolysis & mitochondrial metabolism

In many cancers, tumour cells exhibit aerobic glycolysis[375], using the glucose to lactate pathway even under conditions of oxygenation. Levels of the glycolytic activator, F2,6P, are known to rise in many cancers[376] as are the various proteins associated with production/breakdown. Paradoxically, the antitumor protein TP53-Induced Glycolysis And Apoptosis Regulator (TIGAR, an F2,6 bPase) also rises (following increases in p53) (see below), and this might be expected to inhibit glycolysis. However, TIGAR induction has also been linked to autophagy and apoptosis[377]. There is an interesting interplay between tumour cells and neighbouring cells. In colorectal cancer, glycolytic cancer cells express the MCT4 lactate transporter, which operates to export lactate to 'feed' nearby fibroblasts (CAF) which support tumour growth[378]. However, in other systems, tumour cells can induce glycolytic metabolism in cancer associated fibroblasts ('reverse Warburg effect'). In epithelial cancers, CAFs (despite themselves being untransformed) turn glycolytic and export lactate (via

monocarboxylate transporter 4, MCT4). This provides the (oxidative) cancer cells (expressing monocarboxylate transporter 1, MCT1) with fuel and an acidic environment, both of which favours their growth[379].

Glucose levels are often high in the circulation in obesity, due to insulin resistance, and thus the tumours are provided with the necessary glucose. Abundant fuel supply is associated with a stimulation of glycolysis via expression of the unusual isoforms of glycolytic enzymes, PKM2, LDHA and GLUT1,3, which are common in tumour cells[41,380,381]. Fat deposition can increase tissue distance from vessels, potentially causing hypoxia. Hypoxia leads to energy production being largely glycolytic (producing lactate) and/or employing reductive glutaminolysis. A major driver of these changes is HIF1α, which is stabilized in hypoxia. HIF1α acts as a transcription factor for enzymes that promote a glycolytic phenotype, such as PFK1, GAPdH, aldolase, and isozymes such as glut 1, LDH-A, PKM2[380,381], while c-myc induces synthesis of glutaminase, the amino acid transporter ASCT2, and LDH-A , promoting a the use of glutamine as energy source[382]. Intriguingly, pyruvate kinase M2 (PKM2) also acts as a cotranscriptional activator with HIF1α, leading to a potential positive feedback system[383].

Mitochondrial respiration may also be depressed in obesity by leptin, as shown in colon cancer cell lines[384], possibly indicating a shift to a glycolytic phenotype. Indeed, there is increasing evidence that tumour metabolism can switch between glycolytic and oxidative metabolism depending on tissue of origin, cancer progression and malignancy[385]. This may be dependent on the tumour - Rabin-Court et al claim that cell lines from obesity-associated cancers (e.g. colon cancer) switch to oxidative metabolism of glucose in the presence of insulin, while others (e.g. small cell lung cancer) do not[386].

1.11.2 Free fatty acids

In obesity, adipose tissue expands. This is accompanied by an increased release of fatty acids into the circulation and these certainly fuel some cancer growth[387]. The increased fatty acids appear to reflect a constant level of lipolysis (per unit mass of tissue) in a larger mass of tissue, rather than a stimulation of lipolysis itself, but some workers have identified an increase in monoacylglycerol lipase, which promotes fatty acid release from adipose tissue, in several cancers[388]. Another factor in this fatty acid release is PPARγ, a transcription factor associated with adipose tissue development. This promotes glucose uptake and associated fatty acid esterification in white adipose tissue[389]; its level has been reported to be decreased in obesity (possibly via production of miRNA species (miR-27b, 130b, 138)) and decreased esterification would also lead to increased fatty acid release[389]. Besides fuelling tumour growth, fatty acids may promote the production of ROS and prostaglandins (which can promote metastasis) by tumour cells via COX2[390,391]. In murine breast cancer, a shift from glucose metabolism to fatty acid oxidation with a shift to a metastatic phenotype has been noted[392]. Further, adipocytes release fatty acids to nearby cells ('cross feeding' the tumour) which in turn increase oxidative capacity and levels of A-FABP[267]. However, the voracious appetite of tumour cells for fatty acids has been shown to limit $CD8^+$ T cell function in the tumour microenvironment (see *Avoiding Immune Destruction*). More work is certainly needed to further investigate the role of particular fatty acid families (e.g. unsaturated, short chain) and individual fatty acids (e.g. butyrate) in the obesogenic carcinogenesis.

Perhaps paradoxically, cancer cells also upregulate fatty acid synthase, so both increasing fatty acid breakdown and synthesis, and this upregulation is also observed in obesity[393]. The increase in fatty acid synthesis is presumably linked to their biosynthetic need for new membranes[394].

1.11.3 mTOR & AMPK

The mechanistic target of rapamycin (mTOR) is an evolutionarily conserved serine/threonine-protein kinase and a central player in modulating the balance between cellular metabolism and cell proliferation[395]; mTOR is present in two complexes, mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2), comprising different components. mTORC1 has its exclusive partner raptor (regulatory-associated protein of mTOR) and mTORC2 has its partner rictor (rapamycin-insensitive companion of mTOR).

Activation of mTORC1 increases cellular anabolism, increasing fatty-acid biosynthesis and, nucleotide synthesis[396] and reducing autophagy in some tissues[397]. Fatty acid biosynthesis is promoted by the ability of mTORC1 to activate SREBP and induce the production of PPARγ (a transcription factor for lipogenic enzymes)[398]. mTORC activation also leads to increased levels of HIF1α (see below) which, as a transcription factor, can increase the expression of several glycolytic enzymes[399,400] and promote mitochondrial biogenesis[401].

Less is known about mTORC2. When activated, mTORC2 phosphorylates Akt which then phosphorylates (activates) tuberous sclerosis complex protein 2 (TSC2) which then inhibits mTORC1[402] suggesting it may operate as a brake on mTORC1. In the cell, mTOR1 is activated by nutrients, notably amino acids, acting as a nutrient sensor and working via the S6K pathway, while mTOR2 responds to growth factors and acts via the Akt (and PLC) pathways.

mTOR tends to be activated in the obese state, via phosphorylation by Akt resulting from stimulation by leptin and chronic raised insulin levels. This can be observed by increased phosphorylation (activation) of its target, S6K, which promotes protein synthesis mTOR[179,403]. This shifts metabolism towards anabolism, and primes for growth and proliferation. Also upstream of mTOR lies AMP dependent kinase (AMPK), which inhibits it. AMPK is a phylogenetically conserved energy-sensing enzyme which can shift metabolism towards catabolic pathways and away from biosynthesis (in response to raised AMP

levels)[404]. As such, this enzyme (and its physiological activator, LKB1 or its pharmacological activator, metformin) can act as tumour suppressors – typically by downregulating mTOR (by phosphorylation of the mTOR inhibitor, TSC complex). AMPK also activates p53 (see below). The activity state of AMPK varies in obesity - high leptin levels will lead to an inhibitory phosphorylation (via Akt) while adiponectin (which falls in obesity) causes an activatory phosphorylation (via PKA)[149]. Thus obesity leads to mTOR activation and releases one of the normal brakes on cell growth and proliferation. AMPK has other, more direct, effects on metabolism: for example it phosphorylates/inhibits acetyl CoA carboxylase, inhibiting the fatty acid synthesis required for new membrane synthesis/cell proliferation[405]. Notably, exercise is a powerful physiological activator of AMPK[406,407], and this may account for the beneficial effects of exercise in tumour suppression - while decreased physical activity is associated with obesity[408].

1.11.4 TP53

TP53 is the most frequently mutated gene in human cancers and is widely known as a tumour suppressor, affecting genes involved in DNA repair, cell cycle progression and apoptosis[409,410]. In obesity, TP53 activity is controversial. Normally, TP53 is unstable in the cell, but it can be stabilised by phosphorylation by AMPK, under the influence of adiponectin[149]. Thus it might be expected that TP53 concentration falls as adiponectin falls in obesity, and indeed this fall appears to be exacerbated by leptin in cellular models[97]. In contrast, some authors claim that TP53 levels increase in obesity, exacerbating the release of inflammatory cytokines[96]. Whichever effect is dominant, a potential influence of TP53 on tumorigenesis in obesity cannot be ignored. Indeed, polymorphisms of TP53 have been linked to obesity and lean body mass in mice and humans[98,99].

1.11.5 Adipocyte-derived exosomes

Adipose tissue expands in obesity and it can communicate with other cells using nanovesicles (exosomes). Adipocytes secrete exosomes in abundance and, as noted above, these exosomes have been shown to influence gene expression in other tissues at least in mice[268], in particular expression of fibroblast growth factor 21 (FGF21) in liver which can affect fatty acid oxidation and glucose uptake[411]. Adipocyte-derived exosomes are taken up by tumour cells in a number of different settings, particularly in obesity-related malignancies[269–271]. These exosomes contain several proteins and RNA species, including miRNAs (above), many of which target energy metabolism. Obesity changes the miRNA content of exosomes in mice; miR-122, miR-192, miR-27a-3p, and miR27b-3p are all increased[93]. These miRNA species can regulate lipid metabolism[412–414]. When these exosomes were given to lean mice, glucose intolerance and insulin resistance were induced[93]. Further, it is worth noting that tumours can also remodel adjacent adipocytes metabolism through exosomal miRNAs, promoting beige/brown characteristics in breast cancer co-culture[415,416].

In melanoma, it was demonstrated that exosomes from adipocytes were specifically enriched in proteins implicated in fatty acid oxidation (FAO), and these changed the metabolic nature of tumor cells in vitro favor of FAO. In breast cancer, exosomes from adipocyte-derived mesenchymal stem cells activate Wnt signaling within tumour cells. Wnt is involved in determining cellular metabolism, by switching between glycolysis and oxidative metabolism[417] and also regulating cellular metabolism through its effects on expression and of transcription factors including c-Myc. c-Myc further influences lipid metabolism via PPARδ[418] (a regulator of lipid oxidation and influence on oxidative capacity) and PPARγ (a controller of lipogenesis, mentioned above).

This cross-talk between adipocytes and ASCs has highlighted some other key differences in obesity-related cancer metabolism. Recent work in murine models of obesity-

accelerated breast cancer highlighted creatine metabolism as a potential key player, with deletion of the rate-limiting enzyme in creatine biosynthesis, glycine amidinotransferase (Gatm), resulting in reduced tumour growth[419].

Circulating metabolite profiles of obese individuals provide nutrients for cancer cells but the link between obesity and metabolic reprogramming appears to go beyond just providing more fuel. Local and distant cross-talk between adipose tissue and cancer cells appears important, with creatine and glutamine metabolism providing interesting areas for further investigation in obesity-related tumours. We see that metabolic communication can involve miRNAs and further clarifying how adipose tissue communicates, including through metabolite exchange, may help us understand carcinogenic transformation and subsequent metastasis in obesity.

1.12 Obesity and cancer presentation and treatment

In the setting of cancer, the main treatment modalities are surgery, radiotherapy, antineoplastic agents, such as chemotherapeutics, immunotherapies and hormonal therapies. The effects of obesity on cancer treatment modalities are well-documented. However, the treatment of a patient is always tailored to take into account their comorbidities, performance status, stage of disease (TMN) and in some scenarios their genetic makeup (e.g. eGFR mutation). In general, obese subjects tend to present at later stages and with several obesity-related malignancies[420,421], breast cancer being a pertinent example - tumours tend to be larger in the obese population as they may escape notice in larger breasts[422].In addition, excess body fat can prove challenging to our imaging methods, including those used in cancer screening. Particularly in the context of breast cancer, mammography is attended less by obese patients and has been seen to be more difficult in the obese patient, although conclusions relating to its effects on sensitivity and specificity have been mixed[423–425].

Obesity is a significant co-morbidity and may significantly alter the treatment course and ultimate recovery. In some settings, excess body fat makes treatments more challenging, particularly in surgery where it impacts on ventilation intraoperatively and subsequent postoperative complications, such as poor wound healing[426–428]. A similar picture is seen in external beam radiotherapy, where toxicities, particularly skin toxicities, are at higher rates in obese patients[429,430]. Obesity also poses problems for therapeutics, whether it be the effect of the adipose tissue increasing the volume of distribution, its effects on cytochrome P450 enzyme induction, the metabolic activity of the adipose tissue or the complications associated with chemotherapy dosing. It is notable that there is limited concensus on the most appropriate way to dose in obese patients, although weight-based cytotoxic chemotherapy doses have been recommended for obese patients[431]. There are suggestions that hormonal therapy choice should be altered in obese individuals when

treating breast cancer, potentially by choosing a more potent alternative to anastrazole, a commonly used inhibitor of aromatase[432,433]. Furthermore, in vitro work has shown that leptin can interfere with ER+ breast cells response to tamoxifen[434].

However, with every challenge comes opportunity and obesity has already provided some unexpected synergistic opportunities, as obesity changes both the host and tumour biology. For instance, transcriptomic analysis of breast, renal and prostate[435–438] identifies differences between tumours in obese and non-obese individuals. The so-called *obesity paradox* in cancer treatment has become apparent in the last few years, with a number of studies showing that overweight and obese states, although often more prone to tumour initiation, are also associated with improved survival[439–446]. It is possible that some of these observations may be artefactual, depending on the methods used. Methodological considerations include various biases due to relationships between BMI and confounders such as smoking, collider stratification bias, the time of recording BMI (and thus "obesity") during the experimental data collection, the use of 5 year survival rates as a parameter (as these are skewed by early diagnosis[447]) and various other potential biases (reviewed in [448]). Another explanation may simply be that increased adiposity provides an energy reserve which is useful in times of stress, i.e. cancer treatment.

However, if this effect is real, it opens up the exciting possibility that certain therapeutics might work better in obese patients. Recent work on survival in melanoma is particularly instructive[446]. Here, a combined analysis of multiple large, independent cohorts of metastatic melanoma patients showed increased progression-free survival, and overall survival, in obese males treated with targeted (BRAF and MEK) inhibitors or immunotherapies (anti-PD-1 and anti-PDL1 antibodies). The relationship was less significant in women, but this may reflect variation with menopausal status, which was not known. Perhaps the increased response to the targeted agents could be related to activation of, and/or reliance on the BRAF/MEK pathways in these patients. However, it has also been

shown that high-fat diets enhance BRAF V600E-MEK1 activity and promote tumour growth in this melanoma, apparently through an increase in serum acetoacetate[449,450].

The concept that immune checkpoint inhibitors may work better in obese patients is also intriguing as the immune ecosystem is markedly altered in obesity, and this effect has been noted in various tumour types, e.g. non-small cell lung cancer[451], renal cell carcinoma[452] and in a pooled analysis[453] (depicted in Figure 2). However, recent work again in melanoma, suggests the obesity paradox may be restricted to overweight/class I obese individuals and may be related to lean body mass rather than adiposity[454]. This potentially highlights the importance of taking more detailed anthropometric, metabolic and immune measurements when trying to understand this *obesity paradox*.

Obesity also suggests other potential pathways as targets. Differences between cancerous cells and normal tissue form the basis of all oncological interventions, and the differences detailed in the sections above are therefore potential opportunities. With leptin levels being a clear difference between obese and non-obese individuals, and its role as a tumorigenic factor, one might suggest anti-leptin therapy would be useful in treating obesity-related cancers. There have been some studies in this area with development of compounds to bind leptin in the circulation, such as anti-leptin antibodies and nanobodies[455], which have been used experimentally in pre-clinical trials in neurological research (in conditions such as multiple sclerosis & autoimmune encephalomyelitis)[456,457], cardiovascular, gastroenterology and anti-obesity research[458–460] . However, these have not translated into the clinic, nor been extensively investigated as anti-tumour agents.

Rather than decreasing leptin, the supplementation of adiponectin directly or indirectly is another potential option which requires further investigation. Current medications licensed for diabetic management, the thiazolidinediones (e.g. rosiglitazone and pioglitazone), have been shown to elevate levels of adiponectin[461], but have associated cardiac and other toxicities which mediate against their therapeutic use in cancer.

Supplementation of the diet with aged garlic[462], L-cysteine[463] or manganese[464] have all been shown to increase adiponectin levels, although their therapeutic use in cancer remains to be studied.

Lifestyle interventions also can alter these key hormones. Several studies indicate that adiponectin levels increase with exercise[465,466]. Their effect on leptin levels is more debated, although a recent study in mice suggested leptin levels fell on exercise, accompanied by a reduction in the numbers of inflammatory leukocytes[467].

With the cytokine and prostaglandin storm present in obesity, it could be hypothesised that non-steroidal anti-inflammatory drugs (NSAIDs) may have a beneficial effect in this cohort of patients[468]. It has been shown that prescription of aspirin is associated with a reduction in colorectal cancer risk, a key obesity-associated malignancy[468]. In a subsequent study of obese patients with Lynch Syndrome, who are at higher risk of colorectal cancer due to a defect in mismatch repair, aspirin again reduced their increased cancer risk[469].

Angiogenesis blocking agents have been shown to have blunted effects in breast cancer models in obese mice. This was attributed to the elevated effects of IL-6 and fibroblast growth factor 2. IL-6 blockade and metformin normalisation of FGF2 levels restored sensitivity to anti-VEGF therapy[470]. This effect is yet to be explored in humans. Nonetheless angiogenesis, as an important activity of tumours, provides a number of possible alternative physiological targets in obesity.

Finally, weight loss in obese patients per se, by diet or bariatric surgery, has been shown to reduce the risk of obesity-associated malignancies[471–474], although the exact mechanisms for such effects is not known.

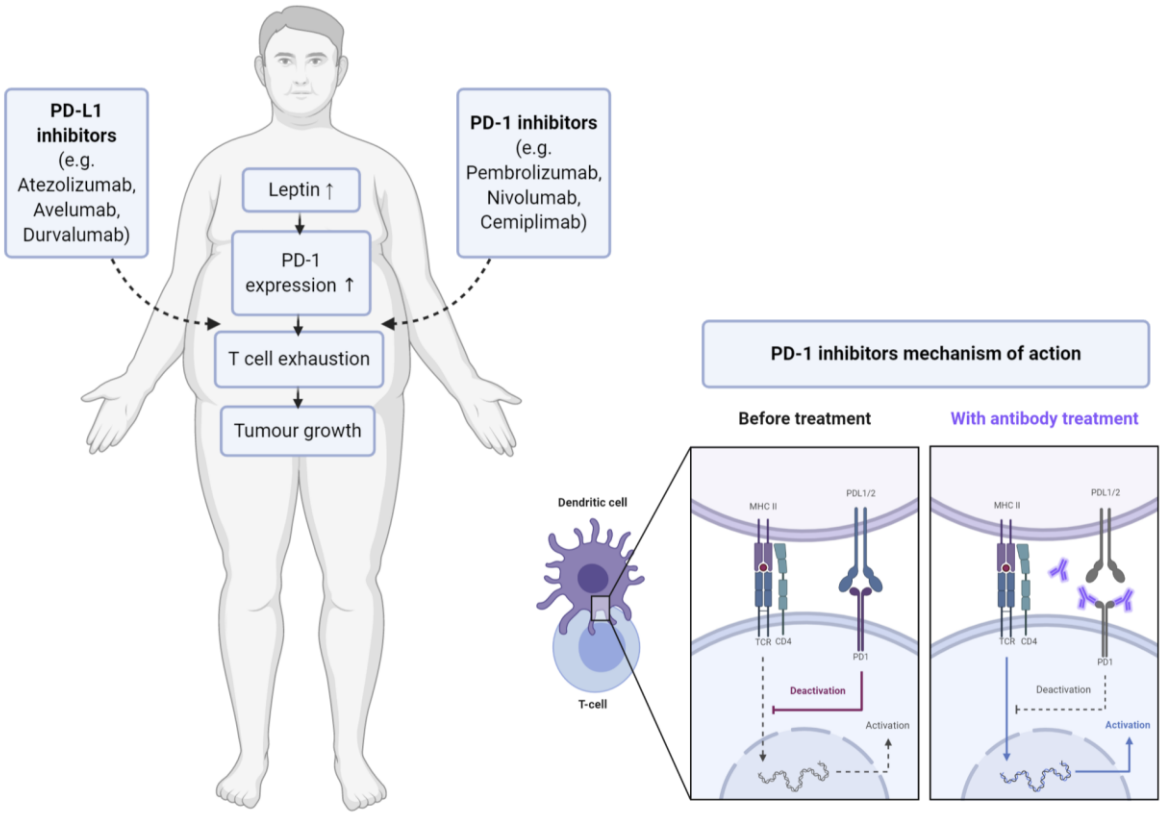


Figure 1.6: *Checkpoint blockade in obese patients.*
Programmed cell death 1 (PD-1) receptor on T cells interacts with programmed death ligand (PD-L1) on a target cell and suppresses T-cell mediated tumor cytotoxicity, allowing a balance between cell growth (self tolerance) and immune elimination. Cancer cells may overexpress PD-L1 to avoid cell death, and thus escape immunosurveillance. However, PD-1 interaction with PD-L1 reduces the ability of $CD8^{+}$-T cells to promote cell killing, and the immune cells are said to become 'exhausted'. Leptin promotes PD-1 expression and T cells from obese donors have demonstrated a significant increase in PD-1 expression. This leads to increased T-cell exhaustion. Antibodies against PD-1 or PD-L1 prevent this interaction and allow tumour cells to be eliminated appropriately.

1.13 Conclusion

The state of obesity has multiple effects on the human body, several of which can influence the carcinogenic process. Indeed, all the Hallmarks of Cancer, both enabling characteristics and emerging hallmarks are potentially influenced by obesity, with leptin and IL-6 appearing to play key roles (Fig. 1.7). This work is meant to act as an overview of this complex and evolving area but it is beyond its scope to act as a completely comprehensive reference text.

The majority of studies described here were carried out in model organisms and cell lines. More work and innovative study design is required to tease apart obesity and carcinogenesis in the clinic. Indeed, there appears to be paucity in the literature of large prospective cohorts monitoring cytokines and other circulating factors over a range of tissues, and subsequently tumours, in obesity. Whether these obesity-associated factors predispose to specific driver mutations remains to be seen. Perhaps weight-intervention measures or other dietary approaches which are relatively low cost, and could reduce tumorigenesis in obesity as well as other major comorbidities, need greater consideration by public health bodies as the physical and psychological burden of cancer carries significant personal and public health costs. Window of opportunity studies in cancer provide routes to the clinic for novel treatment approaches in obese patients. Trials stratifying patients by adipose tissue content, or biomarkers of adiposity that influence carcinogenic processes would be valuable. Particularly, repurposing previously approved drugs to target pathways specifically modified in obesity might be fruitful as adjuncts or alternatives to current treatment regimens. Likely candidates come from across specialities and include: oestrogen receptor antagonists (e.g. tamoxifen), VEGF signalling inhibitors (e.g. pazopanib), insulin response and mTOR modulators (e.g. metformin), JAK inhibitors (e.g. baricitinib) and even IL-6 receptor antagonists (e.g. tocilizumab). Further work is necessary to resolve the conflicts

in the literature surrounding PD-1/PD-L1 inhibitors in obesity, and interventional studies involving the microbiome might be an interesting approach.

Indeed, the links between obesity and cancer remain an intense area of research and debate with many unanswered questions. And, now with data maturing from some of the largest and most extensive cohort studies ever undertaken, we have the exciting opportunity to mine these resources for such, and hitherto unsuspected, associations. To this goal, this work has identified several under-investigated areas for further study, and initially we focus on the explanatory factors behind the BMI association with cancer, looking at fat mass and fat-free mass directly using bioelectrical impedance.

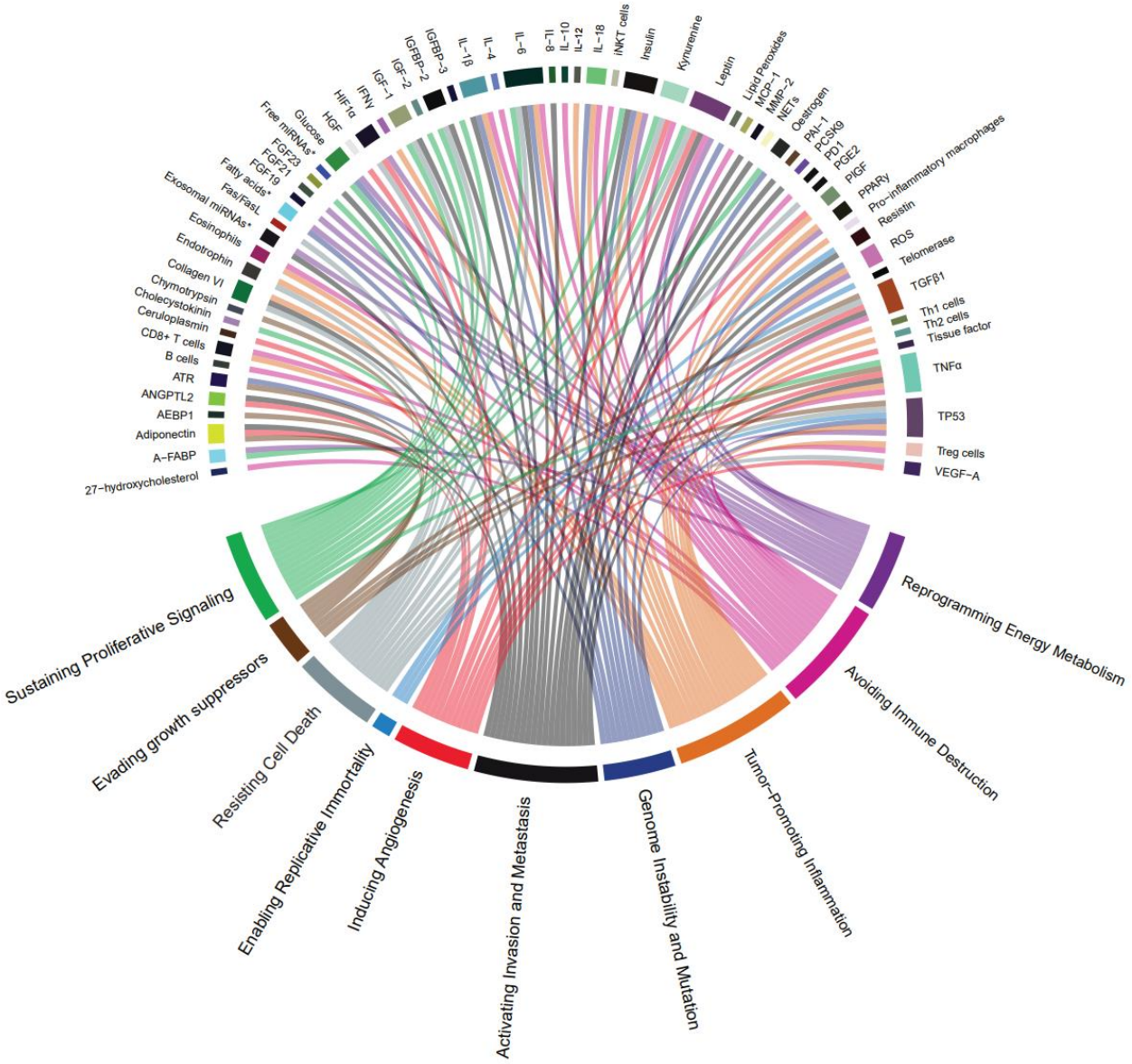


Figure 1.7: *Links between factors affected in obesity and individual hallmarks of cancer.*
Circos plot illustrating the links between obesity-related factors (as enumerated in Table 1.1) and the Hallmarks of Cancer, Enabling Characteristics and Emerging Hallmarks. Factors are located on the top half of the plot whilst Hallmarks are on the bottom. Each hallmark is identified by a particular colour (lower section) and lines from this radiate to any factor suggested to influence it.

Chapter 2: Untangling cancer, body mass index and obesity: anthropometric risk factors in breast, colorectal and prostate cancer

Chapter Summary

Background: The link between obesity and cancer is well established but not well understood. Most of the studies associating obesity and cancer are based on indirect measures of fat, such as body-mass index (BMI). BMI can be skewed by tissues other than fat mass, particularly lean mass, and studies in other fields have highlighted overreliance on body mass index can lead to misinterpretation. This chapter investigates how more direct measures of fat and lean body mass influence the risk of developing cancer.

Methods: Anthropometric measures and self-reported life-style information were collected from a large population-based cohort in the United Kingdom (UK Biobank cohort, n = 502,617). The focus was on BMI, whole-body fat mass (WBFM) and whole-body fat-free mass (WBFFM). Odds ratios (ORs) for post-menopausal breast, colorectal and prostate cancer risk are calculated on these anthropometric measures using the logistic regression models adjusting for salient covariates. Principal component analysis (PCA) investigates the origin of the BMI signal, and its association with cancer, across three tumour types.

Results: Increasing quintiles of BMI, WBFM and WBFFM are associated with an increased risk of post-menopausal breast cancer, with the highest quintile of WBFFM giving the highest odds ratio of the three (1.603 [1.44-1.78], $p = 8.77E^{-19}$). PCA suggests both increased WBFM and WBFFM may both confer increased risk. The upper quintiles of BMI (OR: 0.835 [0.76-0.91], $p = 7.53E^{-05}$) and WBFM (OR: 0.832 [0.76-0.91], $p = 5.22E^{-05}$) appear protective against prostate cancer and PCA shows this signal likely originates from WBFM. In men, the highest quintiles of BMI, WBFM and WBFFM appear to carry increased risk of colorectal cancer and PCA suggests both WBFM and WBFFM might influence risk.

Conclusion: The merit of taking more direct measures of fat mass and fat-free mass as well as BMI when investigating cancer risk is highlighted. The novel use of PCA helps to

disentangle the BMI signal, suggesting it is a worthwhile method to employ in epidemiological studies such as this.

2.1 Introduction

Obesity, in terms of excess adiposity, is generally considered a risk factor in cardiovascular disease[475,476], diabetes mellitus[476], chronic kidney disease[476] and a large number of cancer types[6]. Body mass index (BMI) has been widely used as a surrogate for obesity in such studies.

BMI is a measure proposed by the statistician Adolphe Quetelet in 1835[477] and defined as [mass (*kg*)] / [height (m)]^2. It aims to provide a measure of body mass which is standardised across individuals of different heights. Patients can be categorised by BMI and, in risk analysis, cut-off BMI values are taken both operationally, as useful prompts in the clinic, and analytically, to aid data interpretation. The thresholds for overweight and obese subjects have been debated but currently *overweight* subjects are defined as having a BMI >25 kg/m^2 and *obese* subjects having a BMI >30 kg/m^2[478,479].

Although BMI has been described as a good proxy for assessing overall body fatness[6], a number of publications have raised concerns over its shortcomings[479–483]. For example, as a general measure of adiposity BMI can be misleading in athletes, where increasing muscle (denser than adipose tissue) will also raise BMI. Studies relating BMI to body fat percent measured by under-water weighing have shown a limited correlation of only about 0.5–0.7[484,485].

In addition, there are indications from other fields that it may not be safe to assume that the adipose part of the BMI confers an increased risk of cancer[486] or other conditions. This is clearly demonstrated by the well-described association of atrial fibrillation (AF) with obesity, as established mainly through BMI-based studies[487–490]. In fact, this association appears to be driven predominantly by lean body/fat-free mass as opposed to fat mass[486]. The authors showed that there was no association observed with AF and obesity-related anthropometric measures after adjustment for fat-free mass[486]. This finding was also

confirmed in subsequent cohorts[491]. Although caution must be applied when correcting for intercorrelated variables, this finding casts doubt over the role played by fat in the AF aetiology[486], and more generally has opened up a number of questions surrounding epidemiological studies that have used BMI as a surrogate for obesity.

Within oncology, the majority of epidemiological studies have indeed used BMI to denote obesity, partly due to the ease of data collection over large sample sizes. Others have studied changes in BMI or weight over time and/or used other indicators of adiposity, such as waist circumference and waist: hip ratio. However, relatively few studies have based their findings on more "direct" measures of fat mass, such as bioelectrical impedance, dual-energy X-ray absorptiometry or hydrodensitometry[492–495]. Such data is now becoming available. Large cohorts are being used to collect more sensitive measures of central adiposity, for instance waist: hip ratio, and more direct measures of adiposity, such as bioelectrical impedance or MRI measurements of adipose content. This is important as biologically speaking, adipose and muscle tissue have different autocrine, paracrine and endocrine actions[496,497] and hence different physiological actions on processes such as whole body metabolism, inflammation, and insulin resistance[498,499].

Within the field of cancer treatment, an 'obesity paradox' has also been noted in patients. Although those with a higher BMI are at increased risk of developing cancer, higher BMI at diagnosis does not always translate to a higher mortality, and in some cases has been associated with increased survival. This paradoxical observation has been attributed to individuals in the higher BMI groups having protective muscle reserves, having not yet reached adiposity levels high enough to increase mortality[500]. This highlights a potential deficiency of the weight component of the BMI equation in not discriminating between fat mass and fat-free mass.

When considering cancer prevention, there is a clear signal suggesting higher BMI individuals are at increased risk of some tumour types. For example, in post menopausal breast cancer, a recent study of over 1 million women highlights an increasing risk in breast cancer through increasing BMI groups, with BMI between 30–34.9 having a hazard ratio of 1.53 and a BMI of over 35 being having an adjusted hazard ratio of 1.65[501]. Incremental BMI levels in pre-menopausal women did not show statistically significant associations with the risk of breast cancer. Meta-analyses have identified high BMI as a particular risk factor for oestrogen/progesterone receptor-positive breast cancer in post-menopausal women (exemplar summary risk ratio:1.39 [1.14-1.70]) but not in oestrogen/progesterone negative breast cancer in post-menopausal women (exemplar summary risk ratio: 0.98 [0.78-1.22])[502,503]. Other studies have reported mixed results when looking at ER negative and triple receptor negative breast cancer in post-menopausal women with obesity[502,504–507]. In pre-menopausal women the relationship appears to be different to their post-menopausal counterparts[505,508,509], with even an inverse relationship with BMI being reported in some studies[505,510–512].

The literature on prostate cancer is also equivocal, with some studies demonstrating high BMI as a risk factor[513,514], some suggesting it could be protective[515,516] and some showing no difference[517–519]. Interestingly, recent work has identified an inverse (but non-linear) association between BMI and the controversial screening marker, prostate specific antigen (PSA). PSA in obese men on average is 12.9% lower than men of a normal weight[520]. As a result, it has been suggested that it might be beneficial to account for BMI when interpreting the results of a PSA test - although early trials using this adjustment have not shown added benefit[521].

Colorectal cancer appears to have different relationships with increasing BMI in men and women[522,523], although this is not true in all cohorts[524]. Increasing BMI in men is consistently associated with an increase in the risk of colorectal cancer (exemplar pooled

relative risk: 1.37 [1.21-1.56]). In women no significant association was consistently observed (exemplar pooled relative risk: 1.07 [0.97-1.18])[522], although some studies have shown that increased BMI may increase colorectal cancer risk in women[525,526].

In contrast to studies using BMI as an obesity parameter, a recent meta-analysis of waist circumference reveals that both men and women in the highest category of waist circumference (vs the lowest category) are at increased risk of colorectal cancer (men, pooled relative risk: 1.47 [1.25–1.73] and women, pooled relative risk: 1.30 [1.15–1.48])[527]. This hints at discrepancies when different surrogate measures of obesity are taken within the same cancer type.

In the aforementioned studies, it is not evident where exactly the BMI 'signal' originates within those individuals where increased BMI does associate with cancer risk, and/or whether the origin of this signal is the same across all tumour types. As in AF, fat-free mass may rather be a risk factor for some tumour types. Indeed, from Chapter 1 and *Sustaining Proliferative Signalling* as a key *Hallmark of Cancer*, it is feasible that pathways involved in increasing fat-free mass, and increased cell growth might also increase cancer risk. Perhaps there are signals coming from fat and fat-free mass both culminating to increase cancer risk. To understand cancer risk, It is important to identify the value and the precise origin of the BMI signal, as this may change assessment of obesity in future clinical trials, refocus our efforts when investigating the biological underpinnings of the link between obesity and cancer, and even influence public policy.

The UK Biobank presents an ideal opportunity to investigate the source(s) of the BMI signal, across a number of tumour types. The UK Biobank is one of the world's largest cohort studies of ~500,000 UK residents who volunteered to have their clinical, lifestyle, anthropometric and genetic data collected for research. Notably, the vast majority of participants had bioelectrical impedance measurements, detailing fat and fat-free mass

amounts and distribution, with linked cancer incidence data from a number of sources, including the well-curated UK cancer registries.

Herein we report a study of three cancers in the UK Biobank, which were the most common in participants following anthropometric assessment (>2,000 cases), enabling robust phenotypic and genetic analyses. This includes two cancers that have been associated with high BMI (post-menopausal breast and colorectal cancer), and one cancer that has shown mixed results with BMI over a number of studies (prostate cancer). We set out to unravel these associations. To date, studies in the UK Biobank have focussed on measures of adiposity without assessing the potential role of lean body mass[528,529]. Here we take a more comprehensive and perhaps more unbiased approach looking at the risk conferred by adiposity, lean body mass and body mass index using a genomic approach. Genomics approaches are used to extract signal from noise in large datasets, but as yet are not commonly used in epidemiological studies. We plan to use such techniques to tease apart the signal coming from the weight component of BMI and its association with cancer.

2.2 Methods

2.2.1 The UK Biobank

The UK Biobank is a cohort of ~500,000 UK residents who volunteered to have their clinical, lifestyle, anthropometric and genetic data collected for research. Data was collected in a number of ways, including face-to-face interview, touchscreen assessment and from centralised clinical registers (e.g. the Cancer register and Death register). Participants were aged between 40-69 years at recruitment. UK Biobank obtained informed consent from all participants and all protocols were approved by the National Research Ethics Service Committee. All baseline data used in this study was collected in 22 UK centres, between 2006-2010. This study was conducted under the UK Biobank approved application (#43313, PI Francesca Buffa).

2.2.2 Cancer data

Cancer occurrences were defined by the presence of a cancer international classification of diseases (ICD) code in the UK Cancer register or the UK Death register. To maximise the number of individual cancer cases, we combined ICD9 and ICD10 codes: this was carried out by a clinician to ensure the matching was accurate. We selected post-menopausal breast cancer, colorectal cancer and prostate cancer due to their differing relationships with BMI in previous studies and their relatively large number of cases since the measurement of anthropometric traits. To increase power, we combined the following ICD codes to identify patients with

(i) breast cancer (ICD-10: C50 (Malignant neoplasm of breast), D05 (Carcinoma in situ of breast), ICD-9: 174 (Malignant neoplasm of female breast), 2330 (Carcinoma in situ of breast)),

(ii) colorectal cancer (ICD-10: C18 (Malignant neoplasm of colon), D010 (Carcinoma in situ of colon), 2303 (Carcinoma in situ of colon), C19 (Malignant neoplasm of rectosigmoid junction), C20 (Malignant neoplasm of rectum), D011 (Carcinoma in situ of rectosigmoid junction), D012 (Carcinoma in situ of rectum), 2304 (Carcinoma in situ of rectum), ICD-9: 153 (Malignant neoplasm of colon), 1540 (Malignant neoplasm of rectosigmoid junction), 1541 (Malignant neoplasm of rectum)), and

(iii) prostate cancer (C61 (Malignant neoplasm of prostate), 185 (Malignant neoplasm of prostate), D075 (Carcinoma in situ of prostate)).

2.2.3 Anthropometry and body composition

During the baseline assessment, participants had various anthropometric traits measured directly (height, weight etc) or by bioelectrical impedance (e.g fat mass, fat-free mass etc). Bioimpedance data was measured using a Tanita BC418MA (Tanita, Tokyo, Japan) body composition analyser. Participants stood barefoot on the analyser and held the metal handles. This device produced whole body and segmental readings of fat mass and fat-free, measures generated include whole body fat mass (WBFM) and whole body fat-free mass (WBFFM) and fat mass and fat-free measurements for the right arm, right leg, left arm, left leg and trunk, as well as weight, body mass index and basal metabolic rate. This has been validated in several previous studies[530,531]. Further information on anthropometric data collected is found here[532]. The inter-relationship of all anthropometric measures were investigated through unsupervised clustering (supplementary figures S2.1-S2.4). In this chapter we concentrate primarily on the weight component of BMI comprising WBFM and WBFFM. Interestingly, weight fell into the overall cluster of fat-free mass in men and in the fat mass cluster in women.

2.2.4 Study population

The complete UK Biobank dataset includes 502,620 UK adults (229,165 men and 273,455 women). Individuals were excluded if at recruitment they had any prior cancer diagnosis, as one may expect a diagnosis of cancer to be associated with changes in body composition[533,534]. Participants were removed from the analysis if missing data on anthropometric trait measurements, sex, or age. For variables with missing data, to maximise our statistical power we encoded all missing values in categorical and binary covariates with a new variable. Where relevant, post-menopausal status was defined if an individual answered, “Yes” to the question “Have you had your menopause (periods stopped)?”. The final study population for the post-menopausal breast cancer analysis was 131,546 women (3,952 cases and 127,594 controls). The final study population for the prostate cancer analysis was 191,434 men (5,982 cases and 185,452 controls). The final study population for the colorectal cancer analysis was 218,954 women (1,424 cases and 217,530 controls) and 187,439 men (1,996 cases and 185,443 controls).

2.2.5 Epidemiological Analysis

Logistic regression analyses were carried out using R (version 3.5.3). Anthropometric measures were investigated as continuous and categorical variables. Baseline BMI/WBFM and WBFFM measures of study participants were divided into quintiles, and these were central to our analysis and trend tests (supplementary table S2.5). Correlations between anthropometric measures were calculated using Pearson and Spearman rank correlation coefficients (Spearman discussed below as most appropriate).

Observation time was ended by date of: death, diagnosis of the relevant cancer as outlined above in the Cancer *data* section or end of monitoring period (2016). All exposures were studied individually in a univariate setting and in multiply-adjusted models. In the reductionist models, cancer occurrence was examined using either BMI, WBFM and WBFFM accounting for only age of attendance to the assessment centre (5 year age group). Within the multivariable models we accounted for well-accepted cofounders selected a priori. Post-menopausal breast cancer models were adjusted for age of attendance to assessment centre (5 year age group), Townsend Deprivation Index, ethnicity, alcohol intake, smoking status, family history of breast cancer, diabetes mellitus, parity, oral contraceptive use, use of hormone replacement therapy, age at menarche, age at first full-term birth, physical activity (MET in hours per week) and age at menopause. Multivariable prostate cancer models were adjusted for age of attendance to assessment centre (5 year age group), Townsend Deprivation Index, ethnicity, alcohol intake, smoking status, family history of prostate cancer, red meat intake (beef, lamb and mutton), fish intake (oily fish), diabetes mellitus, physical activity (MET in hours per week), history of prostatitis, and regular use of aspirin, statins, saw palmetto, metformin, selenium supplements, Vitamin D supplements, Vitamin E supplements, testosterone supplements, use of anabolic steroids and 5 alpha-reductase inhibitors. Multivariate colorectal cancer models were adjusted for age of attendance to

assessment centre (5 year age group), Townsend Deprivation Index, ethnicity, alcohol intake, smoking status, family history of bowel cancer, red meat intake (beef/lamb/mutton), physical activity (MET in hours per week), diabetes mellitus, regular use of aspirin, statins, metformin, Vitamin D supplements in men and the same covariates plus menopausal status and use of hormone replacement therapy in women. To assess the impact of missing values, a sensitivity analysis was performed, on participants who had values for all covariate variables, results were not significantly different to those presented below.

Principal component analysis (PCA) was carried out using Python (v.3.7). PCA is an adaptive exploratory method that serves as a descriptive tool that can be applied to a variety of data types. PCA needs no distributional assumptions[535] and is commonly used to explore genomic datasets (e.g. microarray and RNA-Seq). Investigating associations with principal components themselves has yielded new biological insights[536] but as yet, PCA is rarely used in epidemiological studies. Here we use PCA to attempt to untangle the relationship between different components of body composition and cancer risk.

N.B. A note on logistic regression versus Cox proportional-hazards model

In Cox proportional-hazards models, it is common practice to use age as an underlying timescale and also the categorical age of the individuals in the analysis. The inclusion of age into the model through two separate variables raises a possibility of over correction for age. In logistic regression, as detailed below, either categorical age of the individuals **or** continuous age of individuals can be used. The results for logistic regression presented below use categorical age, although analyses carried out with continuous age gave indistinguishable results (data not shown). An additional complication is that age and body composition are likely to be linked, body composition in this cohort will likely vary with age as shown in previous studies[537], further making correction of the models complex. After much discussion, we carried out both Cox and logistic analyses but here present logistic

regression (a type of general linear model) below to obviate possible over-correction by age. Further, we elected not to correct WBFM models for WBFFM/BMI or vice versa due to their significant intercorrelation.

2.3 Results

Logistic regression models were used to investigate the relationship between each of body mass index (BMI), Whole Body Fat-Free Mass (WBFFM) and Whole Body Fat Mass (WBFM) and the risk of post-menopausal breast, prostate and colorectal cancer. Quintiles were compared to the lowest (1st quintile) to give odds ratios for cancer risk. Reductionist and multivariable models yielded similar results and so the multivariable analysis is presented below.

2.3.1 Post-menopausal breast cancer

Our BMI, WBFM and WBFFM logistic regression models to predict post-menopausal breast cancer had several significant covariates, some which just dropped in or out of statistical significance depending on the anthropometric trait tested (supplementary tables S2.6-2.8). Age categories were significant as well as use of hormone replacement therapy, family history of breast cancer, and certain categories in number of births (3 and >5), age of first birth (>30 years), age of menarche (5-10 years), age of last menstrual period (30-40, 40-45, 45-50 years), Townsend deprivation index (fifth quintile) and amount of exercise (fifth quintile). The fourth quintile of exercise becomes significant in the WBFFM model.

Higher quintiles of BMI, WBFFM and WBFM were associated with increased risk of post-menopausal breast cancer (Fig. 2.1). As shown in Fig. 2.1, BMI, WBFM and WBFFM were highly significantly correlated ($p < 0.001$) using Spearman's rank correlation coefficient, details seen in supplementary table S2.9. Notably, BMI and WBFM had an r value of 0.92 whereas BMI and WBFFM had an r value of 0.62. This suggests that WBFM and WBFFM are relatively distinct parameters.

To attempt to untangle the relationship between different components of body composition and cancer, Principal Component Analysis (PCA) including both fat-free and fat

mass measures from the whole body, limbs and trunk was employed. This allows us to identify the main sources of signal affecting cancer risk across three different tumour types.

In post-menopausal breast cancer, the first two PCs explain ~95% of variance (Fig. 2.2a) coming from WBFM and WBFFM. Interestingly, the highest quintile of PC1 gives an increased risk of post-menopausal breast cancer compared to BMI, WBFM or WBFFM alone (Fig 2.2b). Most of the contribution of the principal component 1 (PC1) is given by whole body fat mass, trunk fat mass and whole body fat-free mass (Fig. 2.2c). Principal component 2 (PC2) mainly consists of WBFFM and trunk fat-free mass and shows a negative relationship with WBFM and trunk fat mass (PC2 explained variance = ~12%). The highest quintile of PC2 also appears to carry an increased risk for post-menopausal breast cancer.

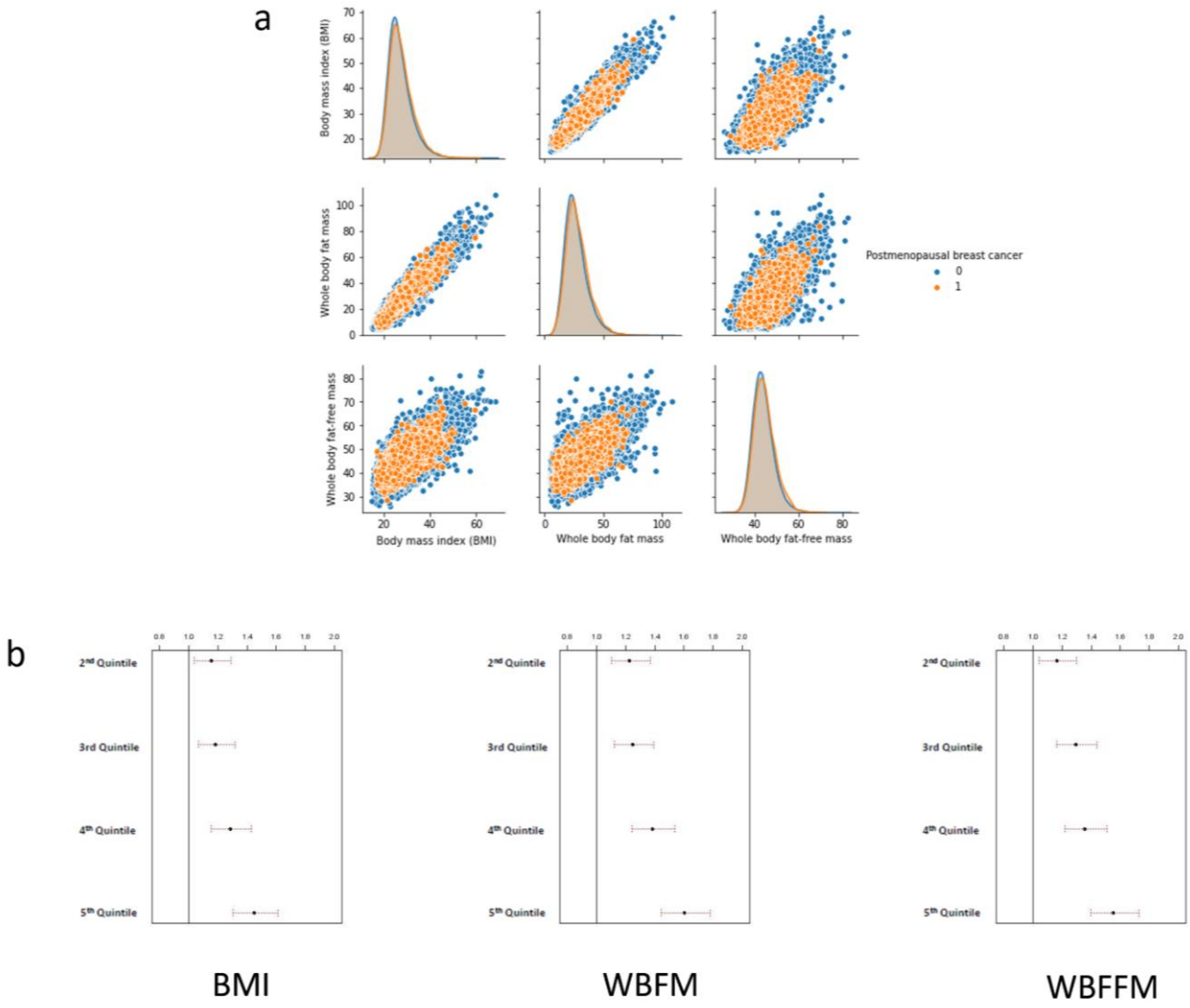
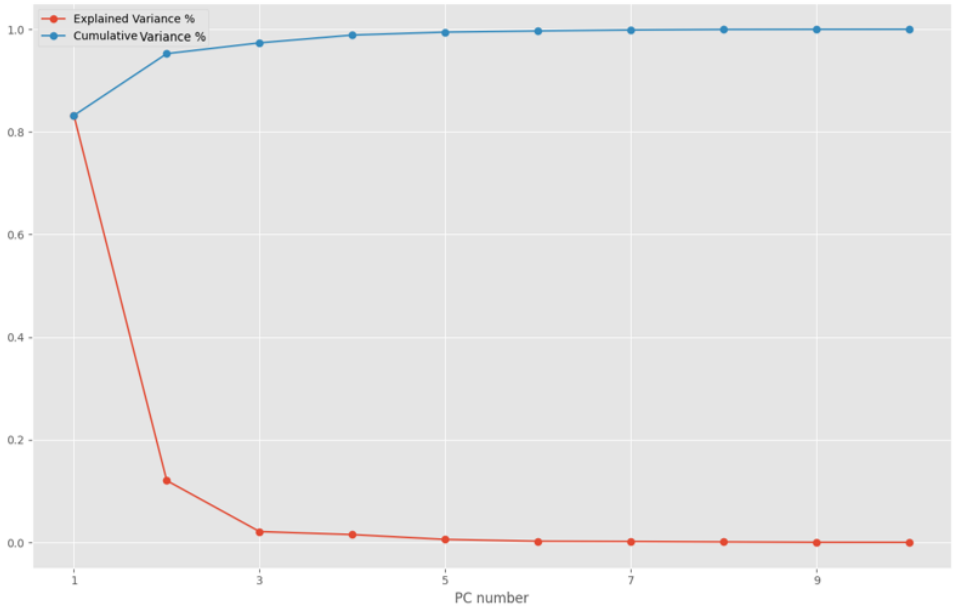
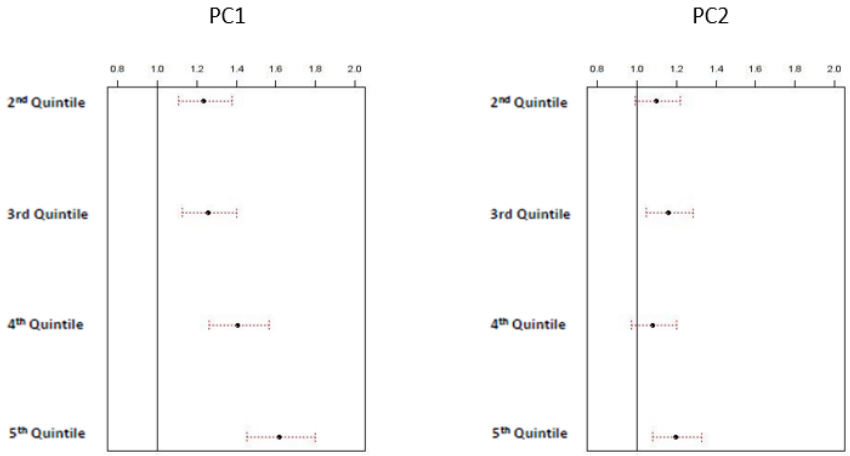


Figure 2.1: *Anthropometric traits in the post-menopausal breast cancer cohort.*
a) Data distribution of BMI, WBFM and WBFFM. In the dot plots, the axes correspond to units of measure BMI [kg/m^2], WBFM [kg], WBFFM [kg] and red dots = cancer cases. Univariate plots (on the diagonal) represent the distribution of measures within the population with the horizontal axis = the anthropometric measure with aforementioned units and vertical axis as frequency count b) risk of post-menopausal breast cancer (odds ratio with 95% confidence interval) in the different quantiles of BMI, WBFM and WBFFM compared to the lowest quintile (OR = 1). n = 131,546 women (3,952 cases and 127,594 controls).

a



b



c

	PC1	PC2
WBFM	0.81	-0.28
WBFFM	0.32	0.78
Leg fat mass (right)	0.14	-0.01
Leg fat-free mass (right)	0.07	0.12
Leg fat mass (left)	0.14	-0.01
Leg fat-free mass (left)	0.07	0.11
Arm fat mass (right)	0.06	0.00
Arm fat-free mass (right)	0.02	0.04
Arm fat mass (left)	0.06	0.00
Arm fat-free mass (left)	0.02	0.04
Trunk fat mass	0.41	-0.25
Trunk fat-free mass	0.14	0.47

Figure 2.2: *Principal component analysis of post-menopausal breast cancer risk.*
Panel A: normalised percentage of variance explained by each principal component. Panel B: risk of post-menopausal breast cancer in the different quintiles of principal component 1 (PC1) and principal component (PC2) compared to the lowest quintile (OR with 95% confidence interval). Panel C: loading values for anthropometric traits in PC1 and PC2. Cohort as in Fig. 2.1.

2.3.2 Prostate

In the male population, our BMI, WBFM and WBFFM logistic regression models to predict prostate cancer had several significant covariates, some which just dropped in or out of statistical significance depending on the anthropometric trait tested. As well as the increasing 5 year age categories, the following covariates were significant in the BMI, WBFM and WBFFM models: ethnicity (Indian, Caribbean, African) townsend deprivation Index (4th quantile), lamb and mutton intake (Never), regular supplementation with saw palmetto or vitamin E, diabetes diagnosed by doctor, family history of breast or prostate cancer. Other variables that had variable significance included ethnicity (White and Black African), smoking status (Never, information not available), diabetes diagnosed by doctor (information not available), the third and fourth quantiles of exercise. Further details can be seen in supplementary tables S2.10-2.12. Notably, exercise appears most significant in the WBFFM model (e.g. 4th quantile, p=0.0390, beta = 0.0959 [SE = 0.0465]).

The upper quantiles of BMI ($p = 7.53E^{-05}$, beta = -0.181 [SE = 0.045597], equating to a odds ratio of 0.835 [0.7634-0.91281]) and WBFM ($p = 5.22E^{-05}$, beta = -0.184 [SE = 0.0454], equating to a odds ratio of 0.832 [0.7613-0.9096]) are significant and appear protective. Whereas the second and third quintiles are significant in WBFFM (Fig. 2.3). Correlations between the different anthropometric traits are detailed in supplementary table S2.9. WBFM and BMI had a slightly less positive coefficient than in the post-menopausal breast cancer cohort, r = 0.89, and the correlation coefficient between WBFM and WBFFM was 0.53 ($p<0.001$).

In prostate cancer, the first two PCs explain ~93% of variance (Fig. 2.4a). Most of the contribution of the PC1 is given by WBFM and WBFFM. The 5th quintile just reaches significance as a protective factor, but the confidence intervals around the OR are very close to including 1 (Fig. 2.4b). PC2 (~18% of explained variance) shows differential directional

relationships with WBFM and WBFFM (Fig. 2.4c). Interestingly, the last quantile in PC2 is similarly protective as the highest quintile of whole body fat mass.

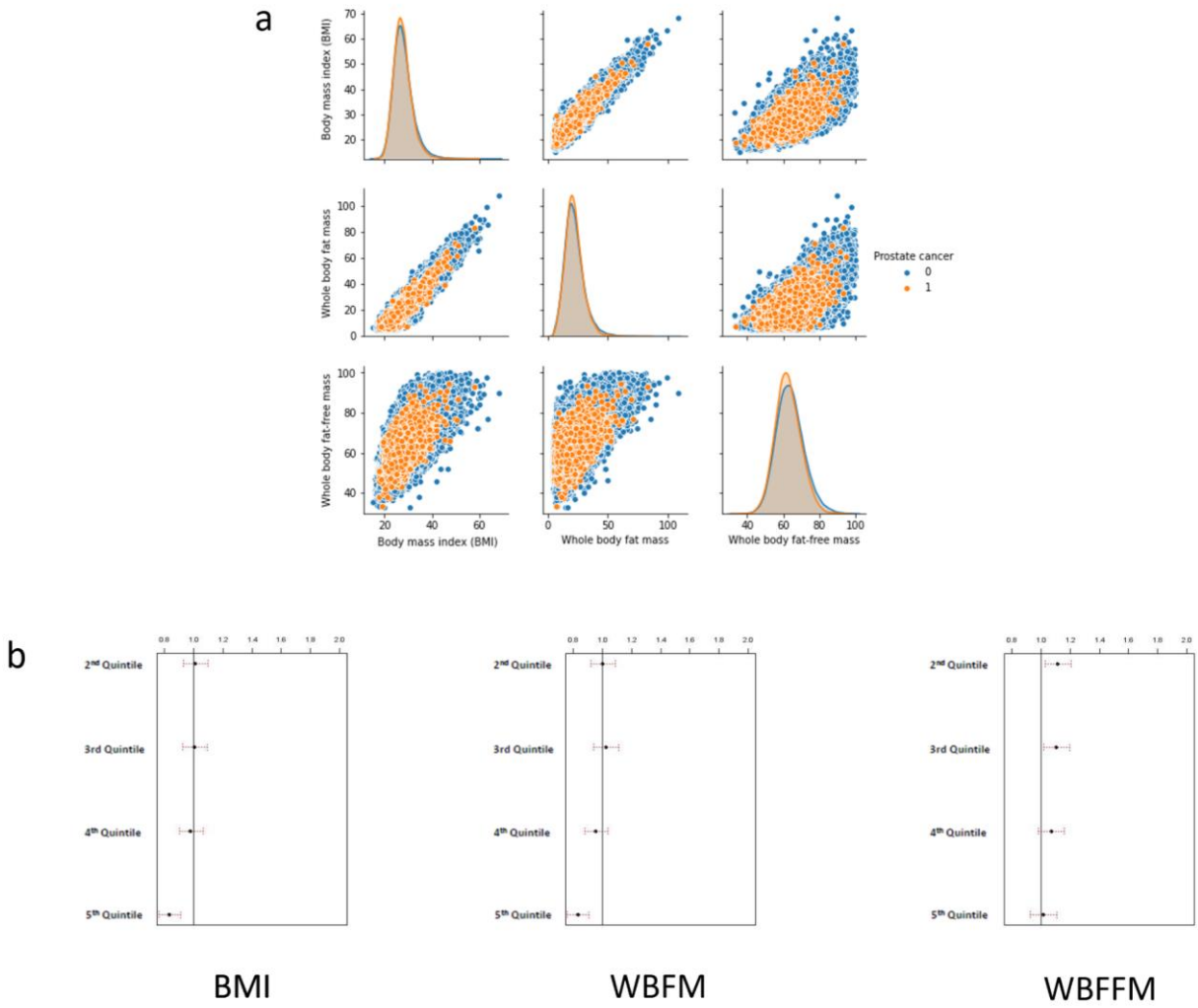
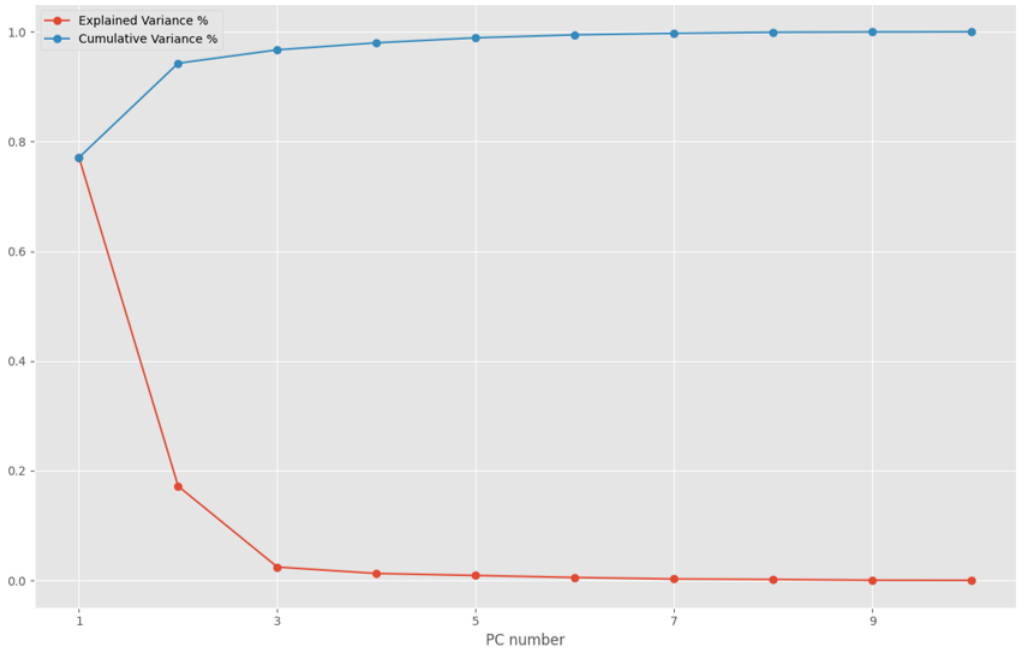
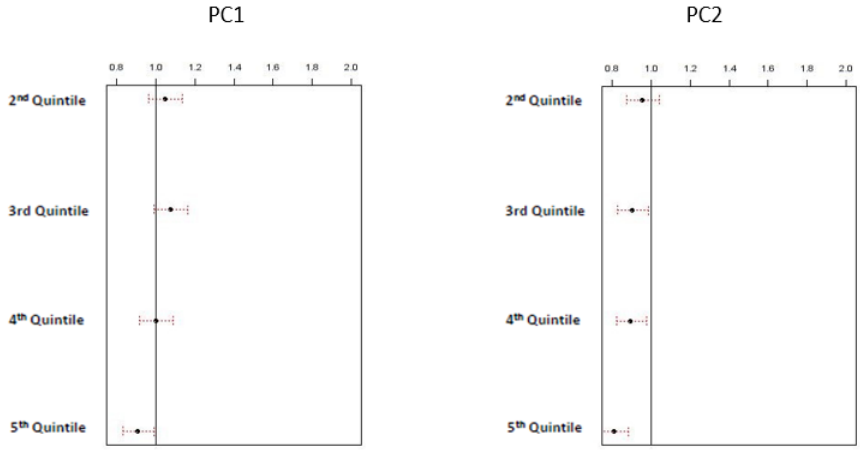


Figure 2.3: *Anthropometric traits in prostate cancer cohort.*
a) Data distribution of BMI, WBFM and WBFFM. In the dot plots, the axes correspond to units of measure BMI [kg/m2], WBFM [kg], WBFFM [kg] and red dots = cancer cases. Univariate plots (on the diagonal) represent the distribution of measures within the population with the horizontal axis = the anthropometric measure with aforementioned units and vertical axis as frequency count b) risk of post-menopausal breast cancer (odds ratio with 95% confidence interval) in the different quantiles of BMI, WBFM and WBFFM compared to the lowest quintile (OR = 1). n = 191,434 men (5,982 cases and 185,452 controls).

a



b



c

	PC1	PC2
WBFM	0.64	0.54
WBFFM	0.57	-0.64
Leg fat mass (right)	0.09	0.08
Leg fat-free mass (right)	0.11	-0.08
Leg fat mass (left)	0.09	0.07
Leg fat-free mass (left)	0.12	-0.07
Arm fat mass (right)	0.04	0.02
Arm fat-free mass (right)	0.04	-0.04
Arm fat mass (left)	0.04	0.02
Arm fat-free mass (left)	0.04	-0.05
Trunk fat mass	0.38	0.34
Trunk fat-free mass	0.26	-0.40

Figure 2.4: *Principal component analysis of prostate cancer risk.*
Panel A: normalised percentage of variance explained by each principal component. Panel B: risk of prostate cancer in the different quintiles of principal component 1 (PC1) and principal component (PC2) compared to the lowest quintile (OR with 95% confidence interval). Panel C: loading values for anthropometric traits in PC1 and PC2. Cohort as in Fig. 2.3.

2.3.3 Colorectal

We investigated colorectal cancer in a sex-specific analysis as previous work has shown different associations in men and women[538]. In our BMI, WBFM and WBFFM logistic regression models to predict colorectal cancer in the female population in the UK Biobank, the significant covariates reassuring were similar in all models. As well as the increasing 5 year age categories these were significant: family history of bowel cancer (BMI p =0.03, beta = 0.165 [SE = 0.0782], WBFM p =0.03, beta = 0.166 [SE = 0.0782] and WBFFM p =0.03, beta = 0.165 [SE = 0.0781]), the upper quantile of exercise (BMI p =0.0417, beta = -0.182 [SE = 0.0892], WBFM p =0.0486, beta = -0.176 [SE = 0.0893] and WBFFM p =0.0478, beta = -0.176 [SE = 0.0891]). The second highest quintile of exercise almost reached significance in (BMI p =0.0852, beta = -0.154 [SE = 0.0895], WBFM p =0.0936, beta = -0.150 [SE = 0.0894] and WBFFM p =0.0927, beta = -0.150 [SE = 0.0893]). Also “prefer not to answer” to *Has a doctor ever told you you have diabetes?* appeared tended towards significance, perhaps suggesting a bias in the population answering this question in this way, although the standard errors on the estimates were relatively large (BMI p =0.0710, beta = 1.469 [SE = 0.814], WBFM p =0.0775, beta = 1.43 [SE = 0.812] and WBFFM p =0.0800, beta = 1.420 [SE = 0.811]). Full details in supplementary tables S2.13-2.15. In terms of anthropometric traits, all confidence intervals included 1 or were very close to including 1 for all anthropometric measures (see Fig. 2.5). Correlations between the different anthropometric traits are illustrated in Fig. 2.5 detailed in supplementary table S2.8. As seen in the post-menopausal cohort correlation analysis, BMI and WBFM are strongly correlated (*r* value of 0.93) while BMI and WBFFM are less so (*r* value of 0.63).

The first two PCs in colorectal cancer in women explains ~95% of variance (Fig. 2.6a). Different quantiles of PC1 do not appear to carry significantly different risk for colorectal cancer (Fig. 2.6b). PC2 shows differential directional relationships with WBFFM and WBFM. The last quintile of PC2 is associated with a mildly increased risk for colorectal

cancer, which is difficult to interpret alone (Fig. 2.6b). Most of the contribution of the PC1 is given by WBFM and WBFFM, with WBFM showing a larger loading value (0.809, Fig. 2.6c) compared to its contribution to PCA1 in men (0.641).

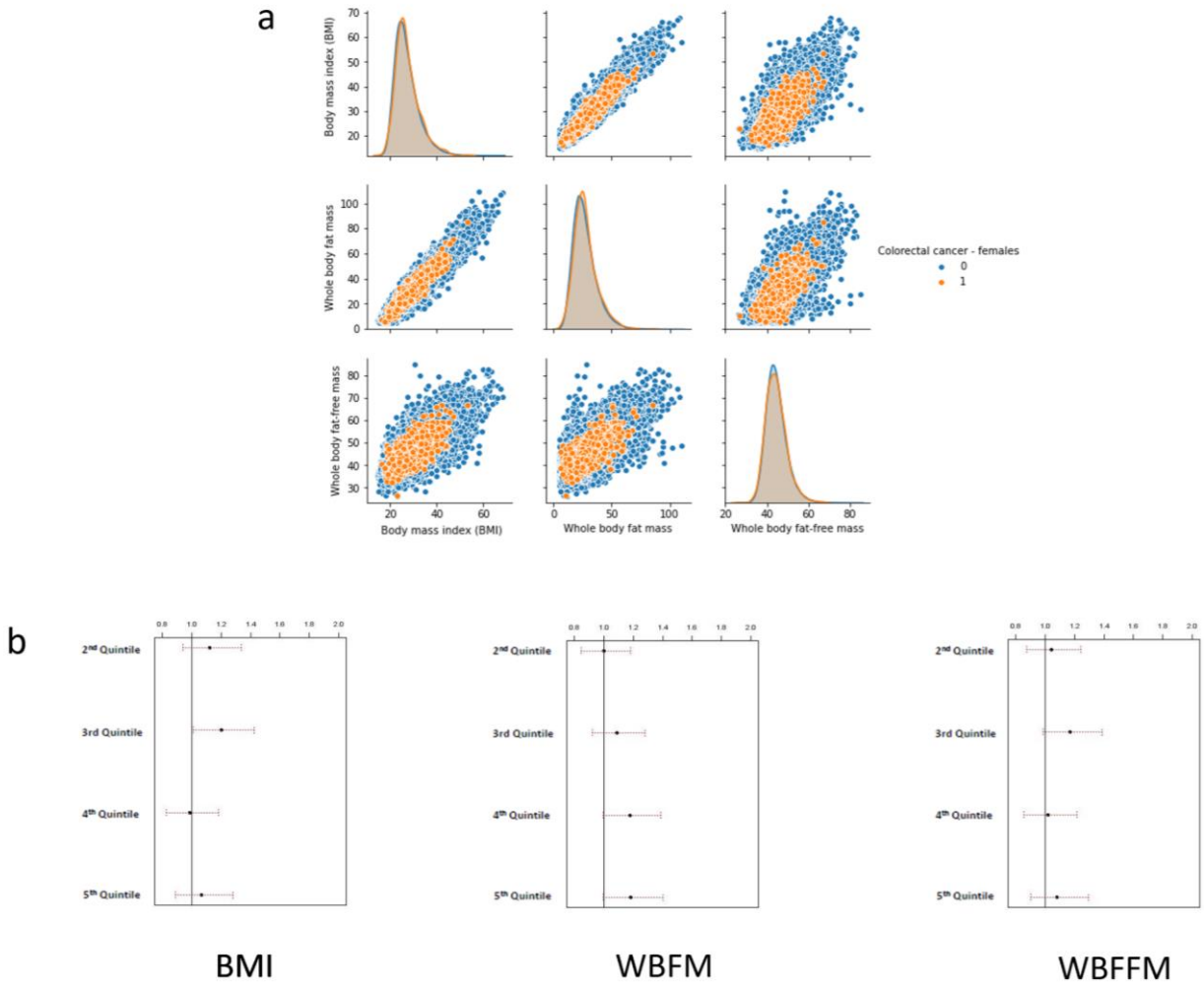
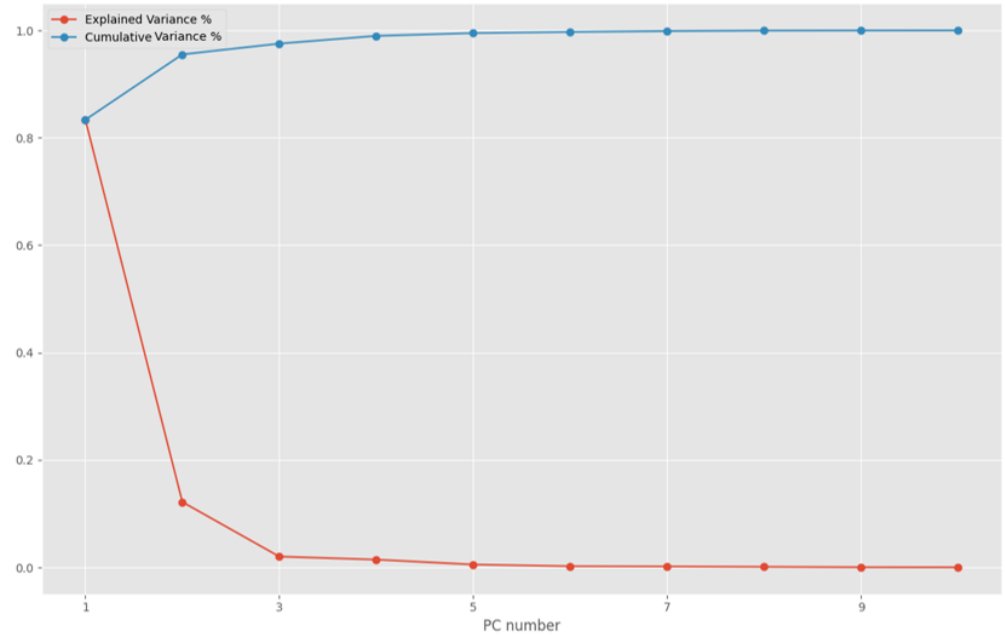
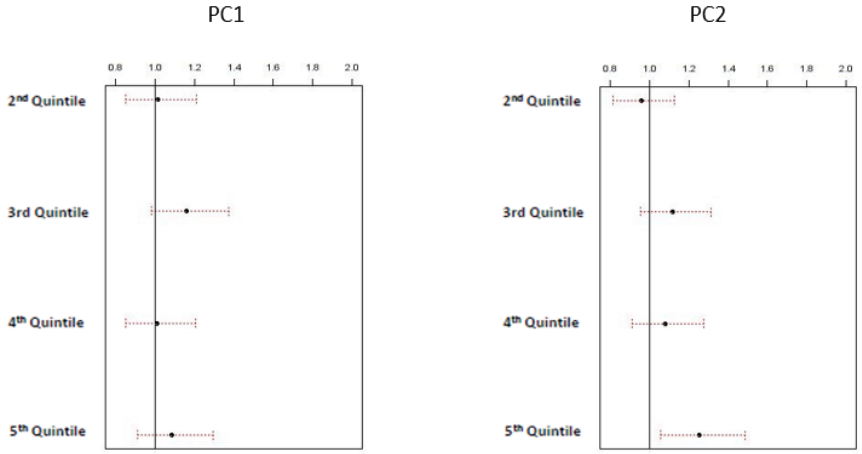


Figure 2.5: *Anthropometric traits in women in colorectal cancer cohort.*
a) Data distribution of BMI, WBFM and WBFFM. In the dot plots, the axes correspond to units of measure BMI [kg/m2], WBFM [kg], WBFFM [kg] and red dots = cancer cases. Univariate plots (on the diagonal) represent the distribution of measures within the population with the horizontal axis = the anthropometric measure with aforementioned units and vertical axis as frequency count b) risk of post-menopausal breast cancer (odds ratio with 95% confidence interval) in the different quantiles of BMI, WBFM and WBFFM compared to the lowest quintile (OR = 1). n = 218,954 women (1,424 cases and 217,530 controls).

a



b



c

	PC1	PC2
WBFM	0.81	-0.28
WBFFM	0.31	0.78
Leg fat mass (right)	0.14	-0.01
Leg fat-free mass (right)	0.06	0.12
Leg fat mass (left)	0.14	-0.01
Leg fat-free mass (left)	0.07	0.11
Arm fat mass (right)	0.06	0.00
Arm fat-free mass (right)	0.02	0.04
Arm fat mass (left)	0.06	0.00
Arm fat-free mass (left)	0.02	0.04
Trunk fat mass	0.41	-0.24
Trunk fat-free mass	0.14	0.46

Figure 2.6: *Principal component analysis of colorectal cancer risk in women.*
Panel A: normalised percentage of variance explained by each principal component. Panel B: risk of colorectal cancer in the different quintiles of principal component 1 (PC1) and principal component (PC2) compared to the lowest quintile (OR with 95% confidence interval). Panel C: loading values for anthropometric traits in PC1 and PC2. Cohort as in Fig. 2.5.

In the male population, the BMI, WBFM and WBFFM logistic regression models to predict colorectal cancer had significant covariates that were similar to what was seen in the female population. These were the increasing 5 year age categories and family history of bowel cancer (BMI p =1.72e-06, estimate = 0.296 [SE = 0.0619], WBFM p =1.78e-06, estimate = 0.295 [SE = 0.0619] and WBFFM p =1.56e-06, estimate = 0.297 [SE = 0.0619]). However, unlike in women, increasing exercise categories were not significant. Other categories that were significant in the models were, certain ethnicities (Irish, Mixed, Any other mixed background [categories as determined by the UK BioBank Consortium]), smoking status (Never), alcohol consumption (Never) and beef intake (2-4 times a week) and lamb and mutton intake (once or more daily). More details can be found in supplementary tables S2.17-2.19. Regular use of the pharmacological agents aspirin (BMI p =0.0732, estimate = -0.111 [SE = 0.0621], WBFM p =0.0782, estimate = -0.109 [SE = 0.0621] and WBFFM p =0.0955, estimate = -0.104 [SE = 0.0621]) and metformin (BMI p =0.0875, estimate = 0.252 [SE = 0.147], WBFM p =0.0898, estimate = 0.250 [SE = 0.147] and WBFFM p =0.0776, estimate = 0.260 [SE = 0.147]) tended towards significance. As did the alcohol consumption categories, daily or almost daily (p values =0.0747-0.0931, estimate ≅ 0.10 , SEs ≅ 0.060 in all models) and one to three times a month (p values =0.0892-0.0995, estimate ≅ -0.159 to -0.164 , SEs ≅ 0.097 in all models).

In terms of anthropometric traits, the third, fourth and fifth quantile of BMI and WBFM reached significance, each with increasing risk of colorectal cancer (see Fig. 2.7). The third and fifth quantiles of WBFFM also carried significantly increased risk of colorectal cancer and p trend < 0.01 (see Fig. 2.7 and table S2.5). Correlations between the different anthropometric traits are detailed in table S2.10. Correlation coefficients were very similar to the prostate cancer analysis cohort.

In colorectal cancer in men, the first two PCs explain ~92% of variance (Fig. 2.8a). Increasing quantiles of PC1 appear to confer increased risk. PC2 (~18% of explained

variance) shows differential directional relationships with WBFM and WBFFM, but different quantiles of PC2 do not appear to carry significantly different risk for colorectal cancer (Fig. 2.8b). The trend in PC1 quantiles and colorectal cancer risk appears more stepwise than WBFM or WBFFM alone. Most of the contribution of the PC1 is given by WBFM and WBFFM (Fig. 2.8c), perhaps pointing to both factors having some element of risk.

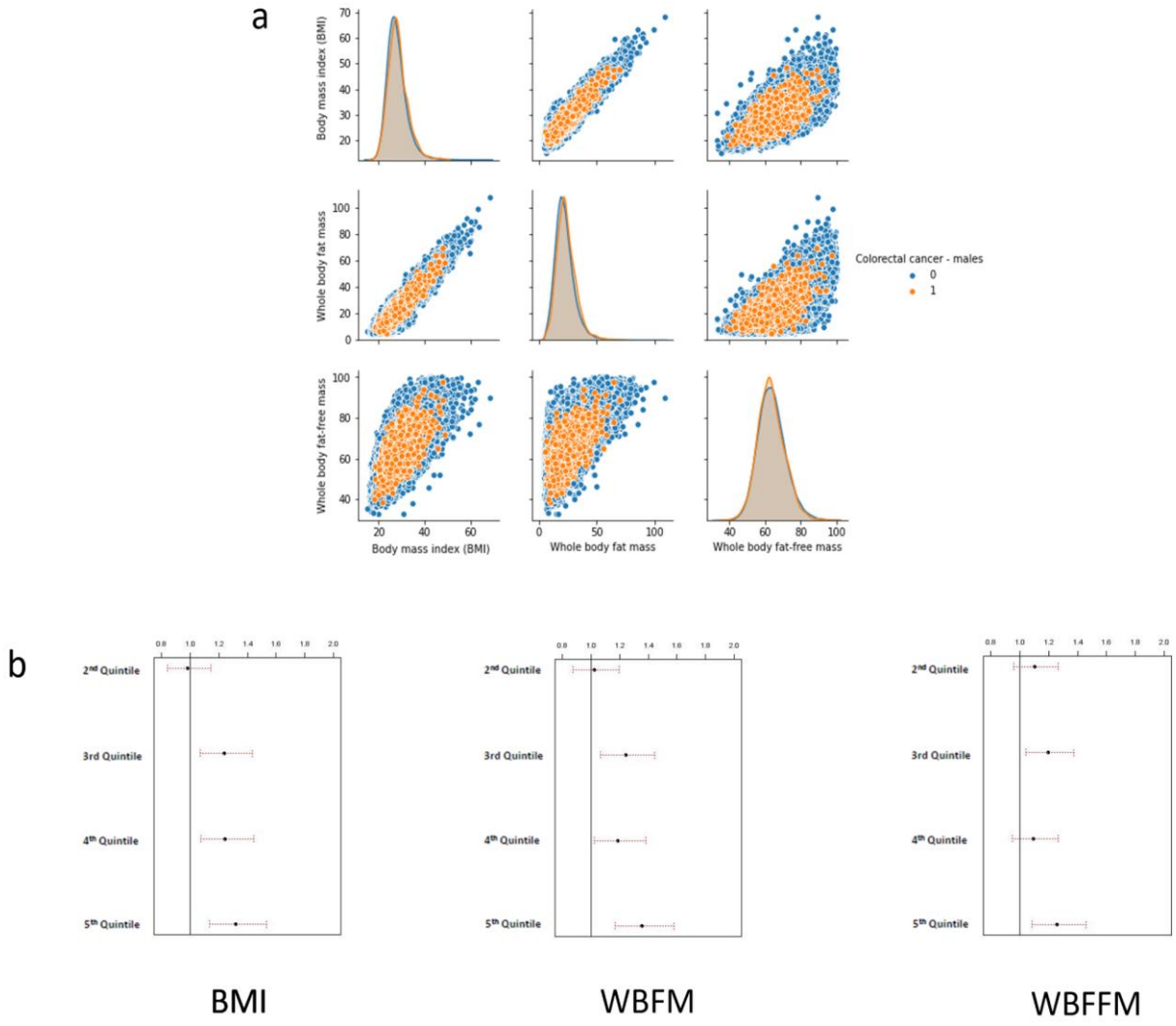
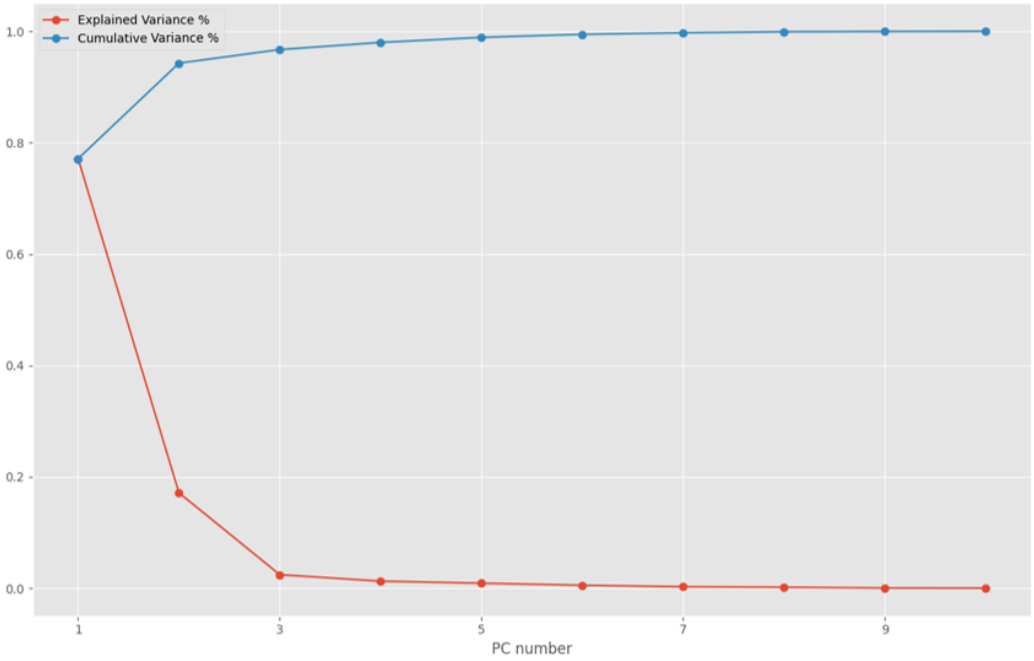


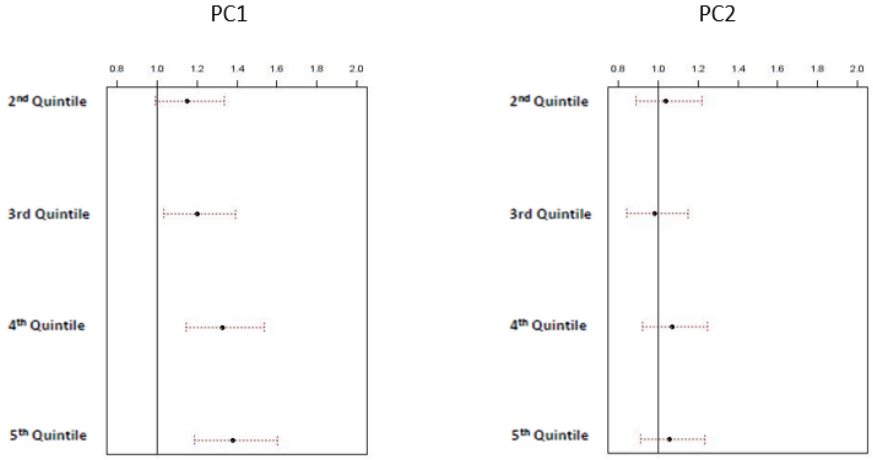
Figure 2.7: *Anthropometric traits in men in colorectal cancer cohort.*

a) Data distribution of BMI, WBFM and WBFFM. In the dot plots, the axes correspond to units of measure BMI [kg/m^2], WBFM [kg], WBFFM [kg] and red dots = cancer cases. Univariate plots (on the diagonal) represent the distribution of measures within the population with the horizontal axis = the anthropometric measure with aforementioned units and vertical axis as frequency count b) risk of post-menopausal breast cancer (odds ratio with 95% confidence interval) in the different quantiles of BMI, WBFM and WBFFM compared to the lowest quintile (OR = 1). n = 187,439 men (1,996 cases and 185,443 controls).

a



b



c

	PC1	PC2
WBFM	0.64	0.54
WBFFM	0.57	-0.64
Leg fat mass (right)	0.09	0.08
Leg fat-free mass (right)	0.11	-0.08
Leg fat mass (left)	0.09	0.07
Leg fat-free mass (left)	0.12	-0.07
Arm fat mass (right)	0.04	0.02
Arm fat-free mass (right)	0.04	-0.04
Arm fat mass (left)	0.04	0.02
Arm fat-free mass (left)	0.04	-0.05
Trunk fat mass	0.38	0.34
Trunk fat-free mass	0.26	-0.40

Figure 2.8: *Principal component analysis of colorectal cancer risk in men.*
Panel A: normalised percentage of variance explained by each principal component. Panel B: risk of colorectal cancer in the different quintiles of principal component 1 (PC1) and principal component (PC2) compared to the lowest quintile (OR with 95% confidence interval). Panel C: loading values for anthropometric traits in PC1 and PC2. Cohort as in Fig. 2.7.

2.4 Discussion

2.4.1 BMI, WBFM and WBFFM across three cancer types

The origin of the BMI signal and its association with cancer across three tumour types was investigated. Similar associations with BMI are replicated from other studies[501,529,522,523,501,529]. In addition, different risk signals from the weight component of BMI (WBFM and WBFFM) in the different tumour types are identified.

Looking at WBFM and WBFFM individually, it is interesting to note that the upper two quintiles of WBFFM carried higher risk for post-menopausal breast cancer compared to WBFM, or even BMI, perhaps suggesting a signal coming from both components. In contrast, the prostate cancer analysis shows that individuals in the highest quintile of BMI were at lower risk of prostate cancer. From the WBFM and WBFFM analyses, this protective effect appears to be originating from WBFM. The previously described complex relationship of BMI with colorectal cancer was also dissected in our analyses. The risk in women showed no clear relationship with WBFM or WBFFM, whereas in men, the highest quintiles of WBFM and WBFFM both appear to carry increased risk, perhaps explaining why BMI shows the clearer relationship (Fig. 2.9).

The data were further analysed using principal component analysis (PCA), to help untangle the weight component of the BMI. These results suggest PCA is a worthwhile method to employ in epidemiological studies such as this. Indeed, PCA appears to help to further separate and explain signals from WBFM and WBFFM. Reassuringly, PCA corroborates the epidemiological models of WBFM and WBFFM in all cancers.

PCA suggests risk signals coming from both WBFM and WBFFM in post-menopausal breast cancer, with the highest quintile of PC1 (comprising both WBFFM and WBFM) carrying a larger increased risk of post-menopausal breast cancer than WBFM or WBFFM

alone. In the PCA for prostate cancer, the upper quantiles of PC2 yield an odds ratio <1, supporting the finding, above, that higher WBFM appears to be protective from prostate cancer. In colorectal cancer, PCA suggests that both WBFM and WBFFM might contribute to risk in men, as in the analysis PC1 appears to have more predictive value than WBFM or WBFFM alone. In women, the relationship overall remains unclear but the uppermost quintile of PC2, which has a high positive value for WBFFM and negative for WBFM, is associated with a mildly increased risk for colorectal cancer.

One complexity in these studies is the relationship of variables with time and thus age. Indeed, the simplest of covariates, such as statin use, is likely to be commoner in those of a greater vintage than those younger in this cohort. Anthropometric traits, such as height and weight, are also likely to change with time. And, although age is incorporated into the model, it is impossible to disentangle its effects from all the other variables, as age by its very nature encompasses a plethora of factors. Interestingly, it has been shown menopause accelerates the fat mass gain and lean mass loss[539,540], making the findings relating to WBFFM in post-menopausal breast cancer even more pronounced.

This work shows WBFFM is associated with one's risk of certain tumours but not others. Coupled with the fact that both WBFM and WBFFM perform differently in the same tumour type, this finding indicates that the effect of WBFFM is unlikely to be simply solely due to its intercorrelation with WBFM. There is, however, a significant correlation between WBFM, WBFFM and BMI - which is why these are not included together in any individual analysis.

WBFFM and post-menopausal breast cancer have been linked before in smaller studies[495,541]. MacInnis et al carried out a prospective cohort study of 13,598 post-menopausal women in Australia (357 cases of post-menopausal breast cancer). This study, despite the small number of cases, found that non-adipose mass, adipose mass, and central

adiposity were all positively associated with breast cancer rates in postmenopausal women. Mellemkjær et al studied a prospective cohort of 23,788 postmenopausal women included in the Danish study Diet, Cancer, and Health during 1993 to 1997 and these women were assessed for cancer between 1993 to 2002. During this follow up period, 633 cases of breast cancer were recorded. Here, whole body fat-free and fat mass measures were normalised for height as fat-free mass index (FFMI) and fat-mass index (FMI). These workers focussed on splitting the cohort by HRT use, rather than using our approach of integrating it as a variable into the model. They found no association between FMI and breast cancer among postmenopausal women who had never used HRT, but a weak positive association with FFMI (although their conclusions were not strong as the 95% confidence intervals of the incidence rate ratios in the upper quantiles of FMI and FFMI both included 1). This might reflect the effect of normalising by height, which carries with it some benefits and drawbacks. Normalising by height makes a measure more similar to the BMI. However due to the actions of fat outlined in *Chapter 1*, it was felt that the raw values for WBFM might be more biologically relevant, whilst WBFFM and WBFM are already scaled for height and age[542]. Mellemkjær et al[541] indeed urged the need for further attention on fat-free mass was urged, as did James et al[543] in a recent review. This study was much larger, with 131,546 post-menopausal women (3,952 incident cases) and the effects of WBFFM, as well as WBFM, become evident.

Several theories have been postulated as to how increased WBFFM may influence cancer risk. One theory involves increased WBFFM being another source of endogenous growth promoting oestrogens, with their synthesising enzyme, aromatase, found in skin fibroblasts, bone, brain and skeletal muscle[541,544,545] as well as in adipocytes. Another hypothesis is that WBFFM reflects nutrition during childhood and adolescence, and hence relates to the action of growth factors including androgens and insulin-like growth factor (IGF-1)[495,546,547]. The role of IGF-1 has been extensively debated but is thought to be pro-

tumorigenic and has been linked to a number of different cancers (including breast and colorectal cancer)[548,549], further discussed in *Chapter 1*. Another factor related to a hormonal axis is the hormone insulin, which causes myogenesis[550] and has been linked to cancer risk[118,551], again discussed in *Chapter 1*. Alternatively, WBFFM may be related to mammographic density which has been described as a risk factor for post-menopausal breast cancer. Indeed, total body adiposity and body fat distribution appear to be independently inversely associated with breast density while breast density appears independent of exercise[552,553], but the relationship of breast density with WBFFM is currently not well-elucidated. Another possibility is that the genomic make up of an individual that influences cancer risk might also simultaneously influence WBFFM.

The increased risk with increased fat-mass (WBFM), demonstrated for post-menopausal breast cancer and colorectal cancer in men, has been attributed to a variety of factors. One is adipocyte hypertrophy and death, which leads to systemic chronic, subclinical inflammation of adipose tissue[333]. This inflammatory cascade releases a number of pro-tumorigenic mediators, which can act locally or can even travel systemically, including IL-6, IL-1RA, Il-18[554]. Leptin and oestrogen released by adipocytes, and increased circulating insulin caused by insulin resistance, may also act as pro-tumour growth factors[16,18,22]. Various theories on how adiposity is linked to carcinogenesis are discussed in *Chapter 1*. However, the protective effect seen by WBFM in prostate cancer, at least in this age group, suggests that this factor is not a universal driver in cancer. This potential protective effect of fat-mass in prostate cancer has been reported previously,[515,516] and previous studies have reported slightly lower prostate specific antigen (PSA) concentrations in men with high BMI[555]. Hypotheses to explain this apparently protective effect include detection bias with obese men less likely to undergo a screening for prostate cancer[556]. However, other hypotheses include a more biological explanation, including hormonal and metabolic alterations in obese men. For instance, obesity is associated with decreased serum

testosterone, and testosterone is heavily linked with prostatic growth and proliferation[557,558]. The androgen hypothesis and testosterone supplementation is much debated in the prostate cancer community[559–561], as much work on the effects of testosterone on proliferation comes from model systems.

2.4.2 Interpretation of estimates of WBFFM and WBFM

In this work, bioelectrical impedance is used to measure fat mass (WBFM) and fat-free mass (WBFFM) in the body. Its accuracy and reproducibility have improved markedly as technological advances have occurred. Nonetheless the accuracy of these measurements (relative to other techniques) is debated, as they can be affected by hydration status, impact of exercise, fluid retention (even to the level of the menstrual cycle), ambient temperature and differences between ethnicities [562–564]. These effects may differ between measuring instruments. In the UK Biobank data, the Tanita BC 418 MA was used. It yielded similar results to dual-energy X-ray absorptiometry (DXA) for fat-free mass in the normal weight and obese groups in different cohort. However, fat percentage (and therefore presumably fat mass) was underestimated in the normal weight, overweight and obese groups by ~4%[565]. This effect was also seen in a previous study[566] , and should be considered when interpreting associations with outcome measures such as cancer risk.

Hydration status might also affect the bioimpedance measures. However the Tanita BC 418 MA instrument used in the UK Biobank appears not to be greatly affected by hydration or dehydration[567,568]. Workers also showed that exercise, bladder voiding and food intake do not significantly impact on estimates of fat percentage[568]. Unfortunately, these workers did not extend their study to the investigation of fat-free mass.

WBFM is a simple parameter as it measures the amount of adipose tissue. Whereas, WBFFM includes cell mass, total body water, and bone mineral mass. One study using

ultrasonic sonography concluded that ~35-38% of fat-free mass corresponds to skeletal muscle mass across three different population groups with a mean age of under 50 years which is lower than the UK Biobank cohort (56.5 years)[569]. Like weight, the composition of fat-free mass is variable between individuals and over time with some elements (e.g. cell mass and bone mineral mass) being intercorrelated. These complexities pose some challenges in directly interpreting WBFFM, and suggest an additional requirement for genetic or laboratory work to dissect any associations.

2.4.3 Other interesting findings: pharmacological interventions

In terms of covariates, there were some notable findings which support or conflict with other cohorts. Perhaps the most notable are those which are modifiable by pharmacological intervention. For instance, metformin was associated with an increased risk of colorectal cancer in men (e.g. BMI model, beta = 0.252 [SE = 0.147], p =0.0875) in this cohort. This was surprising as metformin has been shown to be associated with a reduced incidence of "any cancer"[570] and in vitro work has suggested it has anti-tumour effects[571–573]. However, looking specifically at colorectal cancer, the effect of metformin does not appear to be clear cut. It has been associated with prevention[574,575], no difference[576,577] and even associated with increased colorectal cancer risk[578], as here. The reasons for these different results have been proposed to be related to time-related biases, including immortal time bias[579] (when an endpoint happens before the treatment commences as discussed and addressed in[580]), time-window bias[581] (when the time window for capturing exposure differs between cases and controls in case/control studies). The true effect is unknown and perhaps a one-size fits all approach is not appropriate and more personalisation is needed. For instance the genomics of complex 1 of the respiratory chain may play a role on an individual level, as has been seen in vitro with metformin's close relative, phenformin[582]. Interestingly, one of the commonest side effects of metformin is diarrhoea and presumably some local

inflammation. So perhaps those patients who suffer with this side effect and persevere might be those who go on to develop colorectal cancer. Further this side-effect severely affects compliance,adding another complication in interpreting these studies.

Other supplements significantly affecting the models include saw palmetto, which, in this study, appears to be associated with an increased risk factor for prostate cancer (e.g.WBFM, beta = 0.397 [SE =0.192], p= 0.0389). Cellular studies have shown that saw palmetto inhibits growth of prostatic cancer cells and may induce apoptosis[583], while human trials have seen no effect on prostate cancer risk[584,585].

In this analysis supplementation with vitamin E, already controversial, appears to be associated with a similar increased risk for prostate cancer across the BMI/WBFM and WBFFM models (e.g. WBFFM model, beta = 0.2107 [SE = 0.089], p= 0.017), in line with previous work[586]. A mechanism for the increased risk for prostate cancer with vitamin E supplementation has been proposed from *in vitro* and rat *in vivo* studies. Vitamin E appears to generate supraphysiological levels of reactive oxygen species (ROS) and causes increased DNA damage. Further vitamin E also appears increase expression of cyclooxygenase-2 (COX-2)[587], resulting in overproduction of prostaglandin E2 (PGE2), which has been shown to increase prostate cancer cell growth in model systems and EMT[588,589].

2.4.4 Conclusion

This study is the first to use principal component analysis in trying to untangle the relationship between obesity and cancer risk. Although recently becoming increasingly popular in elements of epidemiology[590], principal component analysis has not yet been applied in this context. Further, this work opens up some interesting questions surrounding cause and effect of body composition and cancer, particularly in relation to WBFFM. In the following chapters we investigate these questions with approaches to mine the genomic landscape, to see if single nucleotide polymorphisms may be simultaneously related to anthropometric traits, including WBFFM, and cancer risk, hopefully giving rise to new hypotheses as to how these two concepts might be linked in the evolving field of anthropo-oncogenomics.

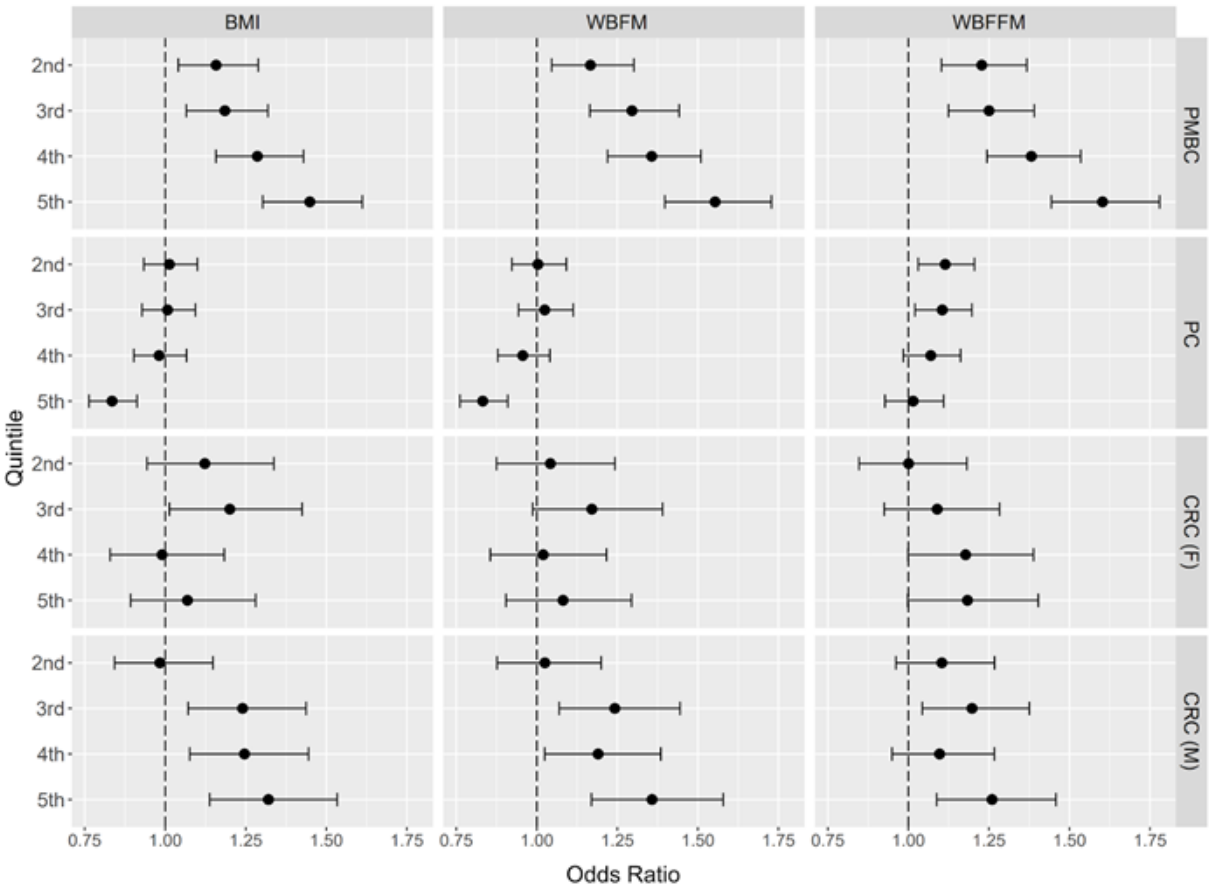


Figure 2.9: *Relation of anthropometric traits to cancer risk across four cancer types.*
Summary of the odds ratios by multivariable logistic regression for the different quintiles of BMI, WBFM and WBFFM in the post-menopausal breast cancer , prostate cancer and colorectal cancer analyses. Data taken from Figs 2.1, 2.3, 2.5 and 2.7.Covariates for each model are detailed in the methods section. Quintiles (2nd, 3rd, 4th and 5th) were compared to the lowest (1st quintile) to give an odds ratios for cancer risk and the 5th quintile denotes the individuals with the highest measurements in the UK Biobank cohort for the anthropometric measure in question (e.g. those with the highest BMI). BMI = body mass index, WBFM = whole body fat mass, WBFFM = whole body fat-free mass, PMBC = post-menopausal breast cancer, PC = prostate cancer, CRC (F) = colorectal cancer in women, CRC (M) = colorectal cancer in men.

Chapter 3: Cancer risk and anthropometric traits - heritable genetic variants in tumour suppressor genes and proto-oncogenes

Chapter Summary

Background: Height and other anthropometric measures are consistently found to associate with differential cancer risk. However, both genetic and mechanistic insights into these epidemiological associations are notably lacking. Conversely inherited genetic variants in tumour suppressors and oncogenes increase cancer risk, but little is known about their influence on anthropometric traits.

Methods: By integrating inherited and somatic cancer genetic data from the Genome-Wide Association Studies catalogue (GWAS Catalog), expression quantitative trait loci (eQTL) databases and the Cancer Gene Census, single nucleotide polymorphisms (SNPs) can be identified which associate with different cancer types and differential gene expression in at least one tissue type. We explore the potential pleiotropic associations of these SNPs with anthropometric traits through SNP-wise association in a cohort of 500,000 individuals.

Results: Three regulatory SNPs for three important cancer genes, FANCA, MAP3K1 and TP53, have been identified which associate with both anthropometric traits and cancer risk. Of particular interest, a previously unrecognised strong association has been identified between the rs78378222[C] SNP in the 3'-UTR of TP53 and increased risk for developing both non-melanomatous skin cancer (OR=1.36 [1.31-1.41], adjusted-p=$7.62E^{-63}$), and brain malignancy (OR=3.12 [2.22-4.37], adjusted p=$1.43E^{-12}$) together with increased standing height (adjusted-p=$2.18E^{-24}$, beta=0.073±0.007), lean body mass (adjusted-p=$8.34E^{-37}$, beta=0.073±0.005) and basal metabolic rate (adjusted-p=$1.13E^{-31}$, beta=0.076±0.006), thus offering a novel genetic link between these anthropometric traits and cancer risk.

Conclusion: These results clearly demonstrate that heritable variants in key cancer genes can associate with both differential cancer risk and anthropometric traits in the general population, thereby lending support for a genetic basis for linking these human phenotypes.

3.1 Introduction

Height and other anthropometric measures are consistently found to associate with differential cancer risk[591–593]. However, neither genetic and mechanistic insights into these epidemiological associations are available[592,594]. Increased height has been shown to associate with a number of malignancies including skin, breast, colon, rectum, endometrium, ovary, and kidney cancers, as well as Hodgkin's lymphoma and leukaemia[592,595]. Current theories suggest that links between increased height and increased cancer risk relate to the role of growth hormone, insulin-like growth factor and/or insulin as pro-tumorigenic (by up-regulating the Ras-MAPK and PI3K pathways), and hence leading to increased cellular proliferation[596,597]. It has also been proposed that increased height might correspond to increased cell number and therefore an increased probability of malignancy, simply by chance[598]; however, such allometric scaling of body mass explains only part of the observed effect and is not observed among mammalian species of different sizes (Peto's paradox)[599]. Like height, increased body mass index (BMI), a surrogate measure of obesity based on height and weight, has been associated with an increased risk of various cancers, including post-menopausal breast cancer, colorectal and renal cancer[600]. Increased BMI within individuals is normally attributed to increasing accumulation of fat mass. This adiposity is hypothesised to cause a state of systemic inflammation, a shift of metabolite and adipokine release, and an increase of circulating insulin, secondary to insulin resistance[25,601], a combination that may increase cancer risk. However, the exact mechanism(s) linking high BMI with increased cancer risk are not understood. Thus, links between height, anthropometric measures and cancer remain an area of intense research.

Anthropometric traits are largely determined by genes that control cellular proliferation, metabolism and apoptosis: these processes are also required for immortalisation of cancer cells and development of tumours[34]. It is known that germline mutations within tumour

suppressor genes and oncogenes affect cancer risk[602]; however, their influence on anthropometric traits is not well-known. Studies on mouse models and rare diseases indicate that mutations in such genes may influence body mass composition[98,409,603]. For example, the tumour suppressor gene ARH1 inhibits cell growth. A deletion in ARH1 is commonly associated with breast and ovarian carcinoma, while mice that are engineered to overexpress ARH1 are significantly smaller than wild type counterparts[604]. Mice possessing knockouts of genes within the tumorigenic hypoxia-inducible pathway vary in size and respond to high-fat diets differently from those without the mutations[605,606].TP53 is the most frequently mutated gene in human cancers and is a key regulator of a number of cellular activities that prevent tumorigenesis, including maintaining genomic stability, controlling cell growth and metabolism[409,410]. Mouse models of TP53 mutations have demonstrated that reduction of p53 activity can increase cancer risk, and also alter metabolism/influence obesity in a complex and signal dependent manner[98,603]. Further evidence for a relationship between body size, genetics and cancer comes from human genetic diseases, such as Turner syndrome. Turner Syndrome is defined by complete or partial chromosome X monosomy and is associated with a distinct clinical phenotype, including gonadal dysgenesis, cubitus valgus and short stature[607]. Short stature in this condition has been attributed to haplo-insufficiency of the SHOX tumour suppressor gene[608]. Deletions and mutations in SHOX have also been proposed to explain short stature within the general population[609]. Another example comes from patients with Simson–Golabi–Behmel syndrome (SGB), caused by mutations in the GPC3 tumour suppressor gene. Patients with SGB tend to have pre- and post-natal overgrowth and thus ultimately taller stature together with an increased risk of embryonic tumours. GPC3-null mice also display overgrowth[610]. These effects, in both humans and mice, are thought to be due to unshackling of the hedgehog pathways, and increasing cellular proliferation[611,612].

These rare diseases provide evidence that major differences in the expression of cancer genes can influence both anthropometric traits and cancer risk in specific cohorts of

patients. However, it is less clear that such associations occur in the broader population. The availability of large phenotypic and genetically linked datasets has allowed us to deploy unsupervised approaches, such as GWAS studies of common SNPs, further to understand the link between genetics and anthropometric traits. Pleiotropy is the phenomenon whereby a single SNP or genetic mutation can influence multiple traits. Until now, it has been difficult to look for pleiotropy between functional SNPs in genes associated with cancer risk and anthropometric traits, as very few cohorts possess both comprehensive genetic data and detailed anthropometric data in the same population. The UK Biobank provides a unique opportunity to investigate such pleiotropic associations in a large prospective cohort of over 500,000 participants.

3.2 Methods

3.2.1 The UK Biobank

The UK Biobank is a cohort of ~500,000 UK is further detailed in *Chapter 2*. This study was conducted under the UK Biobank approved application (#24456, PI Gareth Bond).

3.2.2 Anthropometric data

During the baseline assessment, participants had various anthropometric traits measured directly or by bioelectrical impedance. Bioimpedance data was measured using a Tanita BC418MA body composition analyser. Participants stood barefoot on the analyser and held the metal handles. This device produced measurements of fat mass, fat-free mass and basal metabolic rate. Further information on anthropometric data collected is found here[532].

3.2.3 Cancer data

Cancer occurrences were defined by presence of a cancer international classification of diseases (ICD) code in the UK Cancer register or the UK Death register. To maximise the number of individual cancers cases, we combined ICD9 and ICD10 codes of identical cancers. This was done by a clinician to ensure the matching was accurate, for instance malignant neoplasm of brain was defined as C71 (ICD10) and 191 (ICD9). With the goal of increasing power, when appropriate we merged ICD codes into clinically relevant groups. We ran individual cancers (combined ICD9 and ICD10 codes) and clinically relevant groups in our analysis.

3.2.4 Genetic data

Blood samples were collected when participants were recruited, and DNA extracted[613]. DNA was then genotyped on either the Affymetrix UK BiLEVE Axiom array or the Affymetrix UK Biobank Axiom array (Santa Clara, CA, USA). Imputation was based upon a merged reference panel of ~90 million biallelic variants, from the 1000 Genomes Phase 3[614] and the UK10K[615] haplotype panels. Imputation was performed using IMPUTE2 as described[616,617], producing 488,295 genotyped participants.

3.2.5 Sample quality control

In addition to the standard quality control, we carried out further quality control steps to ensure robustness of our analyses. We excluded individuals based on: a mismatched value between self-reported and genetic sex (data-field: 22001 and 31), level of genotype missingness of >0.05 (data-field: 22005), genetic relatedness factor with kinship coefficient of >0.0442, sex chromosome aneuploidy (data-field 22019), outliers for heterozygosity or missing rate (data-field: 22027). We selected the European population based on self-reported ethnicity (data-field: 21000) by excluding non-white ethnic background. This left a study population of 379,358 suitable genotyped individuals.

3.2.6 Identification of functional cancer gene SNPs

We identified genes confirmed to be involved in carcinogenesis in the Catalogue of Somatic Mutations In Cancer (COSMIC) Cancer Gene Census (release v88, 19th March 2019): the current reference record of genes containing cancer driver mutations. We then selected all SNPs annotated to these genes that have been significantly associated with differential cancer risk in genome-wide association studies in the GWAS catalog[618]. Significant associations with cancer risk were defined with a p-values of $<5E^{-08}$ cut-off. Finally, it was required that these

SNPs were also associated with differential expression of the cancer gene in at least one expression quantitative trait loci (eQTL) database. cis-eQTL databases utilised were GTEX, NESDA/NTR and PancanQTL[619–621]. Further details of these databases are found below:

- GTEX - Gene expression, histological and clinical data for 449 human donors across 44 (42 distinct) tissues, including: Adipose - Subcutaneous, Adipose - Visceral (Omentum), Adrenal Gland, Artery - Aorta, Artery - Coronary, Artery - Tibial, Brain - Anterior cingulate cortex (BA24), Brain - Caudate (basal ganglia), Brain - Cerebellar Hemisphere [Frozen], Brain - Cerebellum [PAXgene], Brain - Cortex [PAXgene], Brain - Frontal Cortex (BA9) [Frozen], Brain - Hippocampus, Brain - Hypothalamus, Brain - Nucleus accumbens (basal ganglia), Brain - Putamen (basal ganglia), Breast - Mammary Tissue, Cells - EBV-transformed lymphocytes, Cells - Transformed fibroblasts, Colon - Sigmoid, Colon - Transverse, Esophagus - Gastroesophageal Junction, Esophagus - Mucosa, Esophagus - Muscularis, Heart - Atrial Appendage, Heart - Left Ventricle, Liver, Lung, Muscle - Skeletal, Nerve - Tibial, Ovary, Pancreas, Pituitary, Prostate, Skin - Not Sun Exposed (Suprapubic), Skin - Sun Exposed (Lower leg), Small Intestine - Terminal Ileum, Spleen, Stomach, Testis, Thyroid, Uterus, Vagina, Whole Blood[622].

- NESDA/NTR - Peripheral blood gene expression was measured on 4,896 subjects with European ancestry (1,880 unrelated subjects from NESDA, 559 MZ twin pairs, 594 DZ twin pairs, 51 parent-sibling trios and 557 unrelated subjects from NTR)[621].

- PancanQTL - provides cis-eQTLs and trans-eQTLs in 33 cancer types and 9,196 tumor samples in The Cancer Genome Atlas (TCGA), which are: Adrenocortical carcinoma, Bladder Urothelial Carcinoma, Breast invasive carcinoma, Cervical squamous cell carcinoma and endocervical adenocarcinoma, Cholangiocarcinoma, Colon adenocarcinoma, Lymphoid Neoplasm Diffuse Large

B-cell Lymphoma, Esophageal carcinoma, Glioblastoma Multiforme, Head and Neck squamous cell carcinoma, Kidney Chromophobe, Kidney renal clear cell carcinoma, Kidney renal papillary cell carcinoma, Acute Myeloid Leukemia, Lower Grade Glioma, Liver hepatocellular carcinoma, Lung adenocarcinoma, Lung squamous cell carcinoma, Mesothelioma, Ovarian serous cystadenocarcinoma, Pancreatic adenocarcinoma, Pheochromocytoma and Paraganglioma, Prostate adenocarcinoma, Rectum adenocarcinoma, Sarcoma, Skin Cutaneous Melanoma, Stomach adenocarcinoma, Testicular Germ Cell Tumors, Thyroid carcinoma, Thymoma, Uterine Corpus Endometrial Carcinoma, Uterine Carcinosarcoma, Uveal Melanoma[620].

N.B. It is worth noting eQTL identification involves a direct association test between SNPs and gene expression levels typically measured in *tens or hundreds* of individuals. This approach allows identification of functional loci without reliance on existing knowledge. Identification of eQTLs has been successful when identifying regulatory regions close (within 1Mb etiher side) to the genes themselves, *cis*-eQTLs. However eliciting trans-eQTLs, where SNPs were over 5 Mb downstream or upstream of the gene has proved challenging when dealing with the whole-genome of expression data, relatively low numbers of individuals tested and stringent significance correction issues. At present it is unknown whether the preponderance to *cis* versus *trans* eQTLs reflects biological reality or is artifactual[623–625], however here the focus is on cis-eQTLs.

3.2.7 SNP quality control

SNP exclusions were Hardy-Weinberg equilibrium with p-value less than $1E^{-10}$, a minor allele frequency less than 0.0001, level of missingness more than 0.05 or an imputation score less than 0.8 (as per http://www.nealelab.is/uk-biobank/).

3.2.8 SNP-wise analysis

To carry out the SNP-wise analysis on all functional-CG SNPs we used SNPTEST (version 2.5.4)[626], and employed the frequentist approach under dominant, additive and recessive inheritance models, using sex, age and genetic principal components (1-20 PCs) as covariates. Genetic PCs were included as covariates to control for hidden population structure. We controlled the genotype uncertainty by implementing the missing data likelihood score test. Within SNPTEST, we used the frequentist approach under an additive inheritance model. P-values of SNP-wise association were adjusted by the stringent Bonferroni correction based on the number of tested traits (28 anthropometric traits, 50 cancer types) multiplied by the number of eSNPs (100). Significant associations were defined as Bonferroni correction p values below the threshold of $1E^{-5}$.

3.2.9 Linkage and lead SNPs

R^2 and D' coefficients were calculated to evaluate the degree of Linkage Disequilibrium (LD) between different loci using LDlink v3.7 (D'> 0.8, R^2>0.4 https://ldlink.nci.nih.gov/)[627]. Lead SNPs were defined as the SNPs that were most strongly associated with the traits in question for each locus[628]. Due to the low value of r2 in some of the potential LD eSNPs we carried out a leave-one-out analysis:

- For each lead SNP:
    - All participants from the cohort carrying the lead SNP were removed.
    - The association for each of the potential LD eSNPs was performed (e.g. lead SNP rs78378222 carriers were removed whilst rs35850753 and rs8753 were tested for their associations with cancer and anthropometric traits).

- Adjusted p-values from the whole cohort association analysis with those obtained in this analysis were compared.
- If the association of the potential LD eSNPs was not significant after the removal of lead SNP carriers (p-values>0.05), we considered the eSNP in Linkage Disequilibrium.

3.3 Results

Through integration and curation of the GWAS catalog, eQTL databases and the Cancer Gene Census we identified 100 SNPs, which have been shown to associate with a differential risk of developing a total of 21 different cancer types and differential gene expression in at least one tissue type of *(i)* proto-oncogenes (8 genes) *(ii)* oncogenic fusion proteins (8), *(iii)* tumour suppressors (16) and *(iv)* 15 genes that span two or more of these groups (supplementary tables S3.1 and S3.2). We define these SNPs as *cancer eSNPs*. In our analyses of the UK Biobank cohort, we found 13 cancer eSNPs associated with differences in anthropometric traits and 31 with differential cancer risk. Interestingly, 7 of these SNPs overlapped and thus displayed some level of pleiotropy between cancer risk and anthropometric traits (Fig. 3.1, supplementary tables S3.3 and S3.4). These 7 cancer eSNPs are found on three different chromosomes and in linkage disequilibrium (see Methods) with at least one other identified cancer eSNP (Fig. 3.2). They have been found to be associated with differential expression levels of *(i)* Fanconi Anemia, complementation group A, FANCA [rs1805007, rs258322], *(ii)* tumor suppressor p53, TP53 [rs78378222, rs35850753, rs8753], and Mitogen-Activated Protein Kinase 3, Kinase 1 [MAP3K1: rs889312, rs1862626]). Both cancer eSNPs associated with FANCA expression do not reside in the FANCA gene but in neighbouring genes; rs258322 is a non-coding variant within CDK10 and rs1805007 is a missense variant within Melanocortin 1 Receptor (MC1R). eSNPs associated with MAP3K1 expression are found in an intergenic region (5q11.2) close to Zinc Finger Protein 296 (ZNF296) and MAP3K1. In contrast, the two cancer eSNPs associated with differential TP53 expression are found in untranslated regions (UTR) of the TP53 gene itself. rs78378222 is found in the 3'-UTR of TP53 and rs35850753 is found in the 5'-UTR of the d133 isoform of TP53. The third TP53 eSNP is in the 3'-UTR of the neighbouring RNA Polymerase II Subunit A gene (POLR2A). The cancer eSNPs that showed the strongest association with anthropometric measures and cancer risk within their loci (lead cancer eSNP) were rs1805007

(FANCA, C>T, minor allele frequency (MAF) in UK Biobank = 0.102), rs78378222 (TP53, A>C, MAF= 0.012) and rs889312 (MAP3K1, C>A, MAF= 0.284).

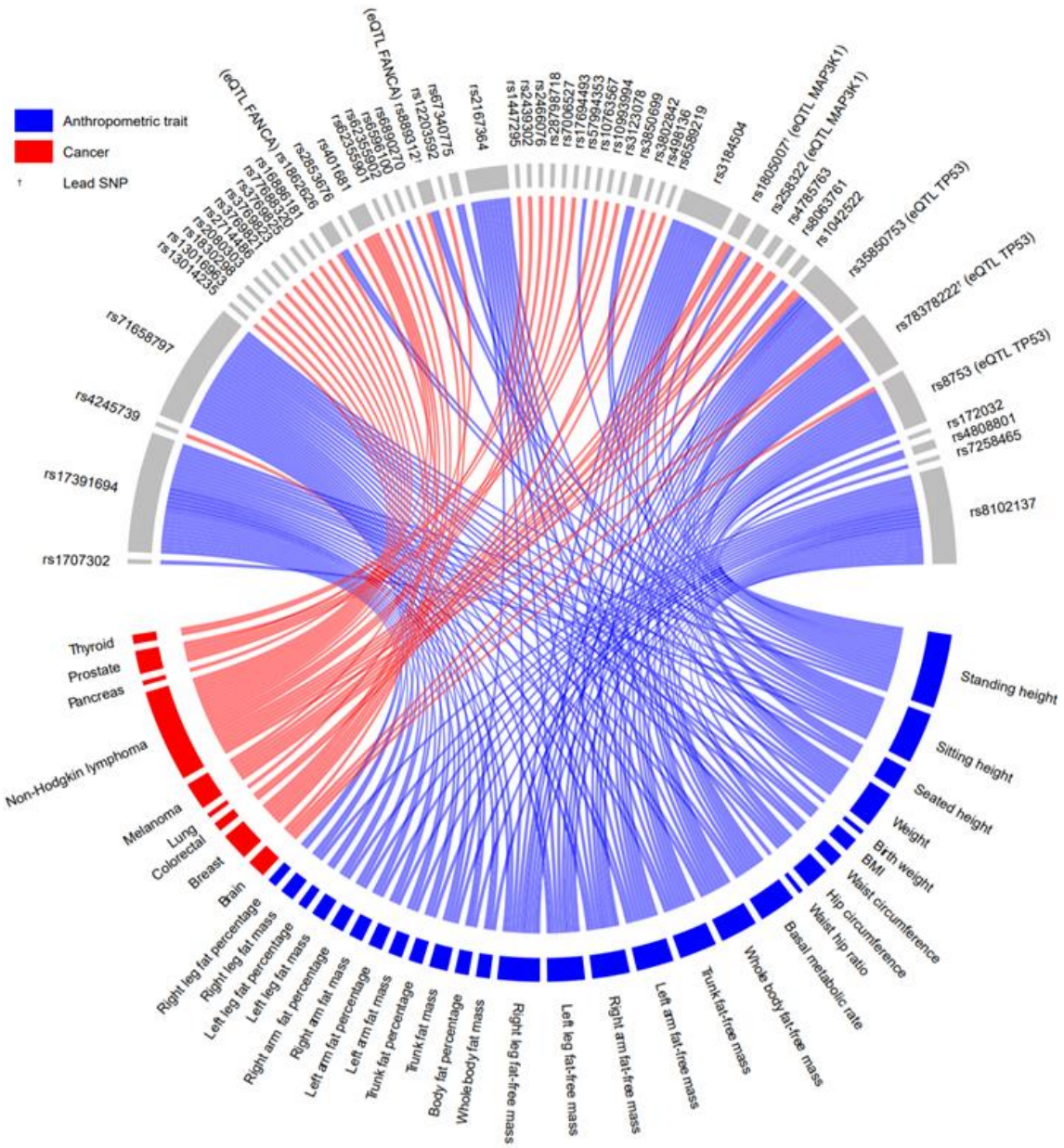
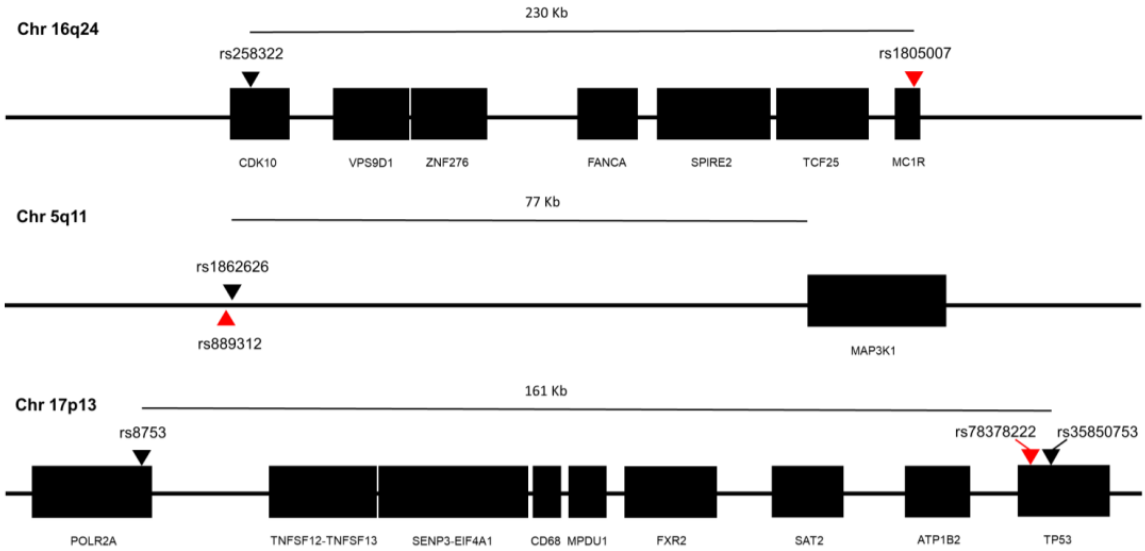


Figure 3.1: *Association between functional SNPs, anthropometric traits and cancer risk (circos plot)*. SNPs are located on the top half of the plot whilst cancer types and anthropometric traits are on the bottom half. Significant associations (Bonferroni corrected p-values <1E^{-5}) are shown as solid lines. Blue lines indicate significant associations with anthropometric traits. Red lines depict significant associations with differential cancer risk. A small subset (see text) are linked by both red and blue lines, and hence represent pleiotropic SNPs. The data sources are as given in Methods. n = 379,358.



rs1805007†/ rs258322	
GWAS catalog traits	Skin (Melanoma, Non-melanomatous skin cancer)
eQTL database	PancanQTL
eQTL gene	FANCA
Gene type	Tumour suppressor
Tissue type	Epithelial
Germline	AML, leukaemia
Somatic	-

rs889312†/ rs1862626	
GWAS catalog traits	Breast cancer
eQTL database	NESDA.NTR
eQTL gene	MAP3K1
Gene type	Proto-oncogene; Tumour suppressor
Tissue type	Leukaemia
Germline	-
Somatic	Luminal A breast

rs78378222†/ rs35850753 / rs8753	
GWAS catalog traits	Brain (Glioma), Skin (Non-melanomatous skin cancer)
eQTL database	NESDA.NTR
eQTL gene	TP53
Gene type	Tumour suppressor; Proto-oncogene; Oncogenic fusion gene
Tissue types	Leukaemia, Epithelial, Mesenchymal, Other
Germline	Breast, sarcoma, adrenocortical carcinoma, glioma, multiple other tumour types
Somatic	Breast, colorectal, lung, sarcoma, adrenocortical, glioma, Spitzoid tumour, multiple other tumour types

Figure 3.2: *Genetic map of the chromosomal regions containing the pleiotropic functional SNPs.* Lead SNPs are highlighted in red. Tables summarise salient GWAS catalog, eQTL, and Cancer Gene Census information for each variant.

In the UK Biobank cohort, the minor allele of rs1805007 (FANCA) was associated with an increased risk of melanoma (OR=1.63 [1.52-1.75], adjusted-p=$2.57E^{-41}$) and non-melanomatous skin cancer (OR=1.36 [1.31-1.41], adjusted-p=$7.62E^{-63}$). The minor allele of rs78378222 (TP53) was associated with an increased risk of brain malignancy (OR= 3.12 [2.22-4.37], adjusted p= $1.43E^{-12}$) and non-melanomatous skin cancer (OR= 1.46 [1.34-1.60], adjusted p= $5.20E^{-18}$). The minor-allele of rs889312 (MAP3K1) was associated with an increased risk of breast cancer (OR= 1.1 [1.07-1.13], adjusted p= $2.82E^{-11}$). Of note, as SNPs were selected due to their noted association with differential cancer susceptibility in GWAS studies, this study provides independent validation of these associations (Fig. 3.3a, supplementary table S3.5).

As regards the anthropometric measures, we identified an unexpectedly high level of pleiotropy and very strong association between the TP53 cancer eSNPs and three measures of height and six of lean body mass, as well as basal metabolic rate, with p-values as low as $7.54E^{-36}$ (Fig. 3.3b). After quality control (QC) procedures, we identified 59 participants homozygous for the minor allele (increased cancer risk) of the lead TP53 eSNP (rs78378222), 9,253 heterozygous for the minor allele and 370,046 homozygous for the major allele. The minor allele carriers were on average taller, had an increased fat-free mass and a higher basal metabolic rate. The associations of these anthropometric traits with the minor allele of the rs78378222 SNP were markedly more significant (up to almost 3 fold) and with greater effect sizes (up to almost 5 fold) than for the other pleiotropic eSNPs (Fig 3.3c, supplementary table 3.3).

The cancer eSNPs for TP53 demonstrated strikingly strong associations with 10 different anthropometric traits in the UK Biobank cohort. In contrast, the cancer eSNPs related to FANCA [rs1805007] and MAP3K1 [rs889312] were associated with standing

height but not lean body mass measures. After QC, we identified 4271 participants homozygous for the minor allele of the lead FANCA eSNP, 68,956 heterozygous for the minor allele and 306,131 homozygous for the major allele. The minor allele for the lead MAP3K1 eSNP was found in 185,184 participants (30,394 homozygous, 154,790 heterozygous). The FANCA eSNP [rs1805007] only associated with standing height (beta= -0.02 ± 0.002, adjusted-p=9.20E^{-15}), while the MAP3K1 eSNP [rs889312] associated with broader range of height measures (standing height beta= -0.02 ± 0.004, adjusted p= 6.11E^{-14} and sitting height beta= -0.01 ± 0.001, adjusted p= 3.34E^{-9}). In order to further validate the significance and magnitude of these associations to the lead cancer eSNPs, a different algorithm was used, PLINK. As in the initial analysis, the covariates sex, age and genetic principal components were included. Importantly, the PLINK analysis confirmed the observed associations and yielded similar adjusted p and beta values for the lead cancer eSNPs (supplementary tables S3.6-8). In addition, validation was sought for these associations with standing height in the independent GIANT consortium dataset. Here, all lead and linked SNPs for each pleiotropic eSNP were examined in the results of the meta-analysis conducted by Wood and colleagues (2014)[629]. eSNPs for MAP3K1 (rs889312, rs1862626) and FANCA (rs1805007, rs258322) were identified, but these workers did not identify correlations for TP53, likely due to the low minor allele frequency. Three of these eSNPs significantly associated with height in the Wood and colleagues meta-analysis: (i) rs258322 FANCA eQTL (p=1.5E-09, Beta=-0.029), (ii) rs889312 MAP3K1 eQTL (p=2.3E-08, Beta=-0.018) and (iii) rs1862626 MAP3K1 eQTL (p=3.4E-08, Beta=-0.018). Interestingly, both linked MAP3K1 eSNPs and a FANCA eSNP passed the significance threshold and, reassuringly, the directions of the allelic associations are consistent with our findings. The other FANCA eSNP only just fell short of the GWAS significance threshold (rs1805007, 4.3E-04, Beta = -0.024). Together, these data clearly link eSNPs for MAP3K1 and FANCA with height.

Several cancer eSNPs from the GWAS Catalog were found to associate with anthropometric traits but their association with cancer could not be replicated in the UK Biobank cohort (at the time of analysis, supplementary table S3.4). Of particular interest are rs71658797[A], and its linked eSNP rs17391694[T] which are associated with an increased risk of lung cancer in the GWAS Catalog (ORs:1.11[1.07-1.15] - 1.18[1.12-1.25])[630]. Both eSNPs affect Far Upstream Element Binding Protein 1(FUBP1), which has been shown to influence c-myc expression[631]. These FUBP1 eSNPs associate with multiple measures including all three measures of height (standing height:adjusted-p = 3.90594E-27, beta = 0.026021 ± 0.002, seated height: adjusted-p = 4.78563E-16, beta = 0.0229991 ± 0.003, sitting height: adjusted-p = 1.28399E-17, beta = 0.0228278 ± 0.003), all measures of fat-mass (whole body fat mass: adjusted-p = 1.17086E-32, beta = 0.0411098 ± 0.004), all measures of fat-free mass (whole body fat-free mass: adjusted-p = 4.20359E-48, beta = 0.0280651 ± 0.002), basal metabolic rate (adjusted-p = 2.57824E-49, beta = 0.0320974 ±0.002), waist circumference (adjusted-p = 1.78737E-22, beta = 0.0303305 ± 0.003), hip circumference (adjusted-p = 1.16764E-33, beta = 0.0426943 ± 0.004) and even the composite measure body mass index (adjusted-p = 1.74418E-26, beta = 0.037434 ± 0.004). Another eSNP, rs8102137[C], which is associated with an increased risk of bladder cancer in the GWAS Catalog (OR 1.13 [1.09-1.17])[632,633] also associates with multiple anthropometric traits. This eSNP affects Cyclin E1 (CCNE1) which plays a vital role in the cell cycle, controlling the G1/S transition[634]. The CCNE1 eSNP associates with all measures of fat-mass (whole body fat mass: adjusted-p = 4.77057E-17, beta = 0.020069 ± 0.002), all measures of fat-free mass (whole body fat-free mass: adjusted-p = 3.40997E-16, beta = 0.010877 ± 0.001), basal metabolic rate (adjusted-p = 1.46999E-17, beta = 0.0128388 ± 0.002), waist circumference (adjusted-p = 5.07311E-14, beta = 0.0162168 ± 0.002), hip circumference (adjusted-p = 5.5749E-12, beta = 0.0168476 ± 0.002). It is notable these beta values for the anthropometric traits with these eSNPs are markedly lower than the minor allele of rs78378222.

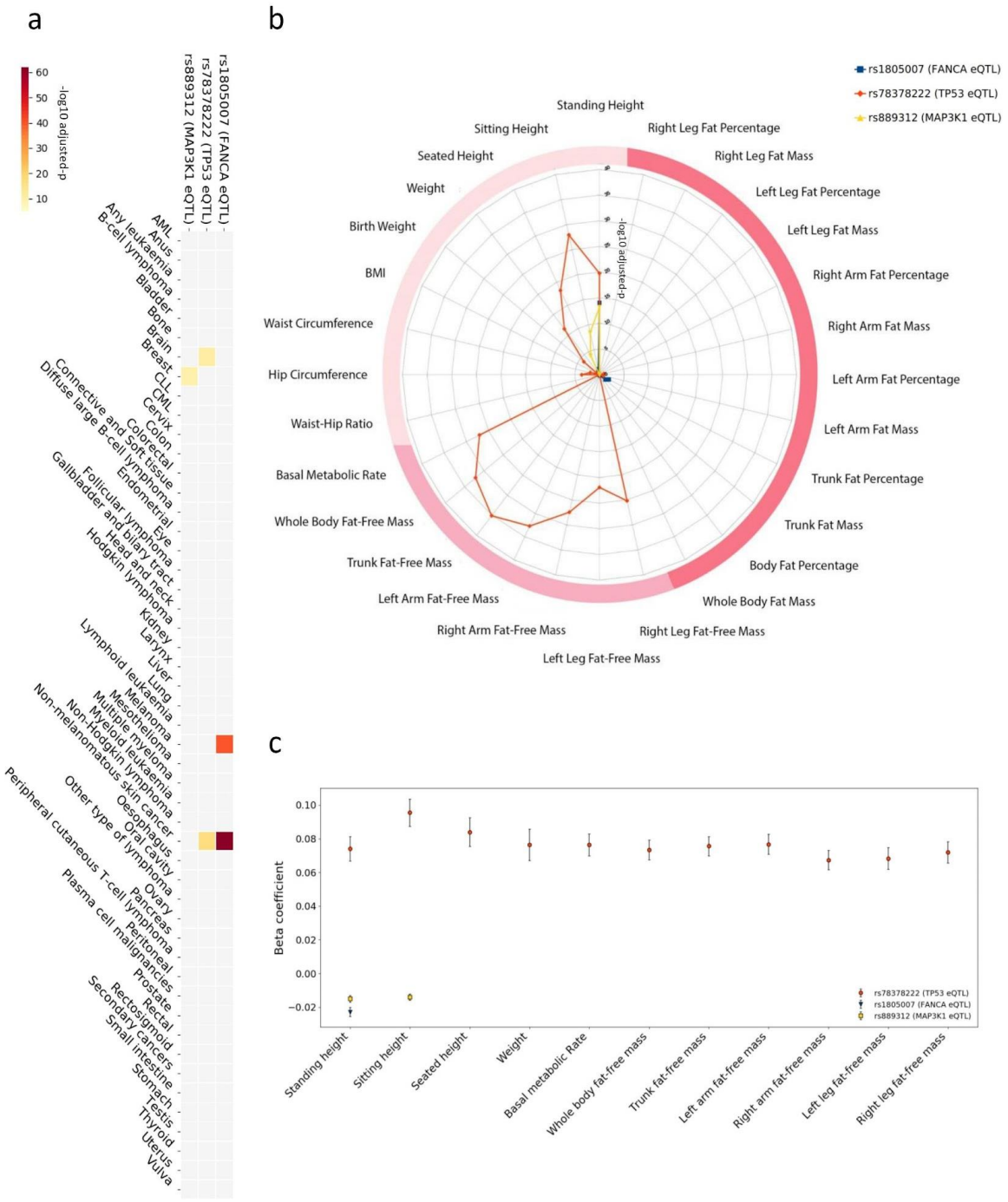


Figure 3.3: *Association of cancer eSNPs with anthropometric traits and cancer risk in the UK Biobank cohort*.

(a) Heatmap illustrating the significance of the associations of the three lead pleiotropic SNPs with differential cancer risk. The colour scale represents the $\log_{10}$ adjusted-p: with the darkest red on the scale being $7.62E^{-63}$ and the lightest yellow being $1E^{-5}$. (b) Radar plot illustrating the level of significance of the associations of the three lead pleiotropic SNPs with anthropometric traits (radial

axis: $-\log_{10}$ adjusted p). The darkest two pink categories are bio-electrical impedance measures (darkest being measures of fat). Traditional bedside anthropometric measures are in blush. Lead cancer eSNP for FANCA (rs1805007) and MAP3K1 (rs889312) associate with standing height. The lead functional SNP related to TP53, rs78378222, associates with multiple measures of height and lean body mass, as well as basal metabolic rate. (c) Plot showing errors on the beta coefficients (y-axis) of the significant associations with anthropometric traits (x-axis). Error bars denote the 95% confidence interval on the beta value.

3.4 Discussion

This is the first comprehensive study providing evidence that functional common genetic variants in oncogenes and tumour suppressor genes may be associated with both anthropometric traits and cancer risk in the general European population and in the same cohort. SNPs exhibiting these pleiotropic associations in our study are found in three different loci: (i) MAP3K1 (two SNPs in linkage disequilibrium), (ii) FANCA (two SNPs in linkage disequilibrium) and (iii) TP53 (three SNPs in linkage disequilibrium). Observations gained from human studies and mouse models designed to alter signalling pathways involving these three genes (MAP3K1, FANCA and TP53) suggest that such associations with anthropometric traits in humans are feasible.

3.4.1 MAP3K1

Mitogen-Activated Protein Kinase Kinase Kinase 1 (MAPKKK1/MAP3K1/MEKK/MEKK1/ SRXY6/Raf, chromosomal location: 5q11.2) is a serine/threonine kinase which is part of the MAPK cascade[635]. Mitogen-Activated Protein Kinase Kinase Kinases (MAPKKKs) act upstream of Mitogen-Activated Protein Kinase Kinases (MEKs) which activate mitogen activated protein kinases (MAPK, ERK) signalling[636] (see Fig. 3.4).

MAP3K1 is unusual in that it can promote cell growth, via NF-κB and ERK activation (acting through primarily through JNK): however, it can also promote apoptosis via caspase activation and phosphorylated MEK4/7 which acts on p53 and BCL-2[637]. MAP3K1 has been shown to affect apoptosis in cancer and seems to act as a switch to integrate signals deciding on cell fate[638].

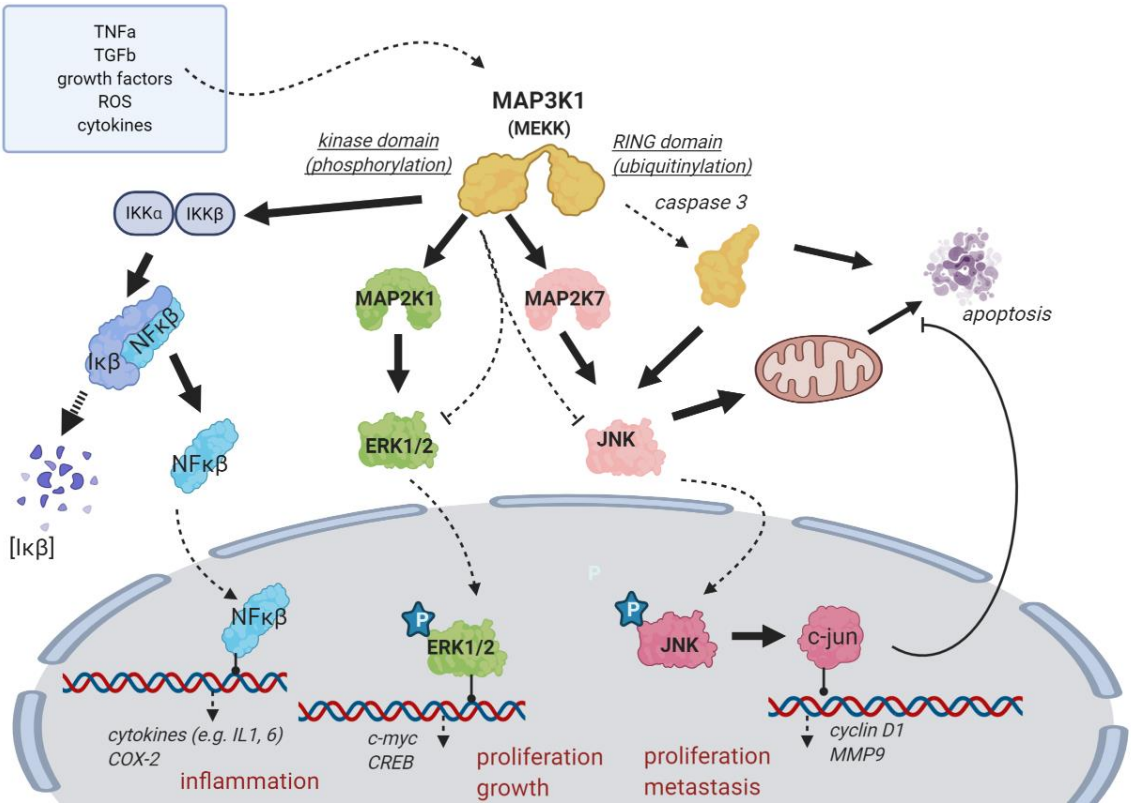


Figure 3.4: *Functions of MAP3K1.*

MAP3K1 (and its proteolysed kinase domain) promotes growth via MAP2K21/Erk kinases, proliferation via MAP2K7/JNK kinases, and inflammation via the NF-κB system. There is a brake on this system by direct inhibitory effects of MAP3K1 on Erk and JNK.

Within the COSMIC database, 1319 of the 51,444 unique samples showed mutations in MAP3K1 over multiple cancer types, including cancer of penis, breast and liver[639]. Mutations in breast cancer and cancer of the penis were most frequent (breast; 8.46%, n = 537/6350, penis 14.29%, n = 1/7). Further, SNPs related to the expression of MAP3K1 have been related to breast cancer risk in a number of different studies with median odds ratios ranging from 1.12 to 1.29 (supplementary table 3.5)[640–644]. As regards influencing anthropometric traits, knock-out studies and analyses of human data revealed that MAP3K1 genetic mutations are related to gonadal dysgenesis and that MAP3K1 integrates developmental signals vital for eyelid closure[645–647]. SNPs influencing the expression of MAP3K1, particularly rs889312, have been shown to be associated with standing height MAP3K1 eQTL (Beta=-0.018). In our study, we see associations with human height, both standing and sitting height (sitting height not present in the current GIANT dataset), as well as with breast cancer, as described in previous studies[640–644]. This study is the first to show both of these associations in the same cohort. The odds ratio for breast cancer risk and the beta value for standing height is similar to previous studies (breast cancer OR= 1.1 [1.07-1.13], standing height beta= -0.02).

3.4.2 FANCA

Fanconi Anemia complementation group A (FANCA) is one of the proteins within the Fanconi Anemia core complex. The Fanconi Anemia core complex plays a key role in interstrand crosslink repair, by monoubiquitination of FANCD2. Further, FANCA can bind in vitro with very high affinity to single-stranded RNA and DNA suggesting a role outside of repairing interstrand crosslinks, and it has been found to play a direct role in double strand break repair by catalysing single strand annealing and strand exchange at a level comparable to RAD52[648,649].

Within the COSMIC database, 1183 of the 52,106 unique samples showed mutations in FANCA over a variety of cancer types including cancer of stomach, endometrium and large intestine[639]. Mutations in endometrial cancer and skin cancer were most frequent (endometrial; 5.55%, n = 54/973, skin 6%, n = 106/1078). FANCA is the most-commonly mutated gene in Fanconi Anaemia (around 65% of cases, chromosomal location: 16q24.3). Mutations in the FANCA gene lead to a nonfunctional Fanconi Anaemia core complex. This disruption to the Fanconi Anaemia pathway leads to a build up of DNA damage, particularly interstrand cross links. Perseverance of these cross links affects DNA replication particularly in cells with rapid turnover. This can lead to stalled replication forks, cell death or genomic mutation leading to carcinogenesis. Patients with Fanconi Anaemia, an autosomal recessive condition, tend to have skeletal, renal and central nervous system issues, including aplastic anaemia (with increased susceptibility to acute myeloid leukaemia), absent radii, short stature, microcephaly, syndactyly congenital deafness and cryptorchidism.

SNPs related to the expression of FANCA have been associated with melanoma and non-melanomatous skin cancer in a number of different studies. The median odds ratios for melanoma was 1.5 and the median for non-melanomatous skin cancer ranged from 1.4 to 1.55 (supplementary table 3.5)[650–654]. As noted above, the human phenotype of Fanconi Anaemia

shows effects on body appearance, and targeted disruption of FANCA in mice leads to phenotypes mirroring the changes in humans, including growth retardation, microphthalmia and hypogonadism[655,656]. SNPs influencing the expression of FANCA have been shown to be associated with standing height FANCA eQTL (rs258322, Beta=-0.029). In the current study, associations are seen with standing height, as well as melanoma and non-melanomatous skin cancer, as described in previous studies[650–654]. For the first time in the same cohort, it is shown that eQTL SNPs for FANCA can affect both standing height and skin cancer risk.

The strongest associations observed with both anthropometric traits and cancer risk are loci related to TP53 (chromosomal location: 17p13.1). p53 is a key regulator of a number of cellular activities that prevent tumorigenesis including maintaining genomic stability, controlling cell growth and metabolism[410].

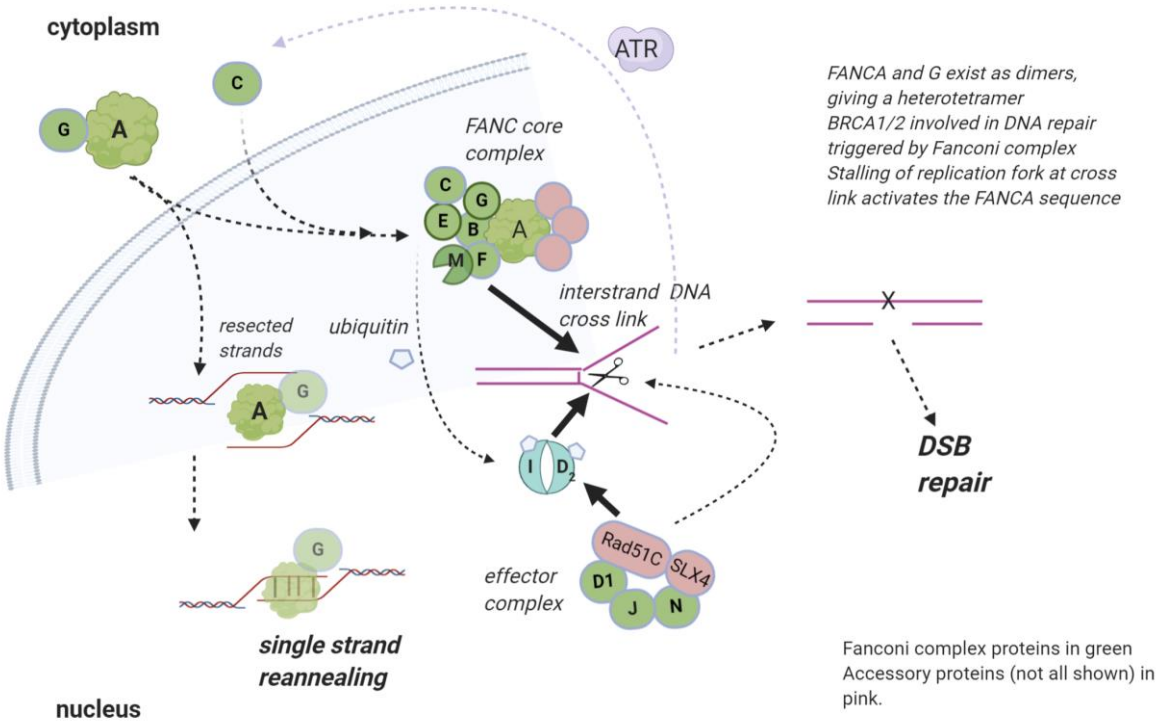


Figure 3.5: *Functions of FANCA.*

FANCA is a cytoplasmic component of the FANC complex, which is activated when FANCA/G move into the nucleus, and involved in resolution of intrastrand cross links in DNA. Failure to resolve these leads to mutation. FANCA (+FANCG) together are also involved in single strand break repair.

3.4.3 TP53

TP53 is known as a tumour suppressor protein, whose main physiological role is to prevent the proliferation of cells with DNA damage, either by preventing division of such cells or by bringing about their death (Fig. 3.6). It has been termed a 'cellular gatekeeper'[410,657], maintaining the integrity of the genome in somatic cells. Its importance is demonstrated by the fact that about half of all human cancers show somatic mutations in p53 (or genes controlling its activity) , making it the most frequently mutated gene in human cancer[658,659].

TP53 is a transcription factor, whose concentration is normally kept low by ubiquitination by MDM2 (Mouse Double Minute 2, human homolog), and subsequent proteolysis[660]. Activation of p53 occurs on DNA damage, sensed either by p53 itself (C-terminal domain), or ATM (ATM phosphorylates and switches off MDM2[661]).

Elevated p53 levels can inhibit cell growth by several mechanisms. As a transcription factor, it moves into the nucleus and promotes the transcription of cell cycle inhibitors (p21, which inhibits in G1, 14-3-3-σ which inhibits in G2/S), slowing cell proliferation and promoting senescence[662,663]. p53 acts as transcription factor for pro-apoptotic proteins, e.g.p53 upregulated modulator of apoptosis (PUMA) and Phorbol-12-Myristate-13-Acetate-Induced Protein 1 (NOXA), which promote cell death in DNA compromised cells[664]. It can also promote apoptosis directly. A modified form of p53, mono-ubiquitinylated near its C terminus can make its way out of the nucleus to the mitochondrion, to promote the action of the Bcl2 family in releasing mitochondrial cytochrome c and triggering the apoptotic cascade[665,666].

The final aspect of p53 activity is its effect on metabolism. In general it promotes oxidative and catabolic metabolism rather than the glycolytic metabolism found in many cancers[667–670]. P53 induces transcription of proteins involved in oxidative phosphorylation, such as complex I and complex IV of the respiratory chain, producing an oxidative (rather than glycolytic) phenotype, promotes transcription of TIGAR (which diverts glucose into the

pentose phosphate pathway), inhibits transcription of MCT1 needed for lactate export and stimulates production of mitochondrial glutaminase (providing) an alternative fuel[671–674].

P53 also has a direct effect on metabolism. It interacts with the kinase AMPK and these together lead to the inhibition of mTOR (a master regulator for cell proliferation) and the promotion of catabolism[675]. AMPK phosphorylates p53, leading to its stabilisation, and in turn p53 stimulates the kinase activity of AMPK (via sestrins 1 and 2) and the inhibition of mTOR through phosphorylation of tuberous sclerosis complex (TSC). In its role as transcription factor, p53 also promotes transcription of PTEN, which again inhibits mTOR[676,677].

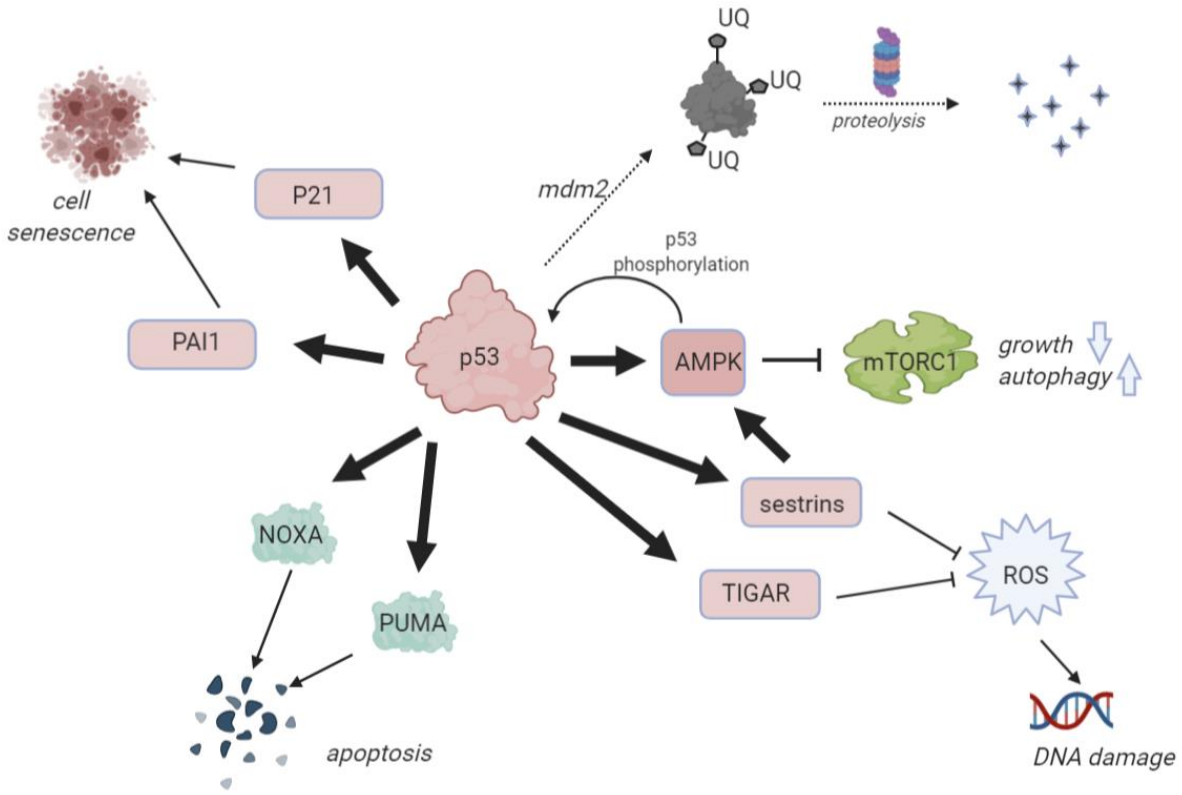


Figure 3.6: *Functions of TP53*.

p53 levels in cells are kept low by mdm2/MDM2 ubiquitination, leading to its proteolysis. If allowed to rise, or activated by DNA damage, it will suppress growth (via inhibition of mTOR, and by inducing growth inhibitory proteins, such as p21), promote apoptosis (by inducing NOXA and PUMA) and decrease glycolytic metabolism (via AMPK and TIGAR).

Mouse models of TP53 mutations have shown that reduction of p53 activity can increase cancer risk and alter metabolism, and can influence obesity in a complex and signal-dependent manner[98,603]. Within the COSMIC database, 45,299 of the 162,512 unique samples showed mutations in TP53 across a wide variety of cancer types, including in 49% of ovarian cancers (n = 4989), 47% of oesophageal cancers (n = 5567) and 44% of cancers of the large intestine[639]. SNPs involved in the expression of TP53 have also been related to gliomas and non-melanomatous skin cancers, particularly basal cell carcinomas, in a number of different studies[650,678–681]. The median odds ratios for gliomas ranged from 2.53 to 2.73 and the median for non-melanomatous skin cancer ranged from 1.41 to 2.24 (supplementary table 3.5). Further, one particular polymorphism in p53 has previously been linked to increased body weight in mice[98].

Consistent with these mouse studies, heritable (germline) mutations in TP53 in humans lead to a dramatic increase in cancer risk (Li-Fraumeni Syndrome, LFS)[602,682]. LFS is rare in the general population, with an estimated incidence of 1 in 5000-20,000. Patients with LFS are at risk of developing a variety of cancer types including carcinoma of the breast, osteosarcoma, soft tissue sarcoma, adrenocortical carcinoma and leukaemias, particularly acute lymphoblastic leukemia[683]. It is worth noting, similar symptoms can also be seen in a mutation in MDM2, which controls the level of p53 in the cell[684].

In recent studies, it has been shown that LFS patients not only have an increased risk of developing cancer but also an increased capacity for oxidative phosphorylation[685], providing a potential link with anthropometric traits and basal metabolic rate. Furthermore, the well-tolerated anti-diabetic drug metformin, which is thought to inhibit mitochondrial complex 1 in the oxidative phosphorylation system has been shown to increase cancer-free survival in a mouse model of LFS and reduce proliferation in cancer cell lines[686,687.] Thus the biguanide metformin is now being trialled in LFS patients with the aim of providing a preventative option for these high-risk patients (ClinicalTrials.gov number: NCT01981525).

Based on the potential link with oxidative phosphorylation, this intervention might also be trialled in humans carrying the minor allele of the TP53 cancer eSNP (rs78378222) and, following this work, a multi-centre trial is being orchestrated with Oxford as a major centre.

The TP53 germline mutations giving rise to LFS are rare in the general population[688]. However, in this study it is shown that a relatively frequent SNPs related to p53 affect both cancer risk and anthropometric traits. Notably, the allele of a SNP in the polyadenylation signal of p53 (rs78378222[C]) which is found in approximately 1% of populations of European descent, has been shown to impair 3'-end processing of p53 mRNA, resulting in a reduction of p53 protein[681,689]. rs78378222[C] confers an increased risk for glioma and basal cell carcinoma, as discussed in more detail above, as well as affecting head circumference, intracranial volume and cataract risk[681,690,691]. Mechanistic work in mice revealed the lengthened p53 transcript provides a miR-382-5p target site whilst compromising a miR-325-3p site[692]. The tissue-dependent differential expression of miRNAs results in p53 downregulation in the brain, where miR-382-5p targets its mRNA. These authors also describe an upregulation of p53 in the murine mammary glands in mice with two copies of the minor allele rs78378222[C] which appears protective against mammary tumours[692]. However, no such association with breast cancer is seen in the human data from the UK Biobank. Human studies have seen associations between the minor allele and an increased risk of neuroblastoma (OR = 2.3 [1.8-2.9], P= $2.03E^{-11}$, number of cases: 2101) and oesophageal squamous cell carcinoma (OR: 3.22 [1.71−6.33], P=$1.34E^{-4}$, number of cases: 405). These associations are also not detected in this study, perhaps due to the structure of the patient cohort: the median age of onset for neuroblastoma is 1-2 years and the survival relatively poor, so these subjects would not appear in the Biobank, while the majority of cases of oesophageal cancer in UK Biobank are adenocarcinoma of the oesophagus rather than squamous cell carcinoma[692].

However, in the current study we validate the associations of rs78378222[C] with non-melanomatous skin cancer and with brain malignancy seen in other cohorts (non-melanomatous skin cancer, OR= 1.46 [1.34-1.60], adjusted p= $5.20E^{-18}$, brain malignancy, OR= 3.12 [2.23-4.37], adjusted p= $1.43E^{-12}$). Post hoc analysis of the histology in the carriers of rs78378222 revealed brain cancers tended to be gliomas (81%) and notably, 16% had no histology recorded. Also, carriers of this allele tend to be taller, leaner and have a higher basal metabolic rate (standing height, adjusted-p= $2.18E^{-24}$, beta= 0.073 ± 0.007, whole body fat free mass, adjusted-p= $8.34E^{-37}$, beta= 0.073 ± 0.005, basal metabolic rate, adjusted-p= 1.13E-31, beta= 0.076 ± 0.006).

It seems likely that reduction in TP53 protein is the cause of increased cancer risk. This could be rationalised by the mechanism described above in mice, where miR-382-5p causes degradation of p53 mRNA and the consequent lack of TP53 protein, meaning it is not possible to mount a sufficient DNA damage response[692,693]. The change in anthropometric traits, particularly the association with increased lean body mass, is likely to be related to the effects of TP53 on mTORC1. TP53 is known to inhibit mTORC1 and the presence of the minor allele, rs78378222[C] reduces TP53 levels and hence increased mTOR activity (Fig. 3.7).

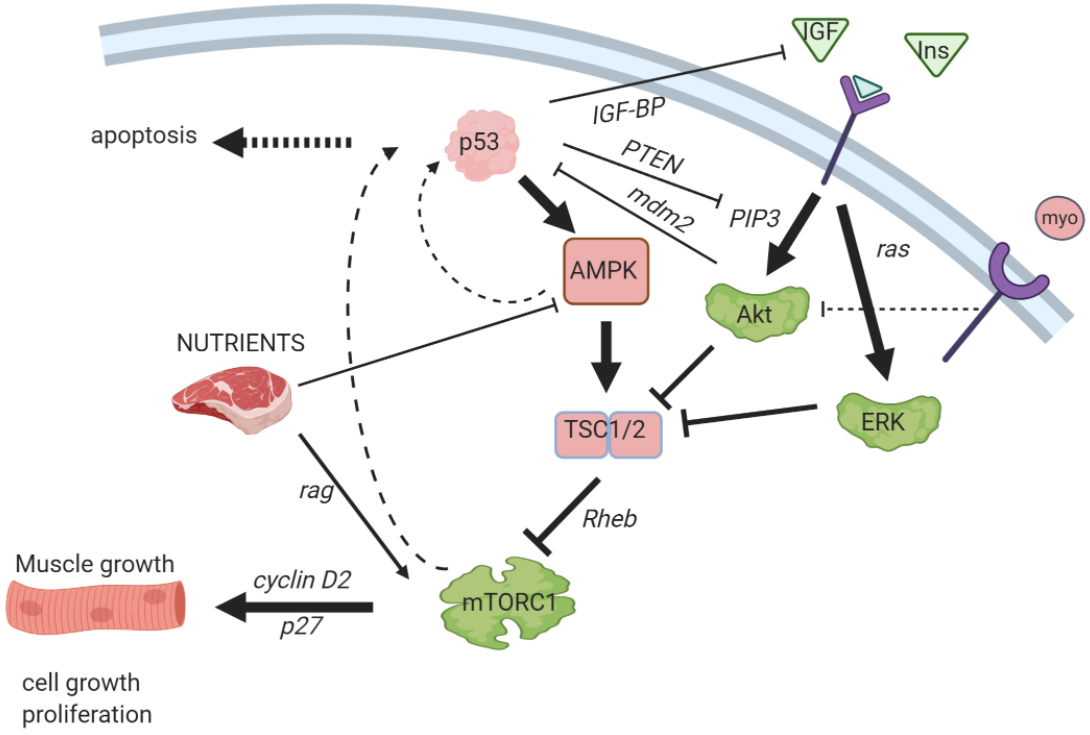


Figure 3.7: *Links between the tumour suppressor TP53 and myogenesis*.
Myogenesis (leading to increased WBFFM) is stimulated by mTORC1, which is a kinase that acts as a sensor for amino acid levels, and promotes cell growth and proliferation. mTORC1 is also activated by the insulin/IGF1/Akt signalling pathway (including the insulin receptor, with associated PIP2 kinase activity), hence the function of insulin as a growth factor. Conversely, AMPK (stimulated by TP53, and by low cellular nutritional status) will inhibit mTOR via phosphorylation of the TSC proteins and hence inhibit growth. Myostatin (myo) opposes the activation of Akt, decreasing mTOR activity and hence also inhibits muscle growth, while TP53 can also promote apoptotic cell destruction. ras, rag and rheb are G-proteins involved in modulating these pathways. Proteins promoting cell growth/proliferation are in green, while those opposing cell division are in red.

The unshackling of mTORC1 has different effects in different tissues, but all in general promote cell growth/proliferation. In bone, mTORC1 is central in regulating multiple aspects of cartilage and thus bone development. mTORC1 activation increases chondrocyte proliferation and osteogenesis[694–696]. This could lead to carriers of the minor allele being taller, as bone formation is increased. In adipose tissue, mTORC1 appears to be involved in browning[697] but its effects on overall fat mass are less clear. In muscle, mTORC1 promotes anabolism[698] which would increase lean body mass. Mice in which muscle-specific mTORC1 has been knocked out exhibit severe myopathy and premature death[699,700]. In carriers of rs78378222[C], the lack of TP53 likely leads to a reduction of mTORC1 inhibition, and this drives muscle anabolism and cellular proliferation.

This explanation is supported by work from Murphy's lab (Maureen Murphy, Wistar Institute, Philadelphia, personal communication) who have shown rs1800371 (another eSNP for TP53) is also associated with increased weight, muscle content, and enhanced performance on treadmill assays in mice. Their work demonstrated that the underlying mechanism was increased mTOR activity in the variant. However, directly demonstrating the link in humans has proved difficult as the minor allele frequency is low in European populations and, while it is higher in African descent populations (MAF 1.2-1.4%), datasets with detailed anthropometric measures are not available for these populations.

It would be interesting to investigate whether the variant rs78378222[C] showed the same properties as rs1800371. One could check (i) if the mice with the minor allele also had increased length and lean-body mass compared to the homozygote major mice (this could be done using MRI) and (ii) compare mouse embryonic fibroblasts from the wild type and carriers in assays for proliferation, gene expression and metabolomics to see if there was evidence of mTORC1 disinhibition. However, access to these mice/their tissue is not yet available.

3.5 Conclusion

The association between increased height and non-melanomatous skin cancer/brain malignancy were established prior to this study. However, mechanistic explanations were lacking[592,701]. This work demonstrates that this may be a more general phenomenon and that functional loci in well-characterised tumour suppressors and oncogenes alter both cancer risk and anthropometric traits. A strong, previously unknown, association between the rs78378222[C] SNP in the polyadenylation site of p53 with both increased risk for developing non-melanomatous skin cancer/brain malignancy and increased height, lean body mass and basal metabolic rate is identified, thereby offering a novel genetic link between these anthropometric traits and cancer risk. However, by focussing this analysis on eSNPs that affect the expression of proto-oncogenes and tumour suppressor genes, a large proportion of the genome remains untouched. This is addressed in *Chapter 4*.

Chapter 4: Genetic pleiotropy, obesity and anthropometric traits

Chapter Summary

Background: In previous work for this thesis it was demonstrated that individual SNPs can be pleiotropic for both anthropometric traits and cancer risk. However, this analysis was limited to inherited genetic variants that affect the expression of tumour suppressor genes and oncogenes. Here, the analysis is expanded across the entire genome to identify pleiotropic SNPs that associate with both phenotypes. After establishing trends from the epidemiological study, an attempt is made to identify SNPs or groups of linked SNPs that tend to be inherited together (haplotypes) which pleiotropically associate with cancer risk and anthropometric traits. It is hoped to establish whether any of these associations reflect the trends seen in the present, or other, epidemiological studies.

Methods: Genome-wide association studies (GWAS) were carried out on all anthropometric traits and clinically curated cancer groups in the UK Biobank. The analysis was then focussed on three cancers which showed differential association with body mass index, whole body fat mass and whole body fat-free mass in previous epidemiological studies. These were post-menopausal breast cancer, colorectal cancer and prostate cancer.

Results: Twenty two haplotypes were pleiotropically associated with prostate cancer and anthropometric traits in this cohort. Five haplotypes exhibit positive effects on both traits (i.e. were associated with *increased* prostate cancer risk and *increased* birth weight). eQTL analysis yielded SNPs within these haplotypes which influence expression of CD81, C11orf21, ASCL2, TH and IGF2. Eighteen haplotypes pleiotropically associate with both post-menopausal breast cancer and anthropometric traits. Two of the three haplotypes that were positively associated with post-menopausal breast cancer risk and anthropometric traits also positively associated with measures of fat-free mass. Notably, one SNP within these haplotypes, rs537626[C], which falls within a non-coding transcript exon variant of the

lncRNA CUPID1, has previously been validated as associated with breast cancer in another cohort.

Conclusion: This work identifies pleiotropic haplotypes which associate with both cancer risk and anthropometric traits in prostate and post-menopausal breast cancer. However, it was not possible to identify any haplotypes that associated with both colorectal cancer and anthropometric traits. The potential effect of rs537626[C] on the function of CUPID1, as a regulator of DNA damage repair might explain its association with an increased risk of post-menopausal breast cancer, but how it may influence fat-free mass needs further investigation.

4.1 Introduction

In previous work for this thesis it was demonstrated that individual SNPs can be pleiotropic for both anthropometric traits and cancer risk[99]. This analysis was limited to inherited genetic variants that affected the expression of tumour suppressor genes and oncogenes. This hypothesis-driven approach revealed some interesting associations, particularly a SNP found in the 3' untranslated region of the most mutated gene in human cancers, TP53, found in ~1% of the European population. This SNP, rs78378222[C], showed positive associations with non-melanomatous skin cancer and glioma together with height and fat-free mass[99]. However, this SNP-wise approach left a large proportion of the genome untouched.

Over the last 10 years, genome-wide association studies of the general population looking at cancer types or anthropometric traits have led to greater understanding of the biology of these phenotypes. GWAS is a technique that harnesses the results of genomic sequencing, which includes millions of SNPs, and relates these to phenotypic traits/outcomes. To determine whether the SNP is associated with a binary trait, e.g. a disease such as cancer, the allelic frequency of each SNP is compared between the case group (the disease group) and the control group (group without the disease). Statistical significance is then assessed to see if allele frequencies vary between the case group and the control group. The "common disease–common variant" hypothesis forms the basis of most GWAS studies, although this will frequently identify SNPs having a small effect size. However, it is also possible that rare mutations with larger effect sizes may accumulate in the population and account for a significant proportion of a common disease[702]. In truth, both hypotheses are not mutually exclusive and may well exist for complex traits, such as height.

The Genetic Investigation of ANthropometric Traits (GIANT) consortium, and others, have shed light on the genetic underpinnings of various anthropometric traits including height[703,704], body fat distribution[705] and body mass index[706,707]. Such studies have identified several key genes that influence body morphology, for instance fat mass and obesity associated gene (FTO). Indeed, SNPs in the first intron of FTO have been associated with an increase in BMI of 0.39 kg/m^2 and the risk of obesity by 1.20-fold across diverse populations. Functional and epidemiological studies suggest FTO confers an increased risk of obesity by subtly changing food preference and intake[708]. Regarding height, loci influencing Insulin-like growth factor-binding protein 3 (IGFBP3), which modulates IGF availability by increasing its half-life and inhibiting (or potentiating) IGF binding to the IGF receptor, has been strongly linked to height[704]. In addition, loci overlapping T-Box Transcription Factor 2 (TBX2) significantly associated with an altered ratio of sitting height to total height, while mutations in this gene have been previously identified as affecting heart development, limb abnormalities and skeletal malformations[704].

Many SNPs have been associated with differential cancer risk, although the number of discoveries in different tumour types has varied[709]. Links to common malignancies such as breast cancer and lung cancer have been well studied[710,711]. Notable discoveries include SNPs in tight linkage disequilibrium in a susceptibility locus on chromosome 15q. A SNP, rs1051730[A] within the CHRNA3 gene, a subunit of the nicotinic acetylcholine receptor (neuronal acetylcholine receptor subunit alpha-3), was strongly associated with smoking quantity, nicotine dependence and lung cancer, demonstrating for the first time a link between genetic preponderances towards tobacco addiction and lung cancer[712]. Although this study looked at linkages with known lung cancer effectors (and thus did not take an unbiased approach to identifying pleiotropic associations), it exemplifies the benefits of looking for pleiotropic effects in both binary and quantitative traits.

Within breast cancer studies, SNPs related to Fibroblast Growth Factor Receptor 2 (FGFR2), e.g. rs2981578[C], have consistently been associated with breast cancer risk[640,644,713–717]. The mechanism of FGFR2 action on breast cancer risk still remains controversial, but it has been shown that signalling by FGFR2 can reverse the activity of the oestrogen receptor (ER1) regulon[718]. Possibly SNPs affecting FGFR2 remove its negative influence on oestrogen signalling, priming the cells for increased proliferation.

GWAS studies have focussed on SNPs associated with single attributes (e.g. cancer risk), but rarely on how individual SNPs could simultaneously influence two attributes in the same cohort, by exhibiting genetic pleiotropy. The wealth of genomic information provided in the UK Biobank allows our work to expand, from considering inherited genetic variants that affect the expression of tumour suppressor genes and oncogenes to a more unbiased approach using GWAS. This approach allows us the exciting opportunity to search for pleiotropy on a genome-wide scale running GWAS on a variety of cancers, and anthropometric traits in the UK Biobank. After establishing trends from the epidemiological studies (*Chapter 2*), an attempt is made to identify SNPs or groups of linked SNPs that tend to be inherited together (haplotypes), and which pleiotropically associate with cancer risk and anthropometric traits. It is hoped to establish whether any of these reflect the trends seen in the present, or other, epidemiological studies. In an attempt to further identify mechanistic links, we will investigate whether any of the linked SNPs within haplotypes are involved in gene regulation, and establish whether any are expression quantitative trait loci (eQTL).

In summary, an attempt will be made (i) to establish if there are pleiotropic haplotypes that associate with both anthropometric traits and cancer risk in the UK Biobank ii) if so, whether the direction of association mirrors that in our epidemiological models and iii) to investigate potential mechanisms by looking for expression quantitative trait loci amongst these haplotypes.

4.2 Methods

4.2.1 The UK Biobank

The UK Biobank is a cohort of ~500,000 UK residents who volunteered to have their clinical, lifestyle, anthropometric and genetic data collected for research. Details on anthropometric, cancer and genetic data are covered in *Chapters 2 & 3*. This study was conducted under the UK Biobank approved application (#43313, PI Francesca Buffa).

4.2.2 Genomic Association Analyses

SNP-only based genetic association analysis of quantitative (anthropometric traits) and dichotomous (cancer occurrence) variables were carried out under an additive model using linear and logistic statistical framework, respectively, in PLINK v2.00a2LM. In principal component analysis, age and genetic principal components (1–10 PCs) were used as covariates to control for hidden population structure in GWAS. Sex was included as a covariate where appropriate, namely when testing anthropometric traits and colorectal cancer.

4.2.3 Pleiotropic Associations

After running a GWAS analysis on cancers and anthropometric traits of interest, we selected the pleiotropic associations following the steps described below:

1. *Anthropometric associations:* SNPs significantly associated with anthropometric traits (23) were selected using a GWAS threshold-p of $5E^{-8}$.
2. *Haplotype calculation for anthropometric associated SNPs:* Haplotypes were calculated using PLINK V. 1.9 (R^2>0.8) for the significant SNPs identified in step 1. All SNPs without attributed rsID could not be given haplotypes (17,766 out of 338,458) and so were excluded from the next step.

3. *Searching for cancer pleiotropic associations:* SNPs which significantly associated with anthropometric traits within haplotypes, were assessed for association with cancers. Associations with post-menopausal breast, colorectal or prostate cancer were deemed significant if a SNP's association-p was below a Bonferroni corrected threshold. This threshold was calculated by correcting 0.05 by the number of haplotype blocks calculated in step 2 (43,499) and the number of tested anthropometric traits (23). The resultant threshold-p was $5.05E^{-8}$.

This workflow is summarised in Figure 4.1.

4.2.4 Expression quantitative trait loci

SNPs associated with differential gene expression were identified from expression quantitative trait loci (eQTL) databases: GTEX , NESDA/NTR and PancanQTL[619–621]. These are described in more detail in *Chapter 3*, including tissue types included and number of samples. SNPs that are in eQTL, are noted as eSNPs.

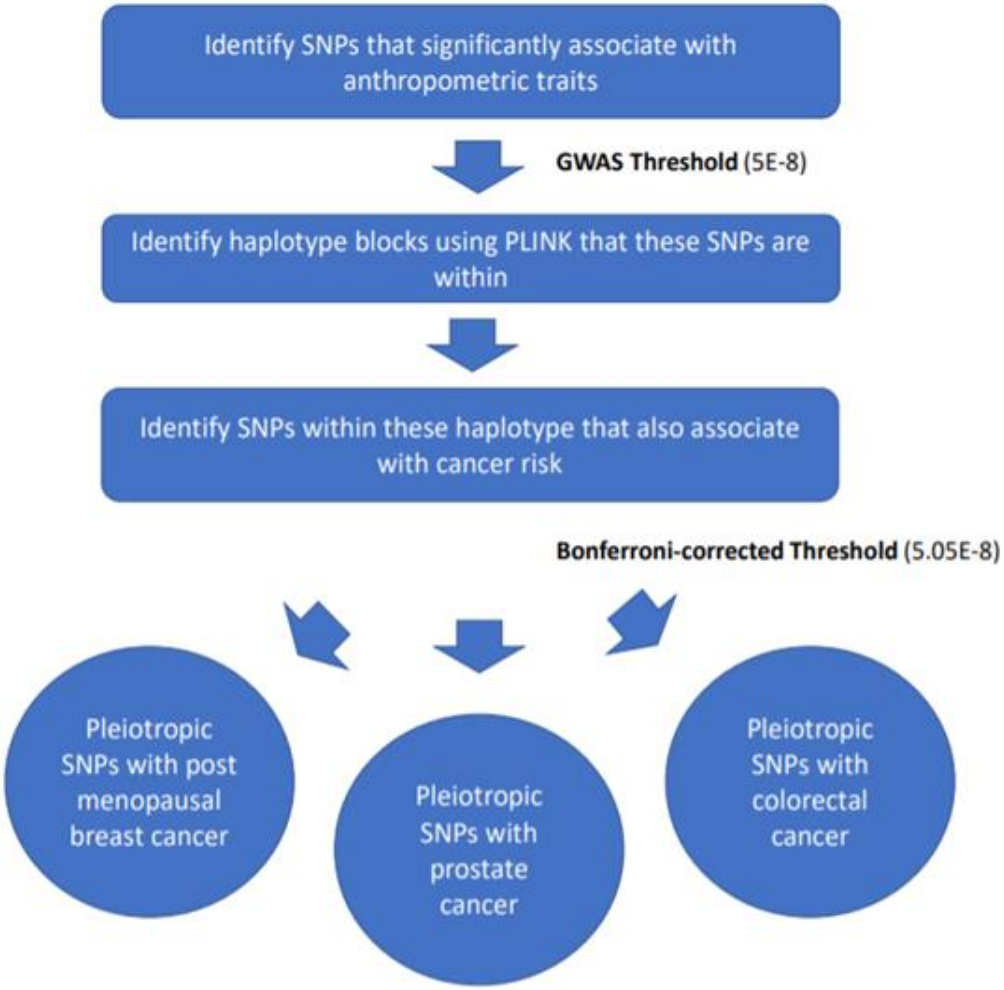


Figure 4.1: *Identification of pleiotropic SNPs.*

Workflow used to identify pleiotropic SNPs for anthropometric traits and cancer risk. Associations with post-menopausal breast, colorectal or prostate cancer were deemed significant if a SNP's association-p was below a Bonferroni-corrected threshold ($5.05E^{-8}$) calculated by correcting 0.05 by the number of haplotype blocks (43,499) and the number of tested anthropometric traits (23).

4.3 Results

4.3.1 GWAS of BMI, WBFFM and WBFM

4.3.1.1 BMI

46,914 SNPs were found to significantly associate with BMI. The one with the lowest p is rs11642015[T] on chromosome 16 ($p = 2.17E^{-213}$, beta = 0.076), see Fig. 4.2. rs11642015[T] has previously been associated with several traits previously including BMI[719,720], urinary sodium excretion[721] and various aspects of blood pressure[722].

Our most significant 11 eSNPs affected the expression of AKT Interacting Protein (AKTIP). These include rs8050136[A] ($p = 3.3071E^{-197}$, beta = 0.074), rs56313538[G] ($p = 6.9579E^{-197}$, beta = 0.074) and rs11075988[A] ($p = 7.73E^{-197}$, beta = 0.074). rs8050136[A] has been associated with BMI and obesity in a number of previous studies[723–726]. Indeed, all eSNPs affecting the expression of AKTIP were also intronic variants for fat mass and obesity-associated gene (FTO), although expression of FTO itself was unaltered in the eQTL databases interrogated.

AKTIP can regulate apoptosis through its interaction with (serine/threonine) protein kinase B (Akt). Mutations in the mouse homologue of AKTIP can lead to fused toes and thymic hyperplasia[727,728]. AKTIP also appears to act as a factor required for telomere maintenance and replication[729]. Therefore, if either or both of these functions are altered, it is easy to see that BMI might ultimately be affected.

The next most significant eSNP in this study, rs1861867[G] ($p = 4.877E^{-102}$, beta = -0.055) affected FTO expression directly. The function of FTO has been extensively studied, although its functional link to human BMI has not been entirely explained. FTO shares commonalities with Fe (II) and 2-oxoglutarate (2OG) oxygenases, which catalyse oxidative

reactions using non-heme iron as a cofactor and 2OG as a co-substrate[730,731]. This allows it to act as a modulator of metabolic processes, as has been shown in model organisms. However, human studies suggest FTO can also influence appetite and food intake[732,733].

182 eSNPs influence the expression of the cytoplasmic peripheral membrane protein, Syntrophin Gamma 2, which interacts with membrane proteins such as neuroligins 3 and 4X[734]. As yet it is unclear how this may influence body mass index but could provide an interesting route for further study. All SNPs associated with BMI are found in supplementary table S4.1.

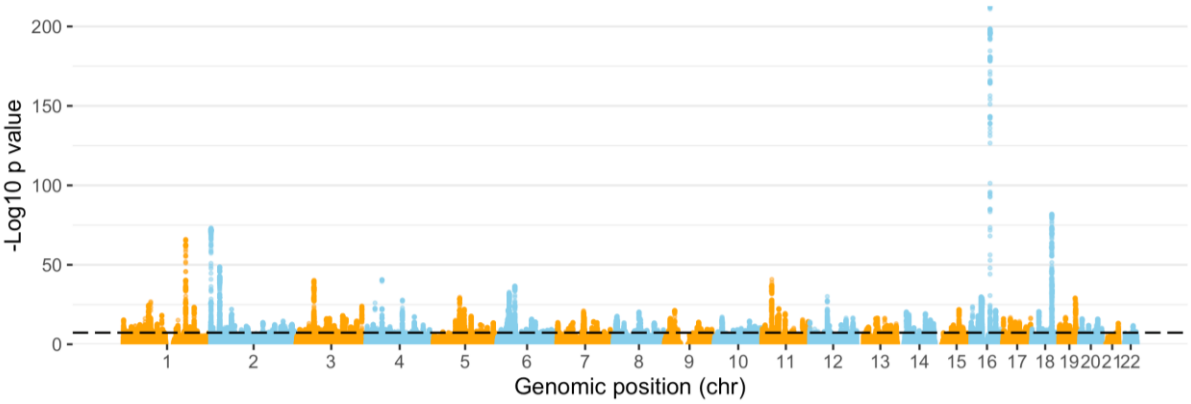


Figure 4.2: *SNPs associated with BMI in the UK Biobank*.

Manhattan plot, showing chromosome locus on horizontal axis, and $-\log_{10}(p)$ value for independence on vertical axis (i.e. a high vertical distance indicates a probable association). Dotted line represents $5E^{-8}$ (widely accepted GWAS threshold).

4.3.1.2 WBFFM

85,487 SNPs were found to significantly associate with WBFFM. The one with the lowest p is rs6567160[C] on chromosome 18, ($p = 5.10E^{-143}$, beta =0.040), see Fig. 4.3. rs6567160[C] is an intergenic variant in RNU4-17P at 18q21, and has been associated with BMI[706,735–737], adiposity[738] and lean body mass in a study of anorexia nervosa in the UK Biobank cohort[739]. However, the potential functional link relating this SNP to WBFFM is unclear.

The most significant eSNP is rs143384[G], $p = 1.5531E^{-133}$, beta = 0.033). This affects the expression of several genes viz: Endoplasmic reticulum-Golgi intermediate compartment protein 3, Eukaryotic translation initiation factor 6 (EIF6), Matrix Metallopeptidase 24 (MMP24) and Copine 1. Perhaps, the most likely to affect whole body fat free mass is the initiation factor, EIF6. EIF6 influences insulin-stimulated translation, including the mTOR pathway, which is involved in cellular proliferation and notably muscle growth (*Chapter 3*)[740]. It is required for full efficiency of oxidative phosphorylation in muscle[741]. EIF6 also impacts lipid metabolism more directly, promoting lipid synthesis in liver and adipose tissue by exerting translational control of adipogenic transcription factors[742]. Any or all of these targets might be expected to affect WBFFM.

Another SNP highly significantly linked to WBFFM is rs9939609[A] ($p = 2.5647E^{-124}$, beta = 0.032). This is another eSNP to AKTIP. As noted above, this protein interacts with the insulin signalling system via protein kinase B (Akt), and hence is linked to anabolism in muscle tissue. Interestingly, expression of AKTIP is inhibited by miR-20b-5p, an exosomal miRNA raised in Type 2 Diabetes (which is strongly associated with obesity), providing a possible link between muscle mass and obesity[743].

This study identifies rs78378222[C], as the eSNP with second largest beta value in the genome for WBFFM (beta = 0.072, p = 2.68E^{-32}). This eSNP affects the expression of TP53[681,689] and was also identified in Chapter 3 as a variant associated with WBFFM. This shows that an unbiased approach by GWAS validates these findings and contextualises them within the whole genome through broader comparison.

GWAS also identifies rs76895963[G] (beta = 0.091, p = 7.04E^{-71}) as strongly associated with WBFFM. This is an eSNP for RP11-264F23.4, RP11-264F23.3 and cyclin D2 (CCND2), and has previously been associated with BMI and appendicular lean mass by other workers using this dataset as part of their analyses[744,745], possibly through effects on cell proliferation through cyclin D2 which has an important role in progression through the G1/S transition[746]. All SNPs associated with WBFFM are found in supplementary table S4.2.

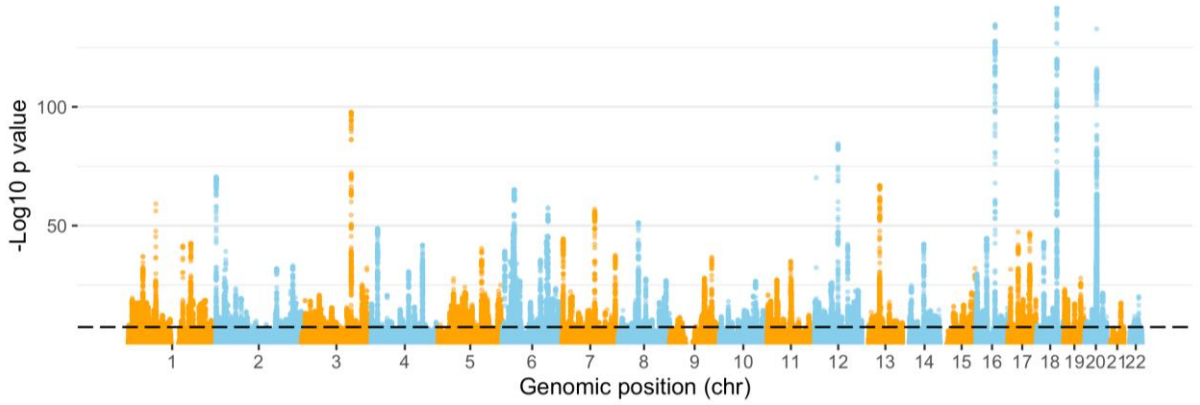


Figure 4.3: *SNPs associated with WBFFM in the UK Biobank.* Conventions as in Figure 4.2.

4.3.1.3 WBFM

41,092 SNPs were found to significantly associate with WBFM. The one with the lowest adjusted-p is rs11642015[T] on chromosome 16, (lowest adjusted-p: $2.88E^{-163}$, beta =0.065), see Fig. 4.4.

The most significant positive associations with eSNPs were obtained with AKTIP (rs8050136[A], adjusted-p = $3.18E^{-150}$, beta = 0.063), RP11-568K15.1(rs6670873[T], adjusted-p = $1.6508E^{-45}$, beta = 0.041); and C6orf106 and C6orf107 (rs11455465[C>CT], adjusted-p = $2.05E^{-44}$, beta = 0.047).

AKTIP will affect fuel storage in the body, and hence fat content, via the insulin response pathway, as outlined in previous sections. The eSNP association with the pseudogene, RP11-568K15.1, is expressed as RNA but, as a pseudogene, produces no protein product. It has a sequence related to a quinone oxidoreductase, Crystallin zeta like 2, pseudogene (CRYZL2P). It can be expressed as a readthrough product with SEC16B, an endoplasmic reticulum protein transporter.

C6orf106 and C6orf107 are a pair of genes on chromosome 6 that have been associated with anthropometric traits in previous GWAS studies[735,706]. As their protein product has become better defined, they have been renamed ILRUN (Inflammation And Lipid Regulator With UBA-Like And NBR1-Like Domains) and UHRFBP1 (Ubiquitin-Like Containing PHD And RING Finger Domains 1-Binding Protein 1) respectively. UHRFBP1 is also known as ICBP90. Both have ubiquitin-like domains, which have been generally associated with trafficking proteins to the proteosome, DNA methylation and potentially affecting the cell cycle[747,748]. Intriguingly, however, recent work has shown that ILRUN is involved in lipoprotein metabolism, decreasing PPARg production and upregulating HDL and VLDL secretion (i.e. triglyceride output) from the liver[749]. This could potentially explain its association with whole body fat mass. All SNPs associated with WBFM are found in supplementary table S4.3.

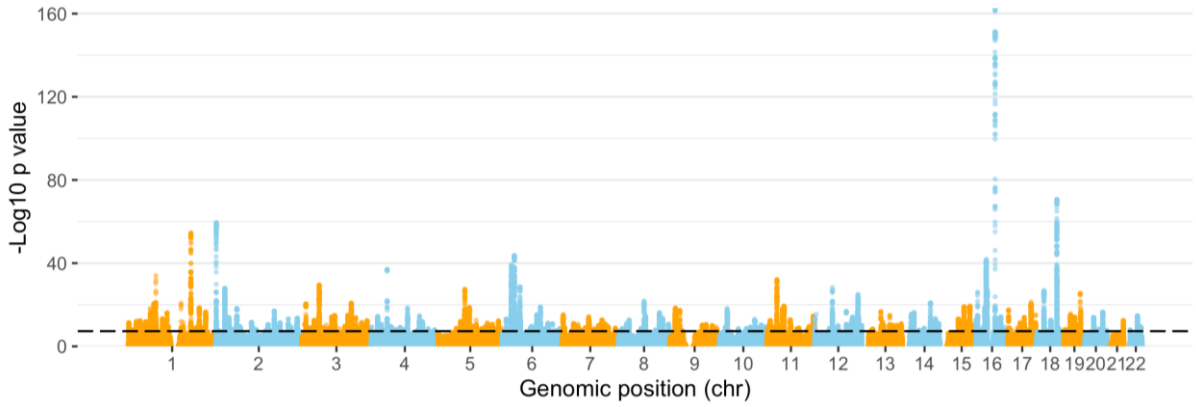


Figure 4.4: *SNPs associated with WBFM in the UK Biobank.* Conventions as in Figure 4.2.

4.3.2 GWAS of Colorectal, Prostate and Post-Menopausal Cancers

4.3.2.1 Colorectal

169 SNPs significantly associated with colorectal cancer. The one with the lowest adjusted-p is rs6983267[T]) on chromosome 8 (lowest adjusted-p: $7.18E^{-21}$, OR =0.83), see Fig. 4.5. The alternate allele, rs6983267[G], has been shown to be associated with an increased risk of colorectal cancer in a large number of different studies[750–758]. These observations are in line with the protective effect of the T variant seen here. rs6983267 is mapped 8q24.21 and is a non-coding transcript intron variant in CCAT2. CCAT2 is a long non-coding transcript (lncRNA), which has been shown to promote activity of the myc oncogene in colorectal cancer models[759]. This SNP has also been linked with prostate cancer in a number of studies[760–762].

Indeed, 35 of these 169 SNPs were located in this small region of chromosome 8. One of these 35, rs6983549[A], was an eSNP with POU Class 5 Homeobox 1B (POU5F1B) within colon tissue in the GTEX database[619]. POU5F1B overexpression occurs in a number of cancers, and induction in model systems has promoted colony formation in vitro as well as both tumorigenicity and tumor growth in vivo. These effects were enhanced by myc overexpression[763–765].

The next two most significant eSNPs are found on chromosome 15, rs58658771[A] (adjusted-p: $1.27E^{-13}$, beta = 1.21) and rs73376930[G] (adjusted-p: $6.50E^{-13}$, OR = 1.20)[619]. Both have been associated with colorectal cancer in different studies[750,756,766–768]. Interestingly, rs58658771[A] is an intergenic variant of Secretogranin-5 (SCG5) while rs73376930[G] is an intronic variant in Gremlin 1 (GREM1). Both affect the expression of the lncRNA, RP11-758N13.1, and GREM1 expression. GREM1 has previously been linked with various stages of colorectal cancer, and in particular metastases. However, its overall effect on colorectal carcinogenesis is not clear cut, perhaps due to its involvement in the complex

TGF$\boldsymbol{\beta}$ pathway[769,770]. The function of lnc-RNA, RP11-758N13.1 is uncertain. All SNPs associated with colorectal cancer are found in supplementary table S4.4.

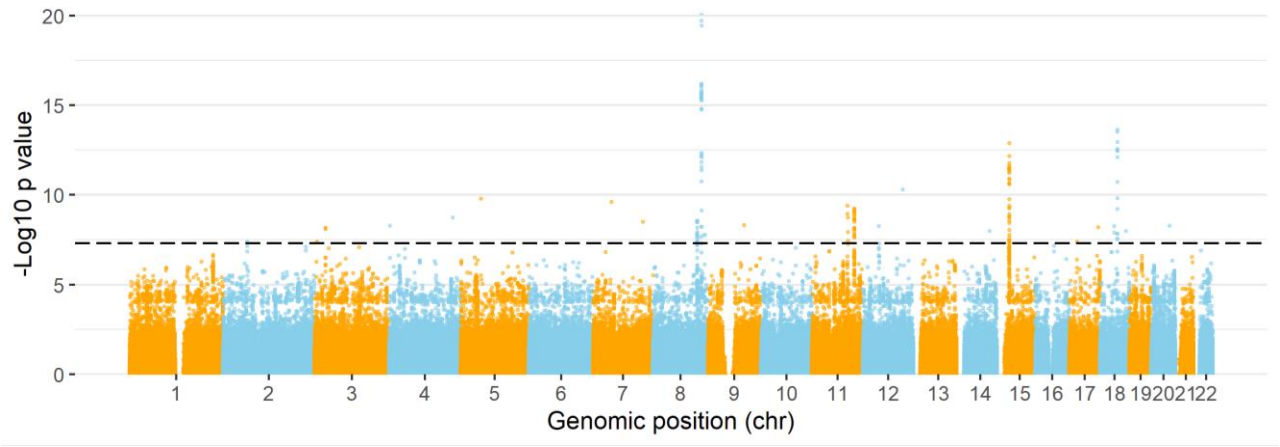


Figure 4.5: *SNPs associated with Colorectal cancer in the UK Biobank.* Conventions as in Figure 4.2.

4.3.2.2 Prostate

2209 SNPs significantly associated with prostate cancer. The most strongly associated SNP, rs77541621[A] (p: 2.12E^{-44}, OR = 1.80), see Fig. 4.6, has been shown in other studies to associate with prostate cancer[771] and is a non-coding transcript exon variant mapped to Prostate Cancer Associated Transcript 1 (PCAT1). PCAT1 produces a long non-coding RNA which promotes cell proliferation by negatively regulating the BRCA2 tumour suppressor protein and positively regulating myc oncoprotein (reviewed in detail here[772].)

The eSNP most significantly associated with prostate cancer is rs11263763[G], p = 4.72E^{-34}, OR = 0.81. Its opposite allele rs11263763[A] has been associated with an increased risk of prostate cancer in other studies[771]. This SNP is an eQTL for RP11-115K3.1, which is an antisense RNA whose function is not yet established.

Other highly significant eSNPs include rs10993994[C], which influences expression of the androgen receptor coactivator Nuclear Receptor Coactivator 4 (NCOA4) and one of the three most abundant proteins secreted by the prostate, Microseminoprotein-β (MSMB)[773]. Both of these genes have been linked strongly to prostate carcinogenesis[762,774–776] and this eSNP, rs10993994[C], has been linked to prostate cancer in a number of different studies[762,771,777]. Other eSNPs affect the well-known proto-oncogene transcription factor myc, (rs1447292[C] p: 5.22E^{-32}, OR = 1.35), and POU5F1B, which also appears in the colorectal eQTL analysis above (rs6983549[A], p = 4.80E^{-32}, OR = 1.22). All SNPs associated with prostate cancer are found in supplementary table S4.5.

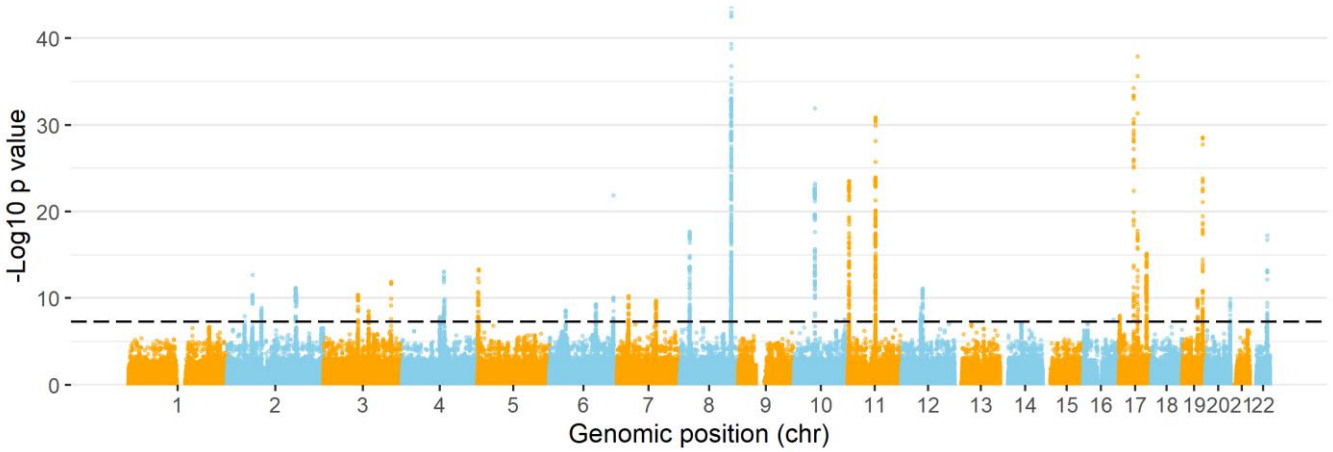


Figure 4.6: *SNPs associated with Prostate cancer in the UK Biobank.* Conventions as in Figure 4.2.

4.3.2.3 Post-menopausal breast cancer

1336 SNPs significantly associate with post-menopausal breast cancer. The most significant, rs4784227[T] (lowest p: $5.04E^{-50}$, OR = 1.28), is an intron variant mapping to a lncRNA, Cancer Susceptibility 16 (CASC16) on chromosome 16, see Fig. 4.7. rs4784227[T] is a risk allele for breast cancer in a variety of populations, within both oestrogen receptor positive and negative subtypes[778–782]. This SNP is also an eSNP for TOX High Mobility Group Box Family Member 3 (TOX3). TOX3 is located in chromosome16q12. Loss of heterozygosity in this 16q region is commonly observed in breast cancers (33.9% primary tumours), including a 2.3% frequency of homozygous deletions[783]. It has been proposed that TOX3 is a tumour-suppressor gene present on the 16q arm which might be involved in unwinding of DNA, although its function is not unequivocally established[784,785].

Other highly significant eSNPs include rs7895676[C] ($p = 8.75E^{-37}$,OR = 1.21), which is associated with changes in expression of NSE4 Homolog A, SMC5-SMC6 Complex Component (NSMCE4A). This protein is involved in both homologous recombination and lengthening and maintenance of telomeres via the alternative lengthening of telomeres pathway[786] potentially impacts one hallmark (*Enabling Replicative Immortality)* and one enabling characteristic (*Genome Instability and Mutation)* of cancer[34] (see *Chapter 1*).

Another eSNP identified in this analysis, rs12653202[C] ($p = 7.66E^{-15}$, OR = 1.16), affects the expression of MAP3K1, which can promote cell growth but also promote apoptosis via caspase activation through p53 and BCL-2[637,638]. Notably, eSNPs of MAP3K1 (*Chapter 3*) were previously identified as associated with anthropometric and oncogenic pleiotropy. All SNPs associated with post-menopausal breast cancer are found in supplementary table S4.6.

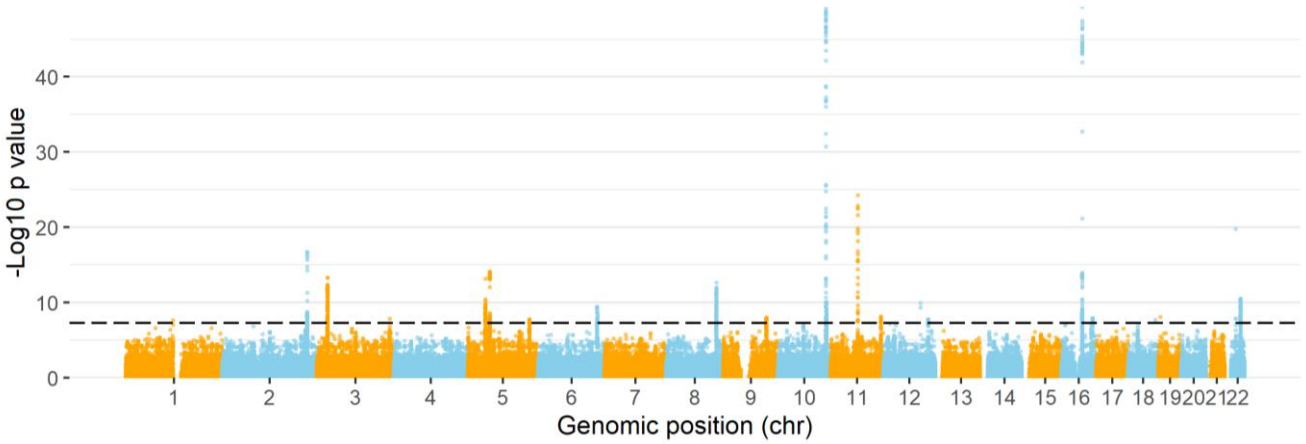


Figure 4.7: *SNPs associated with Post-menopausal breast cancer in the UK Biobank.* Conventions as in Figure 4.2.

4.3.3 Simultaneous anthropogenetic and oncogenic pleiotropy

Above I show associations between individual SNPs and cancer risk, or anthropometric changes, identified by GWAS. These SNPs were then inspected for those which affected both conditions i.e. pleiotropic SNPs.

Although 169 SNPs associated with colorectal cancer (supplementary table S4.4), no haplotypes or SNPs exhibited pleiotropic associations with colorectal cancer and anthropometric traits.

In prostate cancer, 22 haplotypes were associated with both cancer and anthropometric traits (supplementary tables S4.7-4.9). Five haplotypes were discovered to associate positively with both (Fig. 4.8). All haplotypes associated with both an increased risk of prostate cancer and increased birth weight were located on chromosome 11; 2201338:2280751. The strongest pleiotropic association with prostate cancer was observed with rs10743180[G], with an p of $2.94E^{-24}$ and OR of 1.23. The strongest association with birth weight was rs11564709[T] with an p of $2.13E^{-10}$ and beta = 0.035. Other works have identified SNPs rs7126629[C][787] and rs7127900[A], found in one of the haplotypes, as risk alleles for prostate cancer[788–790]. And indeed, another SNP in this haplotype, although not associated with birth weight, has been shown to be linked to increased BMI (rs10840606[G])[735,791,792].

Four haplotypes were associated with both a decreased risk of prostate cancer and reduced anthropometric traits. One haplotype, rs7819881[C] associates with a decreased prostate cancer risk (p, $3.97E^{-08}$, OR: 0.90) and decreased standing height ($1.18E^{-08}$, beta = -0.010). SNPs within the other three haplotypes that also appear protective against prostate cancer in men and associate with a reduced waist-hip ratio. Interestingly, despite being protective against prostate cancer, the same haplotypes were associated with an increased risk of post-menopausal breast cancer (associations in Fig. 4.8 and further information in

supplementary table S4.8). For example, the lead SNP in one of these haplotypes, rs382434[T], was associated with a decreased risk of prostate cancer ($p = 5.47E^{-15}$, OR 0.83), waist-hip ratio ($p = 5.16E^{-09}$, beta = -0.0010) and an increased risk of post-menopausal breast cancer ($p = 3.93E^{-08}$, OR: 1.09). We were able to validate the associations of several SNPs in one of these haplotypes with waist-hip ratio (rs13256367[C], rs378854[T], rs366991[G], rs620861[A]) and prostate cancer (rs445114[C]) within the UK Biobank, and this has been noted in other cohorts from independent studies[735,789,791–795]. 13 haplotypes that associated with anthropometric traits and prostate cancer risk in opposite directions were identified (supplementary table S4.9 & S4.10), i.e. larger cancer risk is associated with smaller size of anthropometric trait and vice versa. These traits include various measures of fat-free mass, standing height, sitting height, hip circumference, waist-hip ratio and birth weight.

To enable functional analysis of these changes, we turn now to eSNPs and their expression profiles. Associations between eSNPs and both prostate cancer and anthropometric traits are also found in both positive and negative directions. Two haplotypes containing eSNPs associated with both increased prostate cancer risk and increased birth weight. One haplotype contained eSNPs for Achaete-Scute Family BHLH Transcription Factor 2 (ASCL2); Chromosome 11 Open Reading Frame 21(C11orf21); Tyrosine Hydroxylase (TH) and Insulin Like Growth Factor 2 (IGF2). eSNPs affecting IGF2 were the most frequent in this haplotype. Diseases caused by issues with IGF2 include Silver-Russell Syndromes 1 and 3. Patients with either of these syndromes have altered growth of various anthropometric traits, particularly slow fetal growth before birth[796]. IGF2 mRNA in chorionic villus samples at 11-13 weeks' gestation has been shown to significantly correlate with birth weight[797].

The second haplotype contained eSNPs for CD81 and C11orf21. C11orf21 was mentioned in connection with the haplotype above and all pleiotropic SNPs in this haplotype

were eSNPs for C11orf21. C11orf21 is a protein of 132 amino acids long that is expressed almost exclusively in the adult heart, and has been proposed to be related to Beckwith–Wiedemann syndrome, signs of which include overgrowth, particularly macroglossia and Wilms tumours[798]. Its function is not well elucidated. SNPs within this haplotype containing eSNPs for C11orf21 (namely rs7126629[C] and rs7127900[A]) have been associated with increased prostate cancer risk in other studies[787–789,799].

One haplotype was negatively associated with both anthropometric traits and prostate cancer. It contains eSNPs for NK3 Homeobox 1 (NKX3-1) and Solute Carrier Family 25 Member 37 (SLC25A37). NKX3-1 encodes a homeobox-containing transcription factor that has been shown to influence prostate cancer risk and progression. This protein is the earliest known marker of prostate epithelium during embryogenesis and is expressed at all stages of prostate differentiation[800–802]. NKX3.1 tends to be lost in prostate carcinogenesis, whilst inducing NKX3.1 expression inhibits proliferation and invasion of the prostate cancer cell line, PC-3. It appears to act by increasing the stability of p53[803,804]. Due to its highly restricted expression to prostate epithelial cells, NKX3.1 is used as a diagnostic biomarker for prostate cancer; it has been reported to have a 98.6% sensitivity for identifying metastatic prostate adenocarcinoma better than Prostate Specific Antigen (PSA) at only 94.2%[805]. SLC25A37 is a solute carrier located in the mitochondrial inner membrane which imports iron for the synthesis of haem and iron-sulfur clusters[806]. It is less clear functionally how it influences prostate carcinogenesis, although its expression has been linked to changes in fatigue scores with prostate cancer patients undergoing external beam radiotherapy[807] .

Other SNPs in this region (e.g. rs1512268[A] and rs142463603[A]) have been associated with altered prostate cancer risk[771]. Intriguingly, risk varies between the SNPs rs1512268[T], and rs1512268[A], in different ethnic cohorts[761,808].

Seven eSNPs within one haplotype (rs11995896[A], rs13274284[A], rs2314687[C], rs28370549[A], rs4872166[C], rs4872167[G], rs7819881[C]) associate with a decreased risk of prostate cancer and reduced standing height (lead SNP, rs7819881[C], prostate cancer p = $3.97E^{-08}$, beta =0.90, standing height p = $1.18E^{-08}$, beta = -0.01). The link with p53 and its downstream effects on mTOR might help explain the reduction seen in standing height, with the importance of mTOR in long bone development having been illustrated in model organisms[809]. Six other haplotypes contained eSNPs that were associated with cancers and anthropometric traits, including eSNPs for genes such as Ten-Eleven Translocation 2 (TET2), which modulates DNA hydroxymethylation by converting 5-methylcytosine (5 mC) to 5-hydroxymethylcytosine (5 hmC) to promote DNA demethylation, and Rho guanine nucleotide exchange factor 38 (ARHGEF38) which has been linked in other studies to prostate cancer aggressiveness[810,811].

As regards post-menopausal breast cancer, 18 haplotypes associate with both post-menopausal breast cancer risk and anthropometric traits using the genome-wide approach. Three haplotypes were positively associated with post-menopausal breast cancer risk and anthropometric traits (Fig 4.9). Two of these showed a positive association between increased fat-free mass and increased post-menopausal breast cancer risk, as seen in the epidemiological analyses (*Chapter 2*). Indeed, the lead SNP in one haplotype, rs1485995[G], was associated with post-menopausal breast cancer (p = $1.78E^{-11}$, OR: 1.11) and Arm Fat-Free Mass Right (p = $2.48E^{-09}$, beta = 0.0083); Leg Fat-Free Mass Right (p = $2.51E^{-12}$, beta = 0.011); Whole Body Fat-Free Mass (p = $8.24E^{-12}$, beta = 0.0096); Basal Metabolic Rate (p = $4.66E^{-11}$, beta = 0.010; Arm Fat-Free Mass Left (p = $1.30E^{-09}$, beta = 0.0089); Leg Fat-Free Mass Left (p = $4.92E^{-12}$, beta = 0.011); Trunk Fat-Free Mass (p = $1.73E^{-10}$, beta = 0.0089. The lead SNP in the other haplotype, rs615029[T], was associated with increases in post-menopausal breast cancer (p = $1.56E^{-17}$,OR 1.19), Whole Body Fat-Free Mass (p = $2.88E^{-08}$, beta = 0.010) and Trunk Fat-Free Mass (p = $2.89E^{-08}$, beta = 0.010). Further, the breast

cancer association of a SNP in this haplotype, rs537626[C], was observed in a previous case-control study of 15,170 individuals with 6993 cases ($p = 1.8E^{-15}$, OR: 1.29)[642].

Interrogation of eQTL databases showed that the haplotypes mentioned above are not associated with changes in gene expression. However,the eSNP rs537626[C] has severe consequences, by being a non-coding transcript exon variant related to long intergenic non-protein coding RNA 1488 (LINC01488) otherwise known as CUPID1 (CCND1 Upstream intergenic DNA Repair 1, non-coding). In another haplotype the lead SNP, rs7117818[T], was positively associated with post-menopausal breast cancer ($p =2.43E^{-08}$, OR: 1.10) and standing height ($p =1.79E^{-11}$, beta = 0.013). No haplotypes were negatively associated with cancer risk and anthropometric traits.

This work identified 16 haplotypes that associate with anthropometric traits and post-menopausal breast cancer risk in opposite directions (supplementary table S4.13-S4.14). Traits include various measures of fat mass, fat-free mass, standing height, seated height, BMI, weight, hip circumference, waist:hip ratio and birth weight. Some of these anthropometric/cancer associations could be validated in the GWAS Catalog in a number of different cohorts[640,644,812,813], and the list of all SNPs associated with post-menopausal breast cancer is found in supplementary table S4.6.

Genes found to be altered in expression related to these pleiotropic SNPs are discussed below. For example, AKT Interacting Protein (AKTIP) was a gene highlighted when considering the most significant SNPs associated with BMI in the GWAS analysis. Here it is also identified as an eQTL containing SNPs within a haplotype that negatively associates with post-menopausal breast cancer but positively associates with a variety of fat and fat-free mass measures and also BMI (rs11075988[A], rs11075989[T], rs11075990[G], rs11075991[T], rs3751814[A], rs56313538[G], rs8050136[A], rs9923233[C], rs9923312[A], rs9936385[C], rs9939609[A]; more information in supplementary table S4.13).

In agreement with the previous SNP-wise study (*Chapter 2*), the current approach also identifies MAP3K1 eSNPs - including rs889312[C] - as associating with both post-menopausal breast cancer and various measures of height (post-menopausal breast cancer, p: $2.08E^{-08}$, OR = 1.10, standing height, p: $9.62E^{-16}$, beta= -0.015, seated height, p:$1.31E^{-08}$, beta= -0.012, sitting height, p: $2.80E^{-12}$, beta: -0.014). Indeed, MAP3K1 appeared a common theme amongst eGenes with identified haplotypes that associate with increased risk of post-menopausal breast cancer but with decreased measures of height.

In a second haplotype, a Ribosomal Protein L37a (RPL37A) eSNP, rs10932692[T], was identified as associating with increased post-menopausal breast cancer risk (p: $5.72E^{-11}$, OR = 1.11) but reduced standing height[646,728] (p: $9.05E^{-09}$, beta = -0.0098).

It is noteworthy mutations in both AKTIP and MAP3K1 cause morphological abnormalities. Mutations in mouse homolog of AKTIP produces fused toes and thymic hyperplasia, and mutations in MAP3K1 are related to gonadal dysgenesis[646,728]. Both appear to influence apoptotic pathways, MAP3K1 via caspase activation and phosphorylated MEK4/7 which acts on p53 and BCL-2[637] and AKTIP through interaction with A- and B-type lamins influencing telomere homeostasis[814]. It is less clear how RPL37A might influence both phenotypes, but as part of the ribosomal machinery perhaps RPL37A eSNPs may alter translation of proteins in pathways that influence both traits.

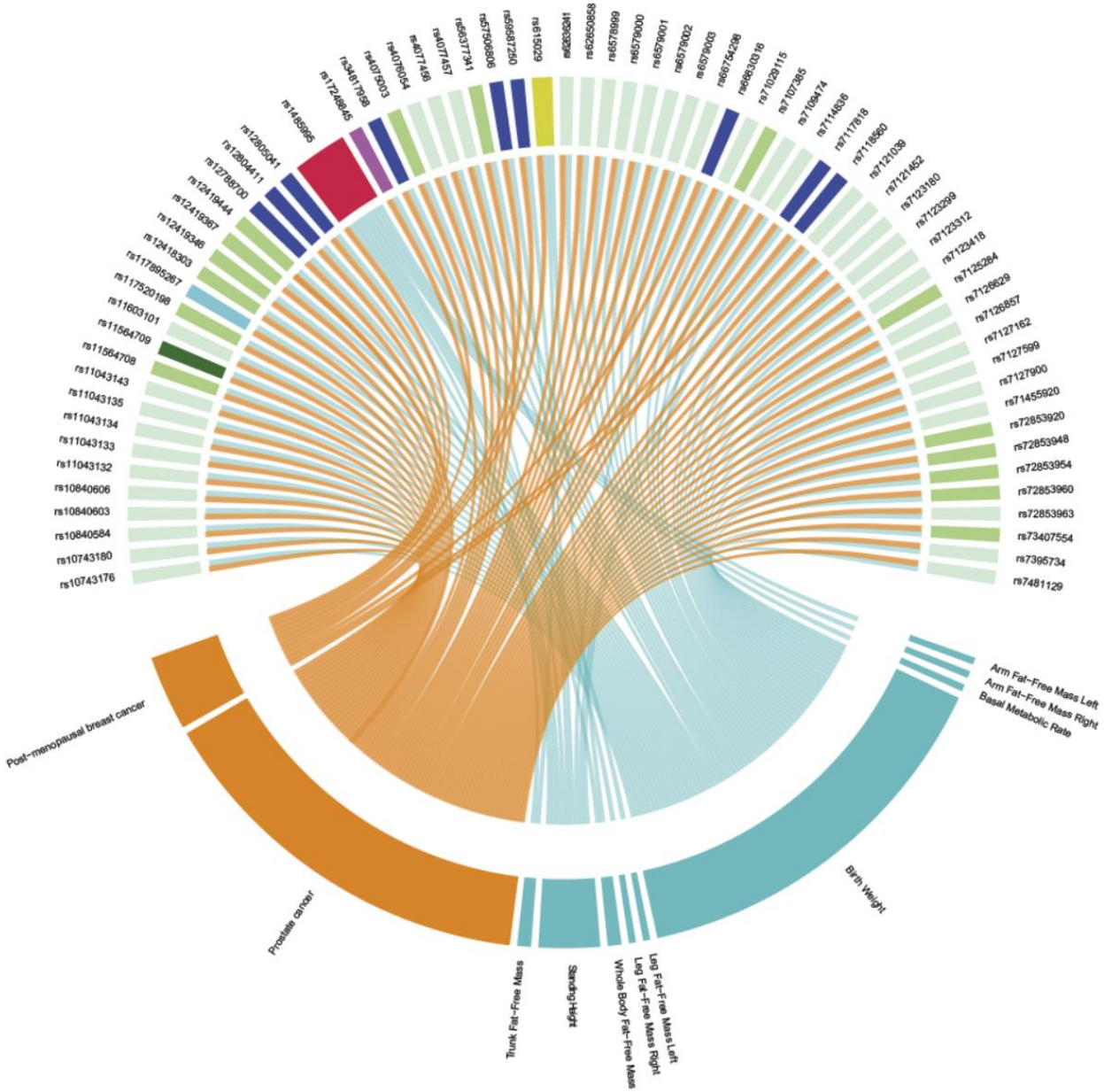


Figure 4.8: *Pleiotropic SNPs associated positively with cancer risk and anthropometric traits*. Circos plot, in which SNPs are located on the top half of the plot whilst cancer types and anthropometric traits are on the bottom half. Significant associations (Bonferroni corrected p values <5.05E−8) are shown as solid lines. Cyan lines indicate associations between SNPs and anthropometric traits. Orange lines depict associations between SNPs and differential cancer risk. Different colours in the outer ring denote diffferent haplotypes.

4.4 Discussion

Above, different associations of BMI, WBFFM and WBFM with three cancer types within the UK Biobank are described, using comprehensive multivariate models (*Chapter 2)*. This chapter shows pleiotropy exists on a genome-wide scale, with individual SNPs being associated simultaneously with differential cancer risk and anthropometric traits in the same large cohort. Indeed, the associations of some haplotypes, and SNPs within them, mirror trends seen in other epidemiological studies. However, this study is the first of its kind to look systematically at BMI, WBFFM and WBFM in three different cancers in the same cohort, addressing a gap in the literature, particularly within breast cancer[543]. To recap salient points for *Chapter 2*, the upper two quintiles of WBFFM carried higher risk for post-menopausal breast cancer compared to WBFM, or even BMI. In contrast, the prostate cancer analysis shows that individuals in the highest quintile of BMI were at lower risk of prostate cancer. From the WBFM and WBFFM analysis, this protective effect appears to be originating from WBFM. In colorectal cancer, the risk in women showed no clear relationship with WBFM or WBFFM, whereas in men, the highest quintiles of WBFM and WBFFM both appear to carry increased risk.

Various ad hoc explanations have been proposed to explain the increased cancer risk with fat mass (WBFM), as shown here for post-menopausal breast cancer and colorectal cancer in men (*Chapter 1*). These include adipocyte hypertrophy and death leading to systemic chronic, subclinical inflammation of adipose tissue[333]. Leptin and oestrogen from adipocytes and increased circulating insulin caused by insulin resistance may also contribute[16,18,22]. Further mechanistic discussions on how cancer and obesity are linked are comprehensively reviewed in[31,815]. The link identified here between WBFFM and post menopausal breast cancer has also been described previously, but in smaller cohorts[495,541].

Its mechanism is unknown, but insulin-like growth factor could well be involved as it has its anabolic effects and has been associated with cancer by other workers[548].

In the current chapter, other techniques are employed in the search for some explanation as to how genetic pleiotropy might be associated with both anthropometric traits (notably WBFM and WBFFM) and cancer. GWAS searching for pleiotropic haplotypes (co-inherited SNPs) associating with anthropometric traits and cancer risk revealed a number of haplotypes associated with both phenotypes in both prostate and post-menopausal breast cancer. The discovery of five haplotypes positively associated with both prostate cancer and increased birth weight may help to explain similar epidemiological observations in other cohorts[816]. Indeed, two SNPs within one of the pleiotropic haplotypes, rs7126629[C] and rs7127900[A], have been associated with prostate cancer in a number of different studies on independent cohorts[787–789,799].

4.4.1 CUPID hits the spot

SNPs within one haplotype associated with both post-menopausal breast cancer (rs615029G>T, p = 1.56E^{-17},OR 1.19) and WBFFM (rs615029G>T, p = 2.88E^{-08}, beta = 0.010). One SNP within this haplotype, rs537626[C], has been previously found to associate with breast cancer in a different cohort[642]. This polymorphism is within a non-coding transcript exon variant of the long non-coding RNA (lncRNA), LINC01488/CUPID1. LncRNAs are present in low copy numbers (hence the requirement for sensitive RNA sequencing) and do not code for polypeptides, although they are extensively processed by splicing (considerably more than mRNA itself) and 3'polyadenylated[817]. Indeed, nearly every non-coding exon is subject to alternative splicing[818]. rs537626[C], and other SNPs in this haplotype, may well alter the function of this lncRNA.

CUPID1 has been shown to be induced after estrogen stimulation and is believed to affect pathway choice for DNA repair[819]. Silencing of CUPID1 inhibited RAD51 recruitment to double-strand breaks, and appears to decrease high-fidelity homologous recombination through a defect in end resection. This results in the more mutagenic non-homologous end-joining (NHEJ) becoming the dominant DNA-repair pathway[819,820]. Homologous recombination might also be reduced due to effects on the cell cycle. CUPID1 is transcribed from DNA within the promoter region of cyclin D, acting as a brake on cyclin D production, via effects on histone acetylation. This slows cell proliferation, particularly in the S phase where homologous recombination is favoured. In its absence, the cells rapidly progress through S phase, and homologous recombination repair is limited, leading to NHEJ becoming the preferred repair pathway (Fig. 4.9). This SNP may thus influence breast cancer risk through the creation of new, and persistence of existing, mutations. CUPID1 also decreases stability of cyclin E, decreasing the production of pro-metastatic agents such as vimentin (via increasing the level of inhibitor miRNA species miR124-3p and miR138-5p) and MMP9[821]. The link between rs537626[C] and increased fat-free mass is less clear as we were unable to confirm alterations in gene expression through interrogation of eQTL databases.

Many of the SNPs identified above, and their associations, mirror the trends in our and others' epidemiological studies, e.g. one haplotype is associated with both an increase in WBFFM and an increase in post-menopausal breast cancer. This perhaps is unsurprising as pathways that control cell growth are involved in both anthropometric traits and cancer. However, a number of SNPs were identified that associate with post-menopausal breast cancer/prostate cancer and anthropometric traits that run counter to the observations in epidemiological studies (supplementary tables S4.13-S4.14). It is possible that inherited factors may not be the overriding influences on these phenotypes, with environmental factors

being dominant in some cases. Of note, we were unable to identify SNPs in colorectal cancer that show associations with both the cancer and anthropometric traits.

4.5 Conclusion

This novel study combines epidemiological and genome-wide approaches, and highlights SNPs that have pleiotropic associations with cancer risk and anthropometric traits. Previous work in this thesis demonstrated this phenomenon in SNPs that are in eQTL related to key cancer genes[99], and this chapter shows this anthropo-oncologic link in a much wider analysis of the entire genome. The paucity of SNPs influencing fat mass is noteworthy, and emphasises the importance of dietary and other lifestyle factors in modifying cancer risk in both colorectal and post-menopausal breast cancer. The unexpected finding that those in the highest quintile for BMI have a reduced risk of prostate cancer, and with other similar trends being shown in other analyses, does raise some challenging public health decisions. Interestingly, obesity (defined through BMI) has been associated with more aggressive prostate cancer by other workers, who attribute it to the development of a low testosterone/high estradiol environment[822].

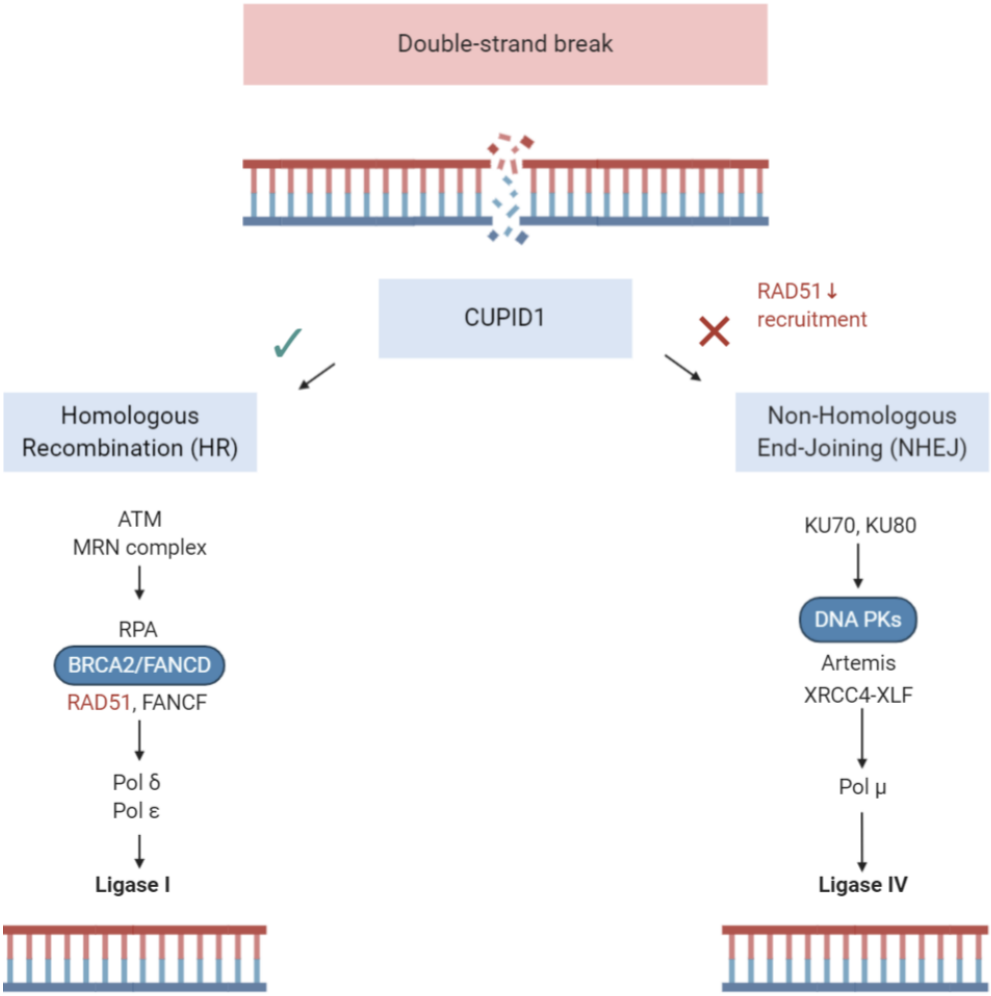


Figure 4.9: *Proposed mechanism by which the LncRNA CUPID1 may influence double strand break repair.*

CUPID1 promotes repair of double strand breaks by homologous recombination (high fidelity) by promoting Rad51 recruitment. rs615029G>T might inhibit the function of CUPID1 and thus favour a switch to error prone non-homologous end-joining (NHEJ).

Chapter 5: An unexpected ending - *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)*

Preface

During my DPhil, the SARS-CoV-2 (SCoV2) pandemic disrupted all aspects of our lives. My research for this thesis was no exception as I attempted to help as a clinician and as an academic (as mentioned in my *SARS-CoV-2 Pandemic Impact Statement*). This chapter describes work I carried out during the global pandemic response in several important areas. We were among the first researchers to describe longitudinal immunological responses to SCoV2 infections in two important patient groups. Further, as cancer and obesity both emerged as important risk factors for SCoV2 outcome, we investigated whether SNPs related to SCoV2 outcome are also associated with pleiotropic effects in these clinical phenotypes.

Chapter Summary

Background: Accurate and sensitive detection of antibodies to SARS-CoV-2 is an essential tool in combating the pandemic. Here we describe the prevalence and evolution of the presence of nasopharyngeal SARS-CoV-2 and serological antibody responses in two important groups of individuals, asymptomatic patients and those with Post-COVID-19 Syndrome (PCS). Since cancer and obesity both emerge as important risk factors for SCoV2 outcome, it was also investigated whether SNPs related to SCoV2 outcome associate with concomitant pleiotropic effects on these two phenotypes.

Methods: Patients and staff were recruited at a hospital in London hospital. To identify SARS-CoV-2 infection, RNA from nasopharyngeal swabs was measured by RT-PCR (Micropathology Ltd, Warwick). Antibody testing for SARS-CoV-2 was undertaken using the 'in-house' double-binding antigen ELISA (Imperial Hybrid DABA; Imperial College London, London, UK).

Results: (a) At the start of the UK SARS-CoV-2 epidemic, the study investigated the frequency of asymptomatic infection. In this rare point prevalence study, it was observed that one third of patients in a London inpatient rehabilitation unit were infected with SARS-CoV-2, but were asymptomatic (n=15/45). Eight of these patients were followed in time, using a novel serological test for antibodies against the receptor binding domain of SARS-CoV-2. The results showed that all of this asymptomatic cohort remained seropositive between 3 to 6 weeks post-diagnosis. (b) The occurrence of post covid syndrome (PCS) in a group of 38 healthcare workers, who had become infected by SARS-CoV-2 while employed at this hospital was also investigated. Over half (n=21/38) of these were diagnosed as suffering from PCS at 7-8 months after the initial infection. No statistically significant difference in was detected in initial RT-PCR viral titres, nor in serial antibody levels taken over 8 months,

between those who did and those who did not develop PCS. No published SNPs were associated with poor SARS-CoV-2 outcome, together with BMI, WBFM, post-menopausal breast, prostate and colorectal cancer risk in the UK Biobank. However, two haplotypes associated with poor SARS-CoV-2 outcome associated with WBFFM (rs1265097[A], p: $3.73E^{-13}$, beta = -0.017 & rs3130279[A], p: $2.30E^{-10}$, beta = 0.012).

Conclusion: (a) This work suggests that 3 to 6 weeks after a positive PCR test is a suitable timeframe to check for a serological response in asymptomatic patients, and that the antibody response is maintained over time. (b) The incidence of PCS was not simply related to either (i) viral load after contracting SARS-CoV-2, or (ii) the level of antibody response during or after viral infection. Indeed, there appears to be a high incidence of PCS in healthcare workers and this should be considered in vaccination scheduling. (iii) Although pleiotropic associations do not currently appear to tie together the link between COVID-19 outcome and obesity or cancer, mining the GWAS Catalog highlights an important clinical point, which suggests that clinicians should carry out early computerised tomography pulmonary angiograms (CTPAs) in patients with COVID-19 requiring oxygen to avoid progression to respiratory failure.

5.1 Introduction

The disease caused by SARS-CoV-2 (SCoV2), 'COVID-19' is complex, displaying variable clinical symptoms and disease severity. On this spectrum, most patients develop mild symptoms. Predominant features of the disease include a dry cough, fever, sore throat and loss of taste and smell[823]. In some cases, patients develop serious complications, including septic shock, severe pneumonia, renal failure, and atypical acute respiratory distress syndrome[824]. Age, and co-morbidities (including cancer and obesity) have all been associated with poorer prognosis, as have sex (males seem more susceptible) and ethnicity (Black, Asian and minority ethnic [BAME] seem more susceptible)[823].

Asymptomatic carriers are well-documented and have been implicated in disease transmission. On the Diamond Princess cruise ship, 634 of the 3711 passengers were diagnosed with SCoV2 infection[825] of which 18% were asymptomatic. In other examples, the proportion of those asymptomatically infected ranges between 40-45%[825,826].

Current testing strategies revolve around the use of reverse transcriptase polymerase chain reaction (RT-PCR) tests from naso-pharyngeal swabs. However, in the absence of viral cultures, RT-PCR results can be difficult to relate to infectiousness[826]. Furthermore, the ability to detect past infection or host-immunity by this route is not well defined[827]. If undertaken within the correct time frame, serological testing can potentially detect both active and past SCoV2 infection[828]. Consequently, the use of serological data has controversially been mooted as "a way out of lockdown" by offering the possibility of "immunity passports"[7]. Trials are also underway to establish whether the plasma from convalescent individuals with high levels of neutralising antibodies may help treat seriously ill patients[829].

SCoV2 was initially described in Wuhan, China in December 2019[830]. At the beginning of March 2020, the number of confirmed cases of SCoV2 in London had risen from 25 to over 11,000[831]. During the third week of March 2020, managers of a rehabilitation facility in London

arranged for all inpatients to be tested for SCoV2. At this time most hospitals in the UK were testing only symptomatic patients, making this study unique. This approach permitted the characterisation of the clinical and immune response in a group of asymptomatic patients, and could give a rough estimate for the burden of asymptomatic carriers within London hospitals at that time. In addition, prospective haematology, biochemical and serological measures over a four-week period (week 3 to week 6 after infection was diagnosed) in a group of these asymptomatic SCoV2 infected individuals were gathered.

Later, as the pandemic continued to spread and we saw the majority of patients surviving the initial infection, it became clear that a cohort of patients who have been infected with SCoV2 have ongoing persistent symptoms after the acute phase of the illness is over. These longer term effects after the initial infection have been labelled Post-COVID-19 Syndrome (PCS), colloquially known as Long-COVID[832]. Indeed, the working definition proposed by the National Institute for Health and Care Excellence (NICE) is '*Signs and symptoms that develop during or following an infection consistent with COVID-19, continue for more than 12 weeks and are not explained by an alternative diagnosis*'[833]. Symptoms include shortness of breath, muscle aches, loss of sense of smell, tiredness and headache. More than five symptoms within a week of diagnosis appears to increase the risk of PCS[834]

Understandably, until now, most of the currently published literature has focussed on epidemiological modelling, clinical complications and treatment strategies for the acute phase of COVID-19 infection. However, as data from various patient cohorts mature, more information is likely to emerge surrounding the phenomenon of PCS. With the global spread of the SCoV2 virus, PCS is likely to have significant personal and economic impacts. Therefore, with a third wave threatening and vaccines coming into use, it is important to recognise risk factors for developing PCS, to further stratify the healthcare workforce and plan vaccination schedules.

It is possible antibody response and/or initial viral load may affect the risk of developing PCS. This work details a cohort of 42 healthcare workers from one hospital in London to investigate if any relationship exists in this cohort between PCS and (i) diagnostic viral Ct values (surrogate for viral load) and (ii) serial antibody levels.

This work utilises a novel approach to antibody detection, different to the tests approved by the UK government and widely used in clinical settings at the time. Conventional antibody immunoassays are based on an indirect format, whereby antibody binding to the antigen is revealed by a labelled antibody to human immunoglobulin. The antibody tests approved at the start of the epidemic, including the Abbott and the Roche platforms, were designed using this approach and based on detection of antibody to the nucleoprotein antigen. This test is relatively simple to manufacture, and is adequate to provide evidence of prior infection if positive. However, it fails to identify *per se* the presence of antibodies likely to confer immunity against reinfection.

The team at Imperial College London have developed an alternative assay, using their previous experience from the Ebola outbreak in West Africa[835]. This detected antibodies against the external components of the virus, which are likely to be to more predictive of neutralising antibodies and hence may indicate protection above the Cytotoxic T lymphocyte response from the primary infection[836]. This assay focussed on the quantification of antibody against the receptor binding domain (anti-RBD) of SCoV2. Below we detail antibody response over time in the context of asymptomatic patients and in Post-COVID-19 Syndrome.

5.2 Methods

Following ethical approval from the rehabilitation hospital, all samples were donated with informed written consent to the local University Communicable Disease Research Tissue Bank (NRES SC/20/0226). The use of the samples for research was approved by the Tissue Bank Steering Committee in accordance with NRES procedures.

Sample populations

5.2.1 Asymptomatic patients

All participants were inpatients in a specialist adult neurological rehabilitation hospital in north-west London. In this facility, each patient is managed in their own room. Between 21/03/20 and 31/03/20, all 45 patients underwent nasopharyngeal swabbing by trained nursing staff. Patients found to be positive were transferred to an isolation ward. As part of routine management, these patients were closely monitored for signs and symptoms of COVID-19 and had weekly blood tests. Additional swabs were taken at least weekly from including at least the 3rd week post-diagnosis swabs. Patients remained in the isolation ward until they had two consecutive RT-PCR negative swabs. Serum for anti-SCoV2 antibody were first taken 3 weeks (18-24 days) after first positive PCR and then weekly for 4 weeks (weeks 3,4,5,6 diagnosis).

5.2.2 Healthcare workers

Healthcare workers, infected with SCoV2 at the same north-west London hospital, as the asymptomatic patients were invited to take part in the study. Invitations were sent by email and details including date of birth, sex, ethnicity, medical history and occupation were recorded. An online questionnaire asked all participants about PCS symptoms between 7-8 months from Covid-19 symptom onset or date of diagnosis in the asymptomatic group. Samples were taken for SCoV2 antibody testing at enrollment, then every week for four weeks and then every month. Both the questionnaire and sample collection after month one were coordinated by Christopher Pereira.

5.2.3 RT-PCR

To identify SCov2 infection, RNA from nasopharyngeal swabs was extracted within 24 h. Diagnostic testing for SCov2 was performed by Micropathology Ltd (Warwick). Nucleic acid was extracted using the Maxwell® HT 96 NA extraction kit (Promega Corp) using the KingFisher FLEX platform (Thermo Fisher Scientific Inc) RT-PCR primers testing were as recommended by the US Center for Disease Control (https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html). RT-PCR was performed on a Roche LightCycler 480 on the following cycle: reverse transcription (50°C, 10min), polymerase activation (95°C, 2min) followed by PCR (45 cycles 95°C for 5sec, 55°C for 20sec). Samples with a Ct value of >40 and no history of positive detection were retested. DNA from baculovirus Adoxophyes orana granulovirus (AoGV) was used as an internal control. Any samples failing to amplify (or showing late/weak amplification) underwent re-extraction and the PCR repeated. Patients with two consecutive negative RT-PCR results were defined as SCoV2 negative.

5.2.4 Detection of anti- SARS-CoV-2 antibodies

Serological screening for SCoV2 antibody was carried out at Imperial College. Antibody to the SCoV2 RBD was detected using a two-step sequential 'in house' double antigen binding assay employing S1 antigen on the solid phase and enzyme-labelled RBD in the fluid phase. Specificity was 100% (95%CI=99.6-100%) (defined by testing 825 sera that pre-dated the epidemic), and sensitivity was 98.91% (96.8-99.8%) (using 276 sera from RT-PCR confirmed individuals)[837]. This DABA approach detects total immunoglobulins against SCoV2, avoiding problems associated with pure IgM or IgG assays. This is important since the timecourse for seroconversion from IgM to IgG in SCoV2 infection is still under debate[838]. Assay cut-off was calculated from Receiver Operating Characteristic Curve analysis, and serum reactivity normalised by using the signal to cut-off (S/CO) ratio, the ratio of optical density (OD) values generated in a sample to the cut-off OD value. A sample was considered to be antibody positive if the S/CO was greater than 1.

5.2.5 Statistical analysis

The two-sided Mann-Whitney-Wilcoxon test (paired or unpaired) was used as appropriate, as was Fisher's exact test (two-tailed). A Bonferroni correction was used to correct for the possibility of type 1 error for extra stringency. Spearman's rank correlation coefficient was used to correlate Ct values and anti-RBD titres.

5.2.6 Pleiotropic association analysis

SNPs associated with SARS-CoV-2 outcome were identified from the literature search of the GWAS catalog, Pubmed, EMBASE [(GWAS) OR (SNP)) OR (Single nucleotide polymorphism)) AND (SARS-CoV-2)) OR (COVID)) OR (COVID-19)] – All fields, Date:

15/1/21). Haplotypes were generated from the UK10K[615]. SNPs were then explored in our GWAS in-house work on the UK Biobank (*Chapter 4*) and in the GWAS catalog.

5.3 Results

5.3.1 Asymptomatic patients: evolution of the presence of nasopharyngeal SARS-CoV-2 and serological antibody responses

SCoV2 RNA was detected in 15/45 patients. Twelve consented to participate: of these four were discharged during the study period and were lost to follow-up. Data are presented on the eight patients who were followed for six weeks (demographic data found in supplementary table S5.1). There were five males and three females, mean age 51 years. Regarding co-morbidities, four of the eight had suffered a stroke and three had undergone neurosurgery. Two had type II diabetes, one of whom also suffered from hypertension, Chronic Obstructive Pulmonary Disease, and chronic kidney disease. All reported either no symptoms of SCoV2 infection or, on questioning, very mild symptoms not requiring medical intervention. Five reported a mild fever, but on repeated tympanic measurements over the study period, no patient had a temperature ≥38°C. Two reported a mild cough lasting less than 24 hours.

Biochemical and haematological markers were evaluated at baseline and throughout the two-month period. The only abnormalities detected were that two patients had a mildly reduced lymphocyte count which returned to normal by the third week after diagnosis. From biochemistry, one patient had mildly deranged liver function tests, attributed to a concurrent course of antibiotics for an unrelated bacterial infection.

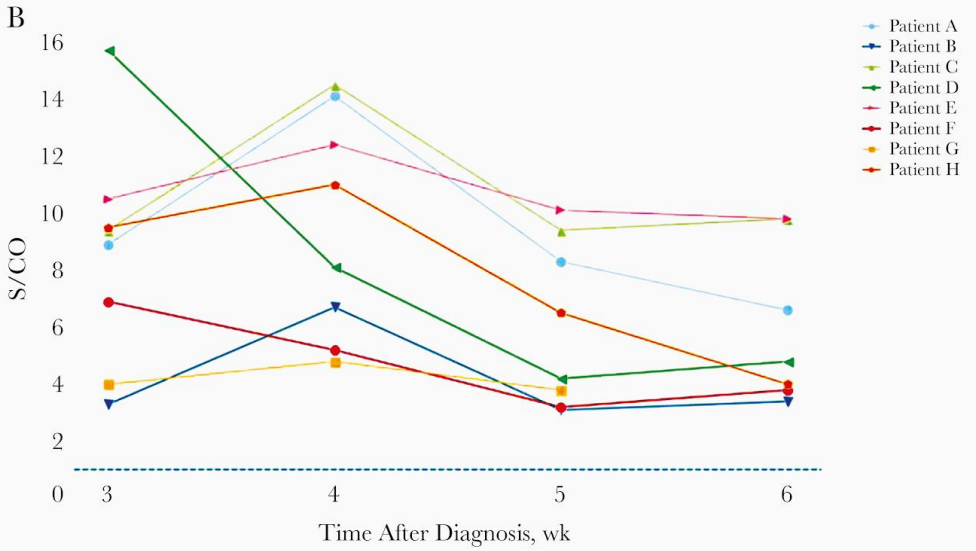
In this point prevalence study, the median cycle at which signal amplification is detected above the threshold, Ct, at initial sampling of these asymptomatic patients was 20.4

(IQR 19.4–22.5). Ct increased significantly at the second testing point (two weeks later) to median 34.7, (IQR 32.8-37) ($p < 0.001$) representing about a 10,000-fold reduction in SCoV2 viral load in the nasopharyngeal swabs. There was a further 10 fold reduction over the next 7 days (median Ct 37.8, IQR 36.2–38.4) with SCoV2 RNA still detected in 7/8 patients. Thereafter SCoV2 was not detected. Participants remained swab-positive for an average of 32 days (range: 15 -46) before two sequential negative swabs. Individual results are presented in Fig. 5.1.

Antibodies to RBD were present in all eight patients at the first serum sample (3 weeks after SCoV2 RNA was first detected). Fig. 5.1 displays individual antibody levels for the eight patients. Antibody levels seem to peak after 4 weeks post diagnosis, and then fall back, but are still present after 6 weeks. There were statistically significant drops in total Ig levels between weeks 3 and 5 post-diagnosis ($p = 0.01$), weeks 4 and 5 post-diagnosis ($p = 0.01$) and week 4 and 6 ($p = 0.049$). The changes between weeks 3 and 4 post-diagnosis and weeks 4 and 5 post-diagnosis did not survive stringent Bonferroni correction. Spearman's rank correlation of Ct values at Week 1 vs anti-RBD serum levels at Week 3, yielded a negative correlation very close to statistical significance ($r = -0.69$, $p = 0.06$). Other correlations between Ct values in subsequent weeks were in the negative direction but were further from statistical significance.

A

	Ct Value for SARS-CoV-2 Assay					
	wk 1	wk 2	wk 3	wk 4	wk 5	wk 6
Patient A	13.84	NA	32.43	35.46	34.19	Negative
Patient B	25.38	NA	40.34	32.63	40.11	Negative
Patient C	21.61	NA	35.99	38.06	36.56	Negative
Patient D	16.9	NA	33.9	37.05	37.73	Negative
Patient E	20.19	NA	32.99	38.71	37.51	Negative
Patient F	20.46	NA	27.39	37.78	37.08	Negative
Patient G	25.82	NA	40.03	Negative	Negative	
Patient H	20.37	NA	35.58	40.13	Negative	Negative



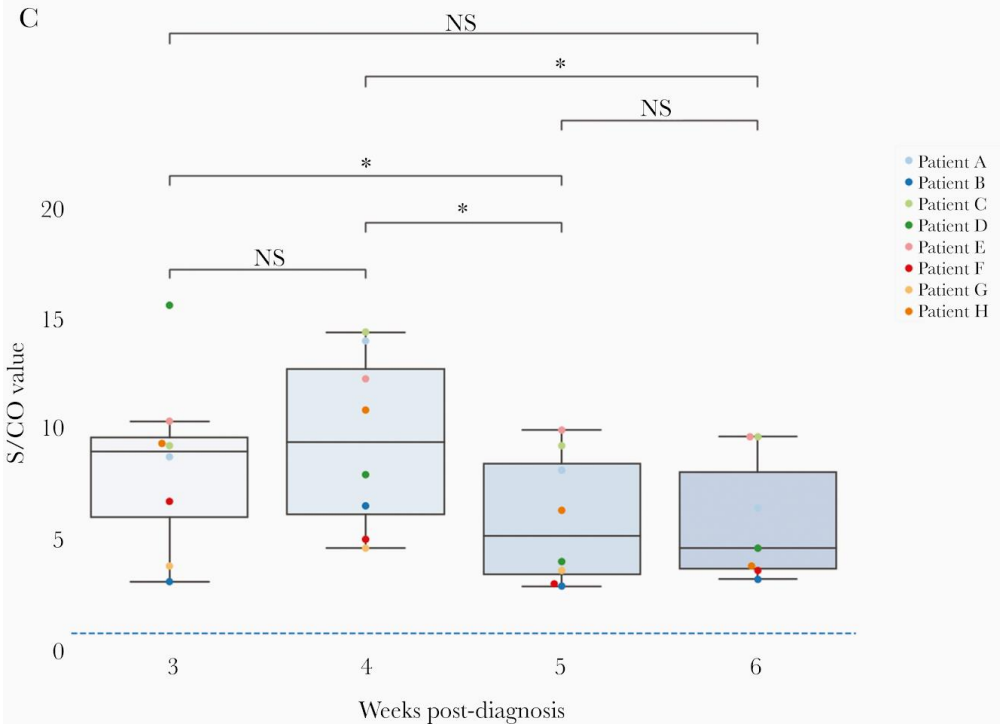


Figure 5.1: *Time course of SARS CoV-2 infection and immune response in asymptomatic patients* a) Level of SARS-CoV2 RNA during infection, measured by RT-PCR on nasopharyngeal swabs as described in Methods. C_{t-} count results for viral RNA over the study period as shown. Values defined as positive for SCoV2 are highlighted red, while green highlights indicate control RNA values (no detectable viral load); b) Serum anti-RBD was measured by the DABA method giving a signal cut-off ratio [S/CO] as described in methods, over a four-week period (weeks 3-6 post-PCR diagnosis); c) Data from (b) shown as box plots at individual time points. Brackets represent significance on a Mann-Whitney test (ns = non-significant, *: $0.01 < p \leq 0.05$).

5.3.2 Nasopharyngeal SARS-CoV-2 levels and serological antibody responses in healthcare workers: relation to the development of Post-COVID-19 syndrome

42 healthcare workers enrolled into this study (female = 36, male = 6) having been diagnosed between 21/3/20-5/5/20. The majority of workers had been symptomatic with the SCoV2 infection but did not require hospitalisation (n=39), and the remainder were asymptomatic but picked up on routine screening. The mean age of the cohort was 43 years (range = 23-67). Further information is found in supplementary table S5.2.

Of the healthcare workers who responded to the PCS questionnaire, 54% (21/38) had signs and symptoms of PCS. The commonest were fatigue (57%), anosmia (29%), and dyspnoea (24%). A third of these individuals had ≥3 ongoing symptoms. This was much higher than the level of PCS in the general populace following SCoV2 infection at >12 weeks post-diagnosis (2.3%)[834].

It is known that Black, Asian and minority ethnic (BAME) individuals are at higher risk of suffering from COVID-19 and 71% of BAME individuals in the study met the definition of PCS at 7-8 months. However, perhaps due to the small sample sizes, this difference was not statistically significant compared to the non-BAME group (Fisher's exact test, p = 0.181). 63% of women developed PCS compared to 17% of men, but this too was not quite statistically significant (Fisher's exact test, p = 0.07).

89% of healthcare workers who gave samples over six months had raised antibody levels (S/CO≥1) at the end of this period (six-months post-symptom onset or RT-PCR diagnosis in the asymptomatic individuals, Fig. 5.2). Interestingly, one individual never developed antibodies and clinical data collected did not yield an obvious reason for this.

SCoV2 viral load at diagnosis, indicated by Ct values from nasal swabs, was not statistically significantly between the PCS and NPCS. There was also no difference in anti-RBD between the PCS and NPCS groups at any time of testing (months 2 to 8 post-symptom onset/diagnosis).

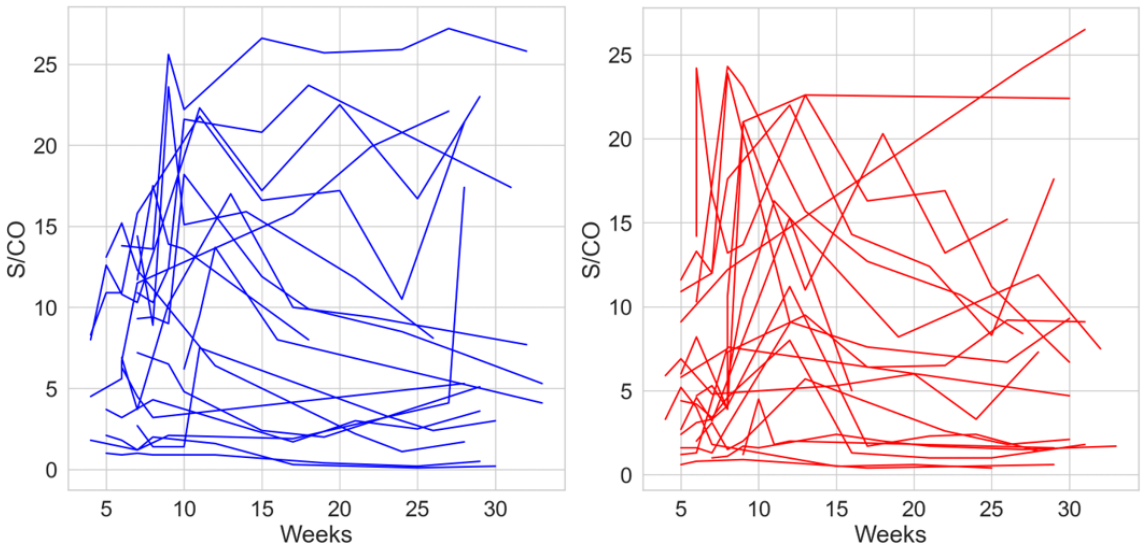


Figure 5.2: *Variation of antibody levels with time in the PCS and NPCS groups*

The right hand pane (red lines) represents individual healthcare workers with PCS and the left hand pane (blue lines) those who did not develop PCS (NPCS). Serum anti-RBD was measured using the signal cut-off ratio [S/CO] as described in 5.1 and methods. Horizontal scale shows weeks following the earlier of post-symptom onset or PCR diagnosis.

5.3.3 SNPs, SARS-CoV-2 and genetic pleiotropy

On the basis of previous work for this thesis, we were curious to investigate the association between COVID-19 and obesity, which has been described as a risk factor for severe disease[839]. As more data becomes available, GWAS becomes an attractive option to try to untangle genetic underpinnings of the severity of SCoV2 infection. Therefore, as in obesity and cancer, I was interested to investigate if SNPs associated with obesity are also associated with severe SCoV2 infection.

Therefore, the major databases (GWAS catalog, Pubmed, EMBASE) were interrogated to identify GWAS studies on COVID-19. This identified 10 SNPs associated with Severe SCoV2 infection (table 5.3), although this number is likely to grow. SNPs and their haplotypes were then investigated for associations with traits the GWAS studies in the UK Biobank (GWAS analyses described in *Chapter 4*). No published SNPs associated with poor SARS-CoV-2 outcome together with BMI, WBFM, post-menopausal breast, prostate or colorectal cancer risk in the UK Biobank. However, two haplotypes were associated with WBFFM (one containing rs3131294[G], lowest p associated to rs3130279[A], p: $2.30E^{-10}$, beta = 0.012 and one containing rs143334143[A], lowest p associated to rs1265097[A], p: $3.73E^{-13}$, beta = -0.017).

SNP	p	OR	Reference
rs11385942[GA]	1.15×10^{-10}	1.77(1.48-2.11)	[840]
rs657152[A]	4.95×10^{-8}	1.32 (1.20-1.47)	[840]
rs73064425[T]	4.8×10^{-30}	2.1 (1.88-2.45)	[841]
rs9380142[A]	3.2×10^{-8}	1.3 (1.18-1.43)	[841]
rs143334143[A]	**8.8×10^{-18}**	**1.9 (1.61-2.13)**	[841]
rs3131294[G]	**2.8×10^{-8}**	**1.5 (1.28-1.66)**	[841]
rs10735079[A]	1.6×10^{-8}	1.3 (1.18-1.42)	[841]
rs2109069[A]	4×10^{-12}	1.4 (1.25-1.48)	[841]
rs74956615[A]	2.3×10^{-8}	1.6 (1.35-1.87)	[841]
rs2236757[A]	5×10^{-8}	1.3 (1.17-1.41)	[841]

Table 5.1: *SNPs associated with severe SARS CoV-2 infection*

Database searches (GWAS catalog, PubMed, EMBASE) were carried out as previously to identify SNPs which associated with severe *SARS CoV-2.* Ten such SNPs were identified. These were investigated to identify further possible associations with anthropometric traits (BMI, WBFM, WBFFM) and risk of cancer (post-menopausal breast, prostate, colorectal cancers) using the techniques

described previously. SNPs within haplotypes with associations with tested traits in the UK Biobank are emboldened.

5.4 Discussion

5.4.1 Asymptomatic SARS-CoV-2 patients & individuals with Post-COVID-19 Syndrome

Evolution of serological and biochemical changes is currently not well defined in asymptomatic SCoV2 patients. In March 2020, at the beginning of the SCoV2 outbreak in London, testing all patients in one institution for SCoV2 RNA was unique as, at the time, guidance was to test only symptomatic patients. Therefore, healthcare planners did not know whether there was infectious virus shedding amongst asymptomatic patients.

This work describes a rare point prevalence study of asymptomatic patients with longitudinal antibody and RT-PCR follow up. This work shows that 15/45 (33%) of rehabilitation patients in one hospital were infected when the number of symptomatic SCoV2 positive patients in London was 11,000 (around 0.01% of London population)[842]. This implies that the numbers who had contracted the virus was far higher in the immediate post-lockdown period than data from early in the UK pandemic suggested.

Our work also shows the time course of anti-RBD induction in these individuals. The only comparable published study is of an asymptomatic cohort of 37 patients from Wuhan, China[843]. Using a chemiluminescence enzyme immunoassay for IgG, these workers found that 40.0% (12/30 of asymptomatic individuals) became seronegative at 8 weeks post-PCR diagnosis[837]. This is in contrast to the results given here where, using a DABA immunoassay, it appears the antibody response is more long-lived. This study provides the first longitudinal estimates of serum total immunoglobulin levels from weeks 3-6 post-diagnosis. All eight asymptomatic patients developed an antibody response within three weeks of first PCR

positive result and this persisted throughout the four-week follow up period. Although antibody levels fluctuated during the study period, 100% of the cohort remained seropositive between 3 to 6 weeks post-PCR diagnosis for anti-RBD. This work suggests that 3 to 6 weeks post-PCR positivity is a suitable timeframe to check for a serological response in asymptomatic patients for confirmation of infection.

In a different cohort of healthcare workers rather than patients, it appears that most individuals infected with SCoV2 retain anti-RBD antibodies over a 6 month period, in agreement with a recently published paper using a similar approach[844]. This work also highlights the high rate of PCS amongst healthcare workers. 55% of healthcare workers in the study had PCS at 7-8 months, and women seemed particularly affected (63% of women in the study developed PCS). However, this cohort at recruitment has a strong female preponderance (36/42).

This persistence of symptoms is much higher than reported in other studies. Sudre et al. reported that symptoms persisted in only 2.3% of the general population at >12 weeks[834]. Since our study cohort consisted of healthcare workers, one might suggest that repeated contact with the virus in clinical areas may play a role in the development of persistent symptoms.

Measurements of viral load and serial anti-RBD measurements up to 8 months from symptom onset (or diagnosis in the asymptomatic individuals) were not statistically different between the PCS and NPCS groups. Although this might seem paradoxical, it may not be surprising as studies into post-viral fatigue, probably the most similar syndrome in the medical literature, have struggled to find consistent correlations with viral load or antibody response[845].

However, this work does highlight an unexpectedly high prevalence of PCS in healthcare workers at 7-8 months, and perhaps a female preponderance. This may affect the healthcare workforce long after the pandemic has passed. And, with a vaccine on the horizon,

but still cases rising, these results should be noted by public health officials involved in planning vaccination programmes in order to minimise detrimental effects on the healthcare workforce.

5.4.2 SNPs, SARS-CoV-2 and genetic pleiotropy

Our investigation into pleiotropic associations of SNPs associated with poor SARS-CoV-2 outcomes in the UK Biobank yielded no SNPs that could easily help explain links between obesity and SARS-CoV-2 outcomes or cancer and SARS-CoV-2. However, mining the GWAS Catalog for other associations did yield some interesting results. Respiratory Failure with SCoV2 infection particularly associates with rs657152[A], an intronic variant in the ABO locus[840]. This SNP had been previously found to be associated with increased fasting circulating IL-6 ($p=2.0E^{-08}$, beta = 0.041), which is raised in obesity (extensively discussed in *Chapter 1*)[846]. Both elevated IL-6 and obesity are associated with deep vein thrombosis[847–849]. rs657152[A] was also found to be associated with an increased risk of deep-vein thrombosis and pulmonary embolism in populations without COVID-19[850].

Carriers of the linked SNP rs505922[C] have higher circulating levels of soluble lectin CD209 which may facilitate the spread of SARS-CoV-2 to T lymphocytes[851]. It has been suggested that this might be related to the accompanying severe infection and respiratory failure[850]. Care must be taken with this conclusion however. Carriers of this variant may be intrinsically more prone to pulmonary emboli without SCoV2 infection, while this infection plays on this pro-coagulable propensity, instigating pulmonary emboli and increased need for supplemental oxygen and ultimately respiratory failure. Although soluble lectin CD209 might play a direct role in promoting an overwhelming infection, it is possible undiscovered pulmonary emboli may play a more major role in patients' increasing oxygen demands. Although these workers associated rs657152[A] with respiratory failure in COVID-19[840], there is no mention in the study whether participants with this variant had pulmonary emboli or

indeed had undergone CT pulmonary angiograms to exclude pulmonary emboli as a cause for respiratory failure, which would be very important in identifying appropriate treatments e.g. choosing between anticoagulants and antibiotics. Perhaps genotyping for this variant on admission might indicate the need for more intensive anticoagulation to avoid respiratory failure.

This analysis highlights the strength of GWAS in generating hypotheses in mechanisms behind conditions and their severity. This work highlights a key practical point, to urge clinicians to carry out CT pulmonary angiograms to exclude a pulmonary embolism in patients with COVID-19 requiring oxygen, to avoid progression to respiratory failure. It is still early in this field with SCoV2 datasets only just maturing; therefore it will be interesting to see if well-known SNPs associated with obesity do in due course turn out to be pleiotropically associated with severe COVID-19. It would be useful to repeat this type of analysis in early 2022 when further genome-wide studies become available from various cohorts. Also, with COVID-19 outcome data becoming available on the UK biobank participants, there might be an opportunity to explore COVID-19 outcome and anthropometric measures, such as BMI, WBFFM and WBFM, and also cancer risk in the same cohort. This subsequent work may uncover shared pathways that influence susceptibility to cancer, obesity and SARS-CoV-2 and pave the way for future laboratory studies.

Chapter 6: Final Thoughts

Previously, links between anthropometric traits and cancer were derived from observations in model animal systems, and on associations in epidemiological studies. A major conclusion from these studies was that obesity was associated with carcinogenesis, based on correlations between cancer risk and BMI (taken as a surrogate for obesity). Previous work on the links between the physiological state of obesity and tumorigenesis is given in the introduction to this thesis, and summarised in a review paper (*Obesity: the perfect environmental storm for carcinogenesis*) submitted to *Nature Cancer Reviews*.

A particular opportunity that was unavailable to my predecessors was for me to use the largest prospective cohort study to date (UK Biobank) to investigate the origin of the BMI signal associating with cancer across a range of malignancies and attempt to identify common genetic roots. This thesis investigates the links between anthropometric measures (BMI etc) and carcinogenesis, and investigates genetic factors underlying both. This work uncovers novel genomic loci that associate with both traits.

By using techniques more commonly used in bioinformatics applied here to epidemiology, including principal component analysis, we were able to identify a novel association between fat-free mass and post-menopausal breast cancer. The ability to disentangle individual risk factors from the composite parameter of BMI highlights the value of tabulating a variety of measurements (whole body fat mass, fat-free mass etc) in the UK Biobank. It also indicates the care that must be taken in deducing causal relationships from BMI. For example, BMI is frequently interpreted as a measure of adiposity but it most certainly is a composite parameter. Looking at WBFM and WBFFM individually, and as a part of principal component analysis, gives interesting insights particularly as cancer risk carried by these two components differs between tumour types.

We then focussed on inherited genetic variants in tumour suppressor and oncogenes, to investigate their influence on anthropometric traits. We discovered several SNPs that vary

the expression of three important cancer genes, FANCA, MAP3K1 and TP53, each associating with both anthropometric traits and cancer risk in the general population. Of particular interest, we identify a previously unrecognised strong association between the rs78378222[C] SNP in the 3'-UTR of TP53 and increased risk for developing non-melanomatous skin cancer (OR=1.36 [1.31-1.41], adjusted-p=7.62E^{-63}) or brain malignancy (OR=3.12 [2.22-4.37], adjusted p=1.43E^{-12}). Simultaneously this SNP associates with increased standing height (adjusted-p=2.18E-24, beta=0.073±0.007), lean body mass (adjusted-p=8.34E^{-37}, beta=0.073±0.005) and basal metabolic rate (adjusted-p=1.13E^{-31}, beta=0.076±0.006), offering a novel genetic link between these anthropometric traits and cancer risk. Indeed, our results clearly demonstrate that heritable variants in key cancer genes can associate with both differential cancer risk and anthropometric traits across the general population (published in the *Journal of Medical Genetics*).

Having investigated pleiotropy in particular SNPs that alter the expression of known proto-oncogenes and tumour suppressor genes (a tiny fraction of the genome), we expanded our analysis to identify pleiotropic SNPs that influence both phenotypes across the entire genome using GWAS. Particularly, we attempt to identify haplotypes that associate with both cancer risk and anthropometric traits that mirror the trends seen in our epidemiological studies (*Chapter 2*). We uncovered 22 haplotypes that associate with both prostate cancer risk and anthropometric traits, and 18 haplotypes that associate with post-menopausal cancer risk and anthropometric traits. In our studies, no SNPs associated with colorectal cancer and anthropometric traits. Unexpectedly, the majority of pleiotropic associations indicate associations that do not match the epidemiological models. For example rs1558901[T] associates with an decreased risk of post-menopausal breast cancer (adjusted-p = 3.14E^{-09}, OR = 0.91) yet increased BMI (adjusted-p = 1.38E^{-196}, beta = 0.073), WBFM (adjusted-p = 3.82E^{-128}, beta = 0.032) and WBFFM (adjusted-p = 9.14E^{-152}, beta = 0.063). No haplotypes showed a positive association with whole body fat mass and with post

menopausal breast/colorectal/prostate cancers. However, we identified two haplotypes by GWAS as being positively associated with post-menopausal breast cancer risk and whole body fat-free mass. Such an association had been identified as a risk factor in our epidemiological study of the same cohort (*Chapter 2*). Also, this genome-wide analysis showed that the TP53 eSNP, rs78378222[C], discovered in our SNP-wise analysis, and found at 1% frequency in the European population, has the second highest effect size in the entire genome on whole body fat-free mass. Fat-free mass has been relatively neglected in previous studies and my work hopefully will highlight the importance of considering it.

A particularly notable finding from this work relates to the rs78378222[C] TP53 eSNP, because of its more obvious mechanistic link to cancer, prevalence in the population and potential for intervention. Based our findings and others in the literature, implying a potential link with oxidative phosphorylation, the well-tolerated anti-diabetic drug metformin, which is thought to inhibit mitochondrial complex 1, will be trialled in humans carrying the minor allele of the TP53 cancer eSNP (rs78378222) across multiple centres, with Oxford as a major contributor. And, although my foray into SARS-CoV-2 testing, antibody responses and genetics was unexpected, I hope that my work will help in some small way to bring an end to current widespread disruption. The link between obesity and SARS-CoV-2 severity remains intriguing, and worthy of further study.

With a SARS-CoV-2 vaccination programme in full swing and an NIHR academic clinical fellowship organised in Oxford to carry on my work, I hope to test some of my mechanistic hypotheses. For instance, it would be valuable to carry out proliferation assays on myocytes with the SNPs we identified in our GWAS work, notably rs78378222[C] and rs615029[T]. I am in discussions with the oncology teams here in Oxford to investigate the possibility of immunotherapy trials related to obesity, and we are also interested to discuss with the surgical teams potentially simple surgical interventions that could form part of a

longitudinal study to investigate cancer risk in individuals who have had white adipose tissue removed.

The major debates as to the mechanisms by which obesity orchestrates tumour formation at different sites are ongoing. However, my work highlights key areas and omissions in the field that could be targeted by myself or other scientists. It points to shortcomings in the use of BMI, and adds several loci to the literature that need further investigation as to mechanisms that may tie together cancer risk and anthropometric traits.

Additional investigations and achievements

As part of my DPhil, I undertook a number of other linked investigations that I have not included in the above chapters. I investigated if mitochondrial mutations in complex 1 (particularly frameshift or deletions) might influence response to metformin in a cohort of breast cancer patients treated with metformin in a window of opportunity study, as seen in[852]. This influence has been shown in *in vitro* model systems with the similar compound phenformin[582]. Lord et al. identified two groups of patients with distinct metabolic responses to metformin: (i) An OXPHOS transcriptional response (OTR) group for where there was an increase in OXPHOS gene transcription and (ii) an FDG response group with increased 18-FDG uptake[852]. Increase in proliferation, as measured by a validated proliferation signature, suggested that patients in the OTR group were resistant to metformin treatment[852]. We hypothesised that mitochondrial mutations in complex 1 might explain the difference between these two groups, and be more common in the FDG response group. After comprehensive analysis using the cancer cell cycle encyclopedia (CCLE) and interrogating the mutational burden in the FDG response group it became clear this was a challenging hypothesis to test as (i) the sample size was modest (41 patients), (ii) the frequency of frameshift or deletions in complex 1 in the CCLE was low and appeared to be negatively selected in patient samples[853] (iii) the lack of paired samples between the tumour and the normal tissue made interpretation of this data difficult. Substitutions in complex 1 are common in the human genome and therefore it is difficult to postulate functional change in tumour cells without a comparable baseline genome (which was unavailable). After consideration of this analysis to my supervisors and independent experts, including Evgeny Izumchenko (University of Chicago), we decided not to further pursue this avenue.

Within the Department of Oncology in Oxford, I assisted in several clinical trials including ATOM (A single-centre pre-operative window of opportunity study of the effects of

atovaquone on hypoxia in non-small cell lung carcinoma), becoming a co-author on a manuscript published in *Clinical Cancer Research*[854]. My participation in IMAGO (A single arm exploratory study examining the feasibility of imaging glioblastoma pH using CEST MRI) has also led to several international presentations[855–857].

As discussed in *Chapter 5*, during my DPhil, the SARS-CoV-2 pandemic disrupted clinical and academic services globally. At this time, I helped on the frontline clinically in London caring for sufferers of COVID-19 and their families. At this time, in the hospital where I was helping to run the COVID-19 ward, I was able to participate in COVID-19 research in collaboration with Imperial College, London. This work involved collecting samples (nasal swabs, gingival, serum) from patients and staff that were infected with SARS-CoV-2. These samples were used in the validation of the in-house double binding antigen ELISA (Imperial Hybrid DABA) which was published in *The Lancet* [837] and in validating lateral flow assays for skin prick tests for SARS-CoV-2[858]. I lead on two projects clinically and academically, discussed in *Chapter 5*, both published in the *Journal of Infectious Diseases*[859,860]. Further I am a co-author of a paper in *The Lancet Infectious Diseases* testing the longevity of the Imperial Hybrid DABA test[861]. We are currently preparing a third manuscript on the long term clinical utility of the Imperial Hybrid DABA test over a 6 month window for submission to *Clinical Infectious Diseases*. I am also acknowledged for my contributions to a paper in review at *Nature Immunology* investigating predictive markers of COVID-19 disease severity[862].

Across Europe, I was principal investigator on two European grants (top-ranked two years running, total approx. €500,000), won five other smaller grants (CRUK - £15,000, UCARE - £10,000, OHSRC - £10,000, Oxford Foundry - £5,000, Pfizer - £50,000) as well as winning two prizes from Santander Universities. I ran these projects without a project manager for 8 months, due to lack of departmental recruitment. Beside these achievements, I was ranked 1st in UK national recruitment for Medical Oncology training and managed to

secure a highly competitive NIHR Academic Clinical Fellowship Medical Oncology here in Oxford.

The aforementioned clinical work and achievements, in addition to the work presented in this thesis, demonstrates how much my scientific, research and logistical acumen has matured over the course of my DPhil.

Acknowledgements

I would like to take this opportunity to thank all those who have contributed in any way, directly or indirectly, to the completion of this work.

- Professor Francesca Buffa for being my primary supervisor and her invaluable suggestions and support.
- Dr Simon Lord for his supervision and direction.
- Professor Gareth Bond for his enthusiasm, GWAS expertise and direction.
- Dr Lisa Heather, for early input on the project.
- Dr Lindy Castell for pastoral support and entertaining academic discussions.
- Professor Mark Middleton for his belief.
- Professor Adrian Harris for academic critique and advice.
- Mr Matteo Di Giovannantonio for everything, particularly his bioinformatics support.
- Dr Michael Fertleman for support, decision making and just generally getting it right.
- Ms Claire Brunt for helping with grant management and just generally being excellent.
- Professor Kevin Spring for miRNA advice and thursday morning discussions.
- Professor Myra McClure for academic discussions and content refinement.
- Dr Lonnie Swift for his common sense and support.
- Dr Mohammed Zuhair for helping me fight COVID-19 and recruit patients for our SARS-CoV-2 research.
- Dr Christopher Pereira for helping carry on data collection on the SARS-CoV-2 staff cohort, particularly around distribution and data collection from the post-COVID-19 syndrome questionnaire.

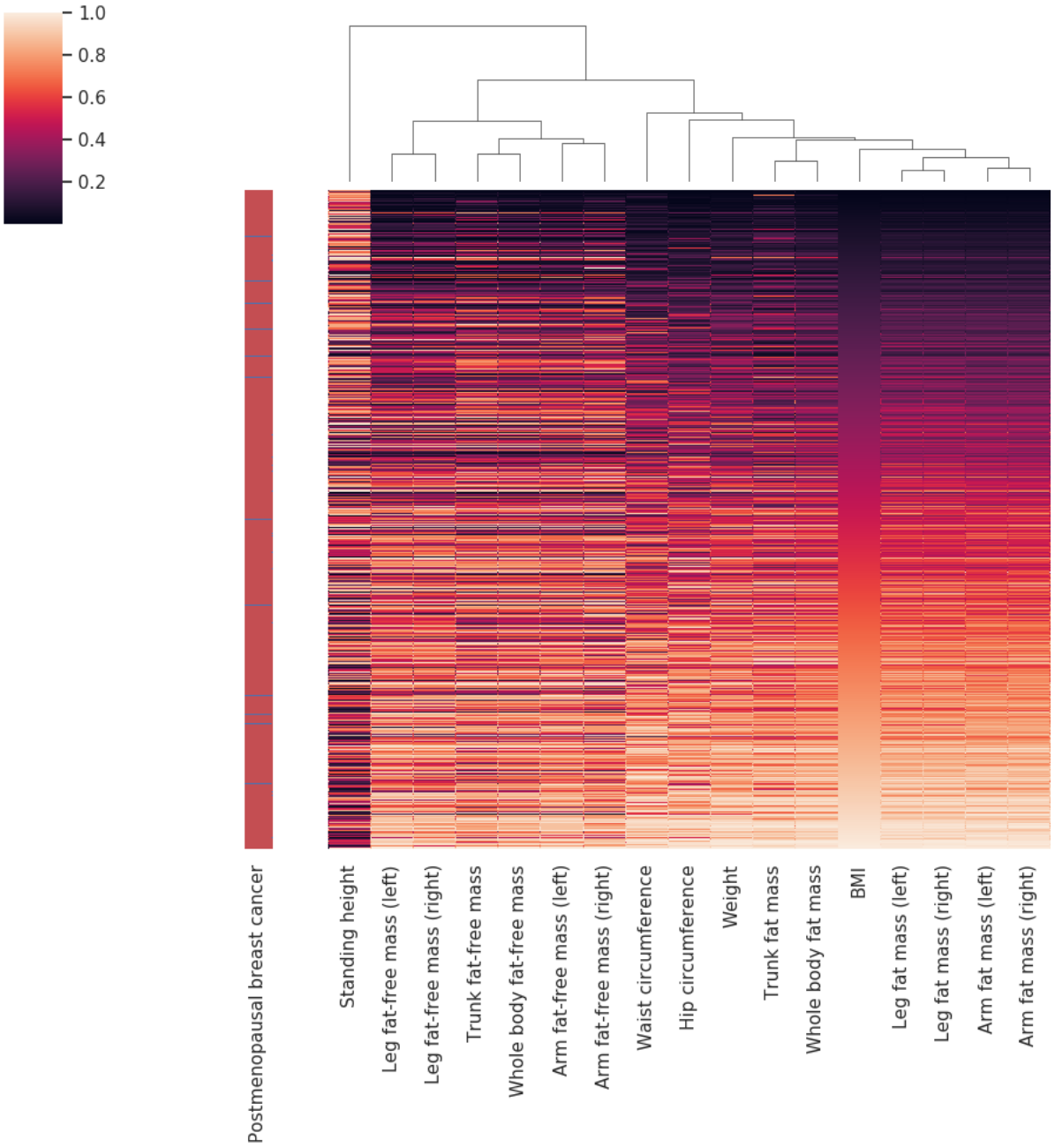
Alessandro Barberis, Ryan Carter, Philip Charlton, Wei-Chen Cheng, Yueachan Chi, Joachim Hagel, Fiona Hartley, Sakura Maezono, Ana Miar Cuervo, Dawn O'Reilly, Jayathilake Pahala Gedara, Clara Pavillet, Karen Sayal, Helen Sheldon, Charalampos Triantafyllidis, Ava Van Ess, Tim Maughan, Karl Morten, Simon Leedham, Richard Bryant,

Pedro Victori, Carolina Rosadas, Rachael Quinlan, Melanie Dani, Ping Zhang, Isaac Kitchen-Smith, Lingyun Xiong, Sarah Blagden, Richard Tedder, Graham Taylor, Alex Feyler, Ana Bakshi, Kevin Roberts, Dan Channer, Jim Wilkinson, Mary McKenna and others I have forgotten.

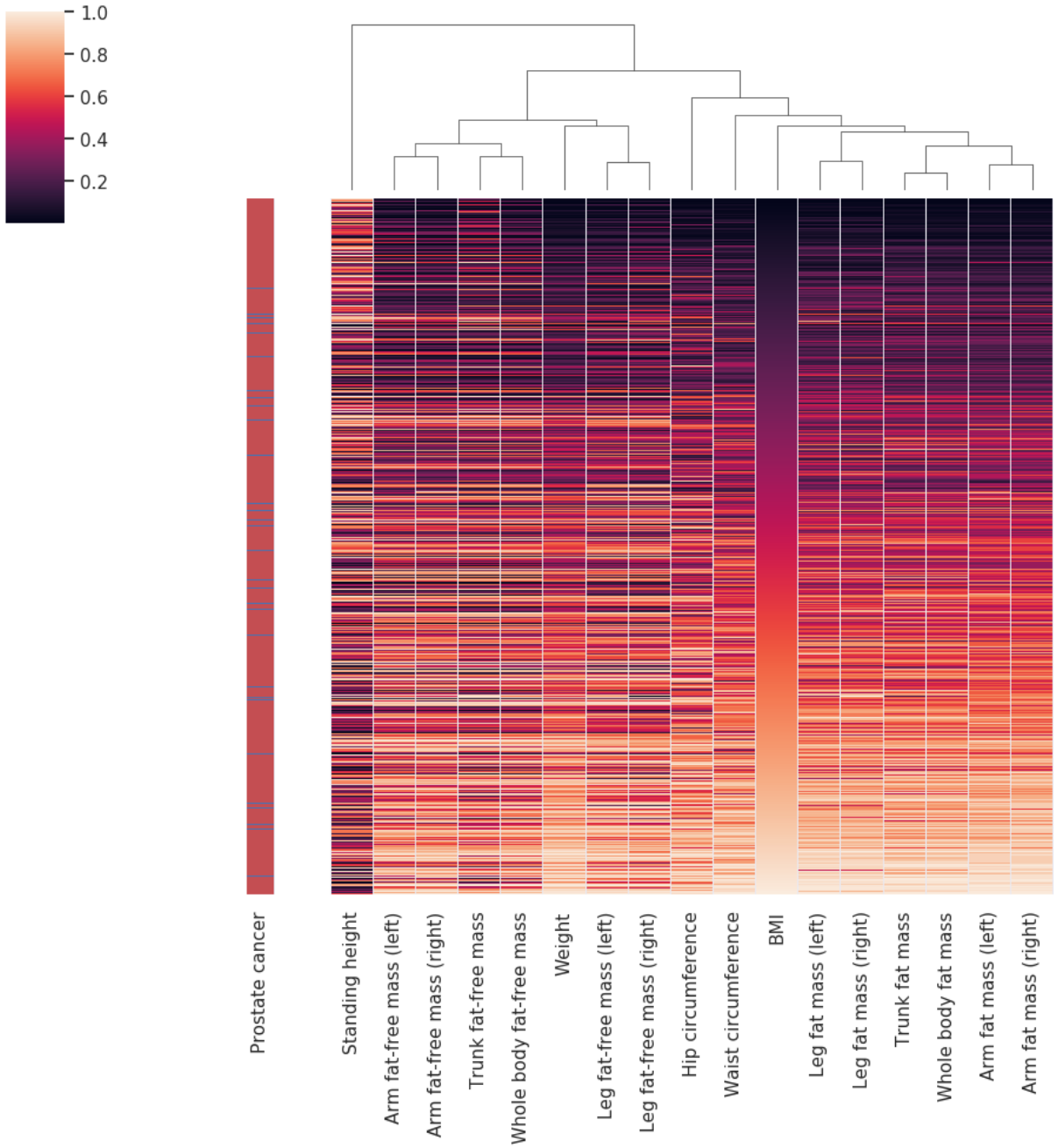
My family and friends for their considerable help and support. Special thanks go to Cas and Dave, without whom I would not have been possible.

I am also truly indebted to Cancer Research UK for their financial support.

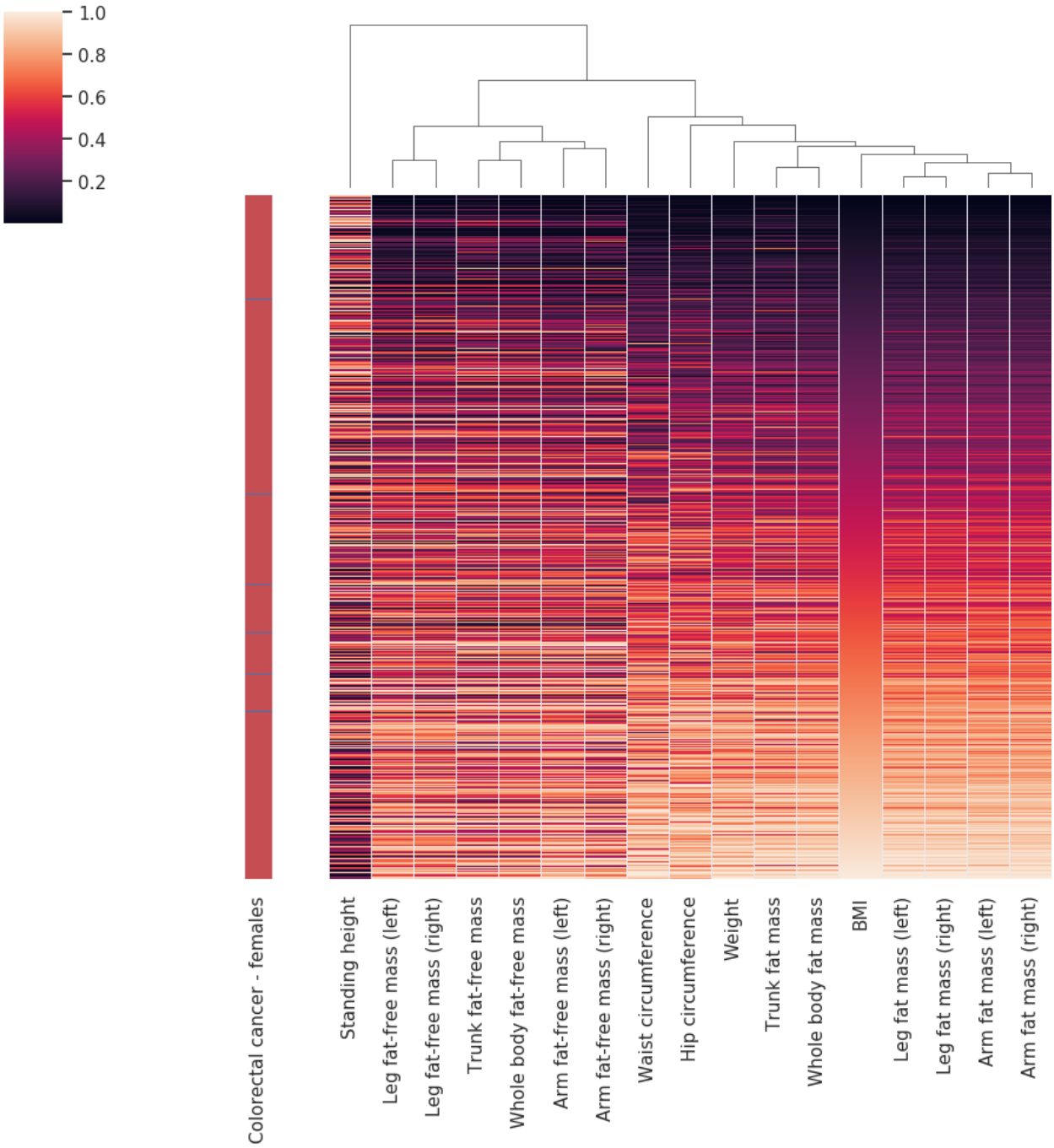
Appendices



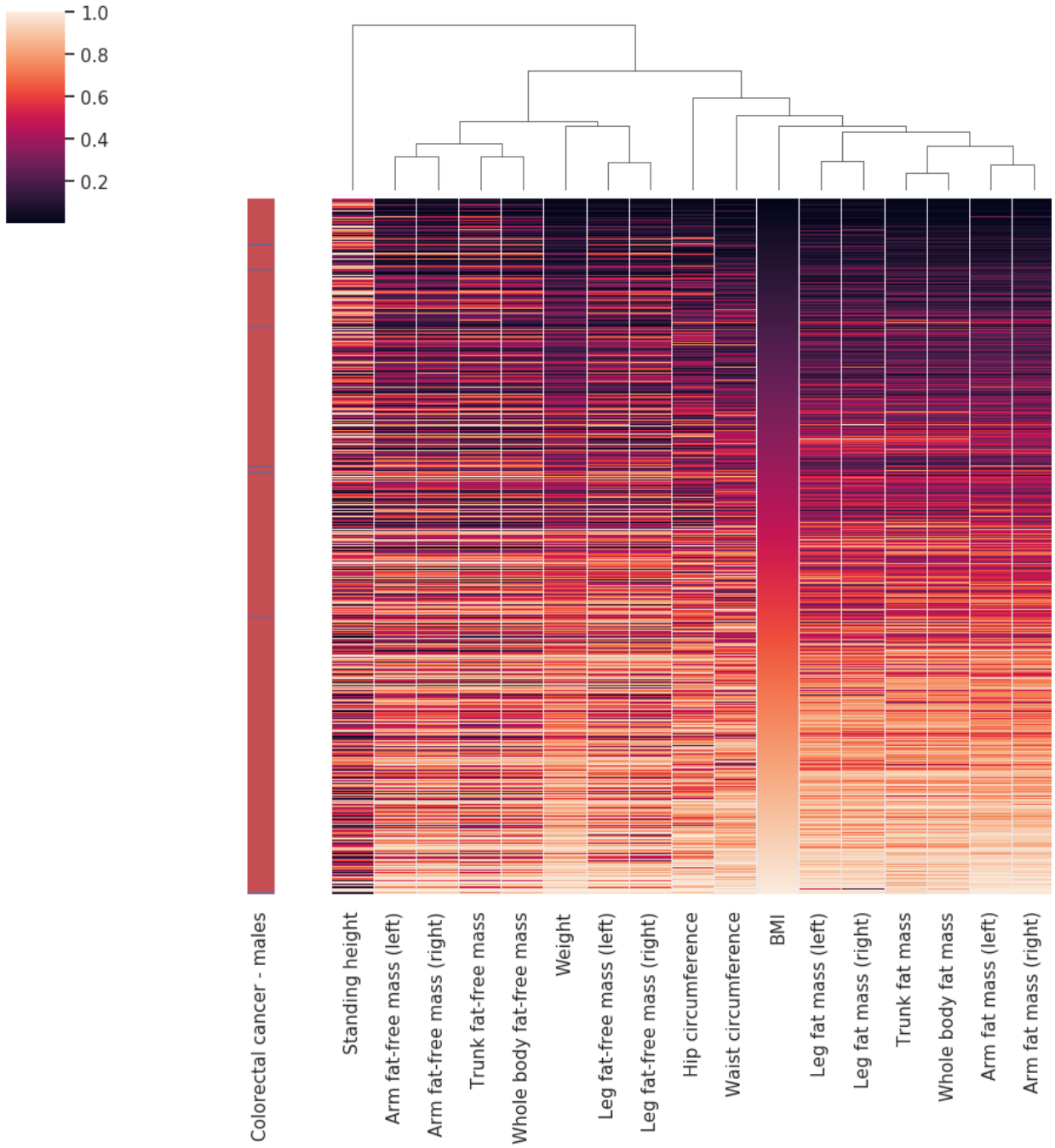
Supplementary figure S2.1: Clustermap with hierarchical clustering using Euclidean distance of normalised ranked anthropometric traits in the post-menopasual breast cancer analysis cohort. Colour indicates normalised rank value ranging from 0 to 1. Blue horizontal stripes in the left column indicate cancer cases.



Supplementary figure S2.2: Clustermap with hierarchical clustering using Euclidean distance of normalised ranked anthropometric traits in the prostate cancer analysis cohort. Colour indicates normalised rank value ranging from 0 to 1. Blue horizontal stripes in the left column indicate cancer cases.



Supplementary figure S2.3: Clustermap with hierarchical clustering using Euclidean distance of normalised ranked anthropometric traits in the colorectal cancer analysis cohort (women). Colour indicates normalised rank value ranging from 0 to 1. Blue horizontal stripes in the left column indicate cancer cases.



Supplementary figure S2.4: Clustermap with hierarchical clustering using Euclidean distance of normalised ranked anthropometric traits in the colorectal cancer analysis cohort (men). Colour indicates normalised rank value ranging from 0 to 1. Blue horizontal stripes in the left column indicate cancer cases.

a

	BMI (kg/m^2)	WBFM (kg)	WBFFM (kg)
Quintile 1	15.0-23.0	5.0-19.1	26.2-40.0
Quintile 2	23.0-25.2	19.1-23.4	40.0-42.3
Quintile 3	25.2-27.4	23.4-27.8	42.3-44.6
Quintile 4	27.4-30.7	27.8-34.0	44.6-47.6
Quintile 5	30.7-68.1	34.0-108.4	47.6-82.7

	p trend
BMI (kg/m^2)	3.44E-12
WBFM (kg)	1.85E-17
WBFFM (kg)	3.87E-19
PC1	2.37E-19
PC2	4.23E-03

b

	BMI (kg/m^2)	WBFM (kg)	WBFFM (kg)
Quintile 1	15.0-24.5	5.0, 15.7	33.0-57.3
Quintile 2	24.5-26.4	15.7-19.4	57.3-61.4
Quintile 3	26.4-28.3	19.4-23.1	61.4-65.2
Quintile 4	28.3-30.8	23.1-28.0	65.2-70.0
Quintile 5	30.8-68.4	28.0-108.1	70.0-100.0

	p trend
BMI (kg/m^2)	1.96E-04
WBFM (kg)	4.26E-05
WBFFM (kg)	0.828
PC1	0.0277
PC2	9.72E-07

c

	BMI (kg/m^2)	WBFM (kg)	WBFFM (kg)
Quintile 1	14.5-22.9	5.0-18.6	26.2-40.4
Quintile 2	22.9-25.0	18.6-23.1	40.4-42.8
Quintile 3	25.0-27.3	23.1-27.6	42.8-45.2
Quintile 4	27.3-30.8	27.6-34.2	45.2-48.3
Quintile 5	30.8-68.1	34.2-109.8	48.3-84.9

	p trend
BMI (kg/m^2)	0.945
WBFM (kg)	0.550
WBFFM (kg)	0.010
PC1	0.436
PC2	4.38E-03

d

	BMI (kg/m^2)	WBFM (kg)	WBFFM (kg)
Quintile 1	15.0-24.5	5.0-15.7	33.0-57.3
Quintile 2	24.5, 26.4	15.7-19.5	57.3-61.4
Quintile 3	26.4-28.3	19.5-23.1	61.4-65.2
Quintile 4	28.3-30.9	23.1-28.1	65.2-70.0
Quintile 5	30.9-68.4	28.1-108.1	70.0-100.0

	p trend
BMI (kg/m^2)	3.49E-06
WBFM (kg)	8.34E-06
WBFFM (kg)	6.31E-03
PC1	3.84E-06
PC2	0.379

Supplementary table S2.5: Thresholds for the quintiles and trend tests across quintiles, as dictated by data distribution, for BMI, WBFM and WBFFM in the post-menopausal breast cancer analysis cohort (a), prostate cancer analysis cohort (b) and colorectal cancer analysis cohorts (c - female, d - male).

	Estimate	Std. Error	Z	p	
(Intercept)	-3.634221	0.078847	-46.092	< 2e-16	***
as.factor(Whole_body_fat_free_mass_female_quantile)2	0.205088	0.054968	3.731	0.000191	***
as.factor(Whole_body_fat_free_mass_female_quantile)3	0.223557	0.054290	4.118	3.82e-05	***
as.factor(Whole_body_fat_free_mass_female_quantile)4	0.323487	0.053724	6.021	1.73e-09	***
as.factor(Whole_body_fat_free_mass_female_quantile)5	0.471998	0.053290	8.857	< 2e-16	***
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)0	0.043642	0.041292	1.057	0.290556	
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)55551	0.012347	0.457752	0.027	0.978482	
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)55553	0.180663	0.498254	0.363	0.716910	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)1	0.103709	0.070606	1.469	0.141874	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)55553	0.298297	1.305751	0.228	0.819297	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)55555	-0.966630	1.006599	-0.960	0.336907	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.012660	0.050527	0.251	0.802148	
as.factor(X1558.0.0.Alcohol.intake.frequency.)2	0.032510	0.047986	0.677	0.498105	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	-0.021432	0.056682	-0.378	0.705349	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.094857	0.054805	-1.731	0.083486	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.023713	0.064144	-0.370	0.711612	
as.factor(X2734.0.0.Number.of.live.births)0	0.666867	0.687924	0.969	0.332350	
as.factor(X2734.0.0.Number.of.live.births)1	0.183111	0.512494	0.357	0.720873	
as.factor(X2734.0.0.Number.of.live.births)2	0.209858	0.512839	0.409	0.682386	
as.factor(X2734.0.0.Number.of.live.births)4	-0.087432	0.046130	-1.895	0.058049	.
as.factor(X2734.0.0.Number.of.live.births)5	-0.184042	0.082998	-2.217	0.026594	*
as.factor(X2734.0.0.Number.of.live.births)6	-0.080965	0.157285	-0.515	0.606719	
as.factor(X2734.0.0.Number.of.live.births)7	-0.653742	0.307353	-2.127	0.033420	*
as.factor(Fhx.breast.cancer..all.)1	0.402437	0.044665	9.010	< 2e-16	***
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	0.121426	0.079126	1.535	0.124885	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	0.436185	0.326972	1.334	0.182200	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	-0.009184	1.041553	-0.009	0.992965	
as.factor(X21000.0.0.Ethnic.background)1	-0.188478	0.586217	-0.322	0.747819	
as.factor(X21000.0.0.Ethnic.background)1002	0.031645	0.103017	0.307	0.758704	
as.factor(X21000.0.0.Ethnic.background)1003	0.025209	0.087899	0.287	0.774264	
as.factor(X21000.0.0.Ethnic.background)2001	-0.472505	0.714246	-0.662	0.508264	
as.factor(X21000.0.0.Ethnic.background)2002	-1.055971	1.002576	-1.053	0.292223	
as.factor(X21000.0.0.Ethnic.background)2003	-0.246350	0.506545	-0.486	0.626731	
as.factor(X21000.0.0.Ethnic.background)2004	0.250789	0.361525	0.694	0.487873	
as.factor(X21000.0.0.Ethnic.background)3	1.570555	1.075893	1.460	0.144354	
as.factor(X21000.0.0.Ethnic.background)3001	0.284506	0.156269	1.821	0.068665	.
as.factor(X21000.0.0.Ethnic.background)3002	-0.205773	0.457689	-0.450	0.653004	
as.factor(X21000.0.0.Ethnic.background)3003	0.442273	1.021452	0.433	0.665026	
as.factor(X21000.0.0.Ethnic.background)3004	0.054122	0.323343	0.167	0.867069	
as.factor(X21000.0.0.Ethnic.background)4001	-0.480971	0.247543	-1.943	0.052019	.
as.factor(X21000.0.0.Ethnic.background)4002	-0.220759	0.284828	-0.775	0.438303	
as.factor(X21000.0.0.Ethnic.background)4003	0.153213	1.019088	0.150	0.880494	
as.factor(X21000.0.0.Ethnic.background)5	-0.109274	0.339733	-0.322	0.747720	
as.factor(X21000.0.0.Ethnic.background)55551	-0.483977	1.011336	-0.479	0.632257	
as.factor(X21000.0.0.Ethnic.background)6	-0.371911	0.228010	-1.631	0.102865	
as.factor(X20160.0.0.Ever.smoked)0	-0.034647	0.033898	-1.022	0.306732	
as.factor(X20160.0.0.Ever.smoked)6	-0.044330	0.273702	-0.162	0.871335	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)1	0.137255	0.034447	3.985	6.76e-05	***
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55551	0.262002	0.325737	0.804	0.421202	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55553	0.927588	0.654001	1.418	0.156095	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)1	-0.077562	0.096020	-0.808	0.419225	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)55553	-0.942101	1.288493	-0.731	0.464678	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)55555	-0.144101	0.185779	-0.776	0.437949	
as.factor(Town_5)2	-0.029049	0.050642	-0.574	0.566233	
as.factor(Town_5)3	-0.057029	0.051071	-1.117	0.264146	
as.factor(Town_5)4	0.051707	0.050249	1.029	0.303464	
as.factor(Town_5)5	-0.125828	0.054104	-2.326	0.020037	*
as.factor(Town_5)6	-0.729023	0.713694	-1.021	0.307028	
as.factor(X2754.0.0.Age.at.first.live.birth)0	-0.069268	0.511264	-0.135	0.892229	
as.factor(X2754.0.0.Age.at.first.live.birth)1	0.040433	0.057703	0.701	0.483481	
as.factor(X2754.0.0.Age.at.first.live.birth)3	0.109680	0.061512	1.783	0.074577	.
as.factor(X2714.0.0.Age.when.periods.started..menarche.)0	-0.066427	0.102872	-0.646	0.518457	
as.factor(X2714.0.0.Age.when.periods.started..menarche.)1	0.153654	0.073497	2.091	0.036562	*
as.factor(X2714.0.0.Age.when.periods.started..menarche.)3	0.065339	0.071357	0.916	0.359845	
as.factor(X2714.0.0.Age.when.periods.started..menarche.)4	1.250501	0.742734	1.684	0.092250	.
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)0	-0.057543	0.069238	-0.831	0.405924	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)1	-0.418795	0.275689	-1.519	0.128742	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)2	-0.233508	0.085464	-2.732	0.006291	**
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)3	-0.213507	0.060092	-3.553	0.000381	***
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)4	-0.117715	0.039972	-2.945	0.003230	**
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)6	0.054636	0.059700	0.915	0.360101	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)7	0.389589	0.365731	1.065	0.286770	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)8	0.049943	1.022362	0.049	0.961038	
as.factor(exercise_5)1	0.048522	0.051555	0.941	0.346619	
as.factor(exercise_5)2	0.042130	0.053084	0.794	0.427394	
as.factor(exercise_5)3	-0.002174	0.053035	-0.041	0.967301	
as.factor(exercise_5)4	-0.114346	0.054565	-2.096	0.036119	*
as.factor(exercise_5)5	-0.136614	0.055292	-2.471	0.013481	*
as.factor(Categorical.age.5.y)2	-0.398154	0.310212	-1.283	0.199320	
as.factor(Categorical.age.5.y)3	-0.644351	0.136788	-4.711	2.47e-06	***
as.factor(Categorical.age.5.y)4	-0.423189	0.059517	-7.110	1.16e-12	***
as.factor(Categorical.age.5.y)5	-0.172931	0.042603	-4.059	4.93e-05	***
as.factor(Categorical.age.5.y)7	-0.034655	0.043096	-0.804	0.421321	
as.factor(Categorical.age.5.y)8	0.134848	0.202339	0.666	0.505127	

Supplementary table S2.6: Logistic regression covariates in the WBFFM model predicting post-menopausal breast cancer

	Estimate	Std. Error	Z	p	
(Intercept)	-3.6334300	0.0797917	-45.536	< 2e-16	***
as.factor(Whole_body_fat_mass_female_quantile)2	0.1538960	0.0556399	2.766	0.005676	**
as.factor(Whole_body_fat_mass_female_quantile)3	0.2587697	0.0545230	4.746	2.07e-06	***
as.factor(Whole_body_fat_mass_female_quantile)4	0.3049090	0.0544297	5.602	2.12e-08	***
as.factor(Whole_body_fat_mass_female_quantile)5	0.4406917	0.0542651	8.121	4.62e-16	***
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)0	0.0372146	0.0412791	0.902	0.367304	
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)55551	0.0256914	0.4578337	0.056	0.955250	
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)55553	0.1734191	0.4975947	0.349	0.727454	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)1	0.1000620	0.0705910	1.417	0.156340	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)55553	0.2556149	1.2991682	0.197	0.844021	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)55555	-0.9756509	1.0066355	-0.969	0.332436	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.0316252	0.0506294	0.625	0.532207	
as.factor(X1558.0.0.Alcohol.intake.frequency.)2	0.0417524	0.0480155	0.870	0.384540	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	-0.0222652	0.0566903	-0.393	0.694503	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.0972986	0.0548356	-1.774	0.076002	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.0244064	0.0641505	-0.380	0.703607	
as.factor(X2734.0.0.Number.of.live.births)0	0.6601463	0.6876706	0.960	0.337068	
as.factor(X2734.0.0.Number.of.live.births)1	0.2004305	0.5123406	0.391	0.695645	
as.factor(X2734.0.0.Number.of.live.births)2	0.2158304	0.5126871	0.421	0.673771	
as.factor(X2734.0.0.Number.of.live.births)4	-0.0810690	0.0461254	-1.758	0.078820	.
as.factor(X2734.0.0.Number.of.live.births)5	-0.1741995	0.0830161	-2.098	0.035872	*
as.factor(X2734.0.0.Number.of.live.births)6	-0.0711411	0.1572427	-0.452	0.650960	
as.factor(X2734.0.0.Number.of.live.births)7	-0.6420615	0.3073132	-2.089	0.036683	*
as.factor(Fhx.breast.cancer..all.)1	0.4044743	0.0446591	9.057	< 2e-16	***
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	0.1293635	0.0791192	1.635	0.102040	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	0.4329577	0.3272011	1.323	0.185764	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	-0.0007932	1.0401954	-0.001	0.999392	
as.factor(X21000.0.0.Ethnic.background)1	-0.1581105	0.5861591	-0.270	0.787360	
as.factor(X21000.0.0.Ethnic.background)1002	0.0179424	0.1029892	0.174	0.861696	
as.factor(X21000.0.0.Ethnic.background)1003	0.0543590	0.0879140	0.618	0.536364	
as.factor(X21000.0.0.Ethnic.background)2001	-0.4849343	0.7142398	-0.679	0.497168	
as.factor(X21000.0.0.Ethnic.background)2002	-1.0453021	1.0025563	-1.043	0.297117	
as.factor(X21000.0.0.Ethnic.background)2003	-0.2839106	0.5064969	-0.561	0.575113	
as.factor(X21000.0.0.Ethnic.background)2004	0.2334872	0.3614456	0.646	0.518291	
as.factor(X21000.0.0.Ethnic.background)3	1.5517605	1.0789484	1.438	0.150373	
as.factor(X21000.0.0.Ethnic.background)3001	0.1715633	0.1556060	1.103	0.270223	
as.factor(X21000.0.0.Ethnic.background)3002	-0.2797643	0.4574152	-0.612	0.540789	
as.factor(X21000.0.0.Ethnic.background)3003	0.3362704	1.0209582	0.329	0.741878	
as.factor(X21000.0.0.Ethnic.background)3004	-0.0294524	0.3230280	-0.091	0.927353	
as.factor(X21000.0.0.Ethnic.background)4001	-0.4716159	0.2475190	-1.905	0.056732	.
as.factor(X21000.0.0.Ethnic.background)4002	-0.2535206	0.2848913	-0.890	0.373527	
as.factor(X21000.0.0.Ethnic.background)4003	0.1613557	1.0189920	0.158	0.874182	
as.factor(X21000.0.0.Ethnic.background)5	-0.1063142	0.3397769	-0.313	0.754361	
as.factor(X21000.0.0.Ethnic.background)55551	-0.4903550	1.0113156	-0.485	0.627770	
as.factor(X21000.0.0.Ethnic.background)6	-0.4060038	0.2279394	-1.781	0.074881	.
as.factor(X20160.0.0.Ever.smoked)0	-0.0369083	0.0338872	-1.089	0.276086	
as.factor(X20160.0.0.Ever.smoked)6	-0.0567844	0.2737060	-0.207	0.835647	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)1	0.1334846	0.0344425	3.876	0.000106	***
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55551	0.2664692	0.3258033	0.818	0.413424	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55553	0.9268360	0.6524977	1.420	0.155479	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)1	-0.0809984	0.0960102	-0.844	0.398868	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)55553	-0.9137193	1.2800097	-0.714	0.475327	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)55555	-0.1525726	0.1857204	-0.822	0.411351	
as.factor(Town_5)2	-0.0317744	0.0506401	-0.627	0.530361	
as.factor(Town_5)3	-0.0609711	0.0510747	-1.194	0.232571	
as.factor(Town_5)4	0.0471310	0.0502606	0.938	0.348382	
as.factor(Town_5)5	-0.1341518	0.0541397	-2.478	0.013216	*
as.factor(Town_5)6	-0.7354761	0.7136580	-1.031	0.302741	
as.factor(X2754.0.0.Age.at.first.live.birth)0	-0.0767215	0.5111131	-0.150	0.880680	
as.factor(X2754.0.0.Age.at.first.live.birth)1	0.0249709	0.0577310	0.433	0.665350	
as.factor(X2754.0.0.Age.at.first.live.birth)3	0.1330397	0.0615406	2.162	0.030632	*
as.factor(X2714.0.0.Age.when.periods.started..menarche.)0	-0.0607213	0.1028711	-0.590	0.555012	
as.factor(X2714.0.0.Age.when.periods.started..menarche.)1	0.1518098	0.0735229	2.065	0.038942	*
as.factor(X2714.0.0.Age.when.periods.started..menarche.)3	0.0635789	0.0713508	0.891	0.372889	
as.factor(X2714.0.0.Age.when.periods.started..menarche.)4	1.1896747	0.7430129	1.601	0.109344	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)0	-0.0572291	0.0692366	-0.827	0.408479	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)1	-0.4276378	0.2757002	-1.551	0.120878	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)2	-0.2454138	0.0854324	-2.873	0.004071	**
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)3	-0.2223127	0.0600787	-3.700	0.000215	***
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)4	-0.1211051	0.0399614	-3.031	0.002441	**
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)6	0.0528190	0.0597020	0.885	0.376312	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)7	0.3833223	0.3657465	1.048	0.294613	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)8	0.0192013	1.0229412	0.019	0.985024	
as.factor(exercise_5)1	0.0391996	0.0515931	0.760	0.447384	
as.factor(exercise_5)2	0.0490739	0.0530906	0.924	0.355308	
as.factor(exercise_5)3	0.0105383	0.0530786	0.199	0.842621	
as.factor(exercise_5)4	-0.0913205	0.0546645	-1.671	0.094808	.
as.factor(exercise_5)5	-0.1062178	0.0554691	-1.915	0.055505	.
as.factor(Categorical.age.5.y)2	-0.3351832	0.3101955	-1.081	0.279895	
as.factor(Categorical.age.5.y)3	-0.5991274	0.1367906	-4.380	1.19e-05	***
as.factor(Categorical.age.5.y)4	-0.3909167	0.0594687	-6.573	4.91e-11	***
as.factor(Categorical.age.5.y)5	-0.1540064	0.0425842	-3.617	0.000299	***
as.factor(Categorical.age.5.y)7	-0.0525192	0.0430576	-1.220	0.222562	
as.factor(Categorical.age.5.y)8	0.1078792	0.2023077	0.533	0.593865	

Supplementary table S2.7: Logistic regression covariates in the WBFM model predicting post-menopausal breast cancer

	Estimate	Std. Error	Z	p	
(Intercept)	-3.582951	0.079392	-45.130	< 2e-16	***
as.factor(BMI_quantile)2	0.146298	0.054524	2.683	0.007293	**
as.factor(BMI_quantile)3	0.169759	0.054363	3.123	0.001792	**
as.factor(BMI_quantile)4	0.251345	0.053874	4.665	3.08e-06	***
as.factor(BMI_quantile)5	0.370861	0.054291	6.831	8.44e-12	***
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)0	0.037459	0.041275	0.908	0.364122	
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)55551	0.019043	0.457798	0.042	0.966821	
as.factor(X2784.0.0.Ever.taken.oral.contraceptive.pill)55553	0.164801	0.497603	0.331	0.740502	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)1	0.103714	0.070596	1.469	0.141802	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)55553	0.281326	1.297427	0.217	0.828338	
as.factor(X3591.0.0.Ever.had.hysterectomy..womb.removed.)55555	-0.976187	1.006673	-0.970	0.332188	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.030366	0.050673	0.599	0.548995	
as.factor(X1558.0.0.Alcohol.intake.frequency.)2	0.041278	0.048025	0.860	0.390063	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	-0.021452	0.056696	-0.378	0.705153	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.097418	0.054862	-1.776	0.075783	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.026650	0.064173	-0.415	0.677934	
as.factor(X2734.0.0.Number.of.live.births)0	0.678811	0.687623	0.987	0.323552	
as.factor(X2734.0.0.Number.of.live.births)1	0.214197	0.512345	0.418	0.675894	
as.factor(X2734.0.0.Number.of.live.births)2	0.226009	0.512690	0.441	0.659337	
as.factor(X2734.0.0.Number.of.live.births)4	-0.081381	0.046121	-1.765	0.077647	.
as.factor(X2734.0.0.Number.of.live.births)5	-0.172823	0.083006	-2.082	0.037338	*
as.factor(X2734.0.0.Number.of.live.births)6	-0.075410	0.157249	-0.480	0.631542	
as.factor(X2734.0.0.Number.of.live.births)7	-0.639904	0.307279	-2.082	0.037298	*
as.factor(Fhx.breast.cancer..all.)1	0.407227	0.044653	9.120	< 2e-16	***
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	0.130117	0.079362	1.640	0.101104	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	0.421351	0.327141	1.288	0.197753	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	0.018775	1.039695	0.018	0.985592	
as.factor(X21000.0.0.Ethnic.background)1	-0.165618	0.586138	-0.283	0.777516	
as.factor(X21000.0.0.Ethnic.background)1002	0.013921	0.102976	0.135	0.892462	
as.factor(X21000.0.0.Ethnic.background)1003	0.047276	0.087894	0.538	0.590665	
as.factor(X21000.0.0.Ethnic.background)2001	-0.478908	0.714206	-0.671	0.502510	
as.factor(X21000.0.0.Ethnic.background)2002	-1.039700	1.002516	-1.037	0.299694	
as.factor(X21000.0.0.Ethnic.background)2003	-0.288835	0.506447	-0.570	0.568464	
as.factor(X21000.0.0.Ethnic.background)2004	0.219383	0.361377	0.607	0.543802	
as.factor(X21000.0.0.Ethnic.background)3	1.526591	1.078419	1.416	0.156898	
as.factor(X21000.0.0.Ethnic.background)3001	0.159198	0.155564	1.023	0.306137	
as.factor(X21000.0.0.Ethnic.background)3002	-0.293588	0.457411	-0.642	0.520973	
as.factor(X21000.0.0.Ethnic.background)3003	0.304760	1.021026	0.298	0.765334	
as.factor(X21000.0.0.Ethnic.background)3004	-0.056877	0.322904	-0.176	0.860182	
as.factor(X21000.0.0.Ethnic.background)4001	-0.468093	0.247512	-1.891	0.058598	.
as.factor(X21000.0.0.Ethnic.background)4002	-0.248166	0.284963	-0.871	0.383825	
as.factor(X21000.0.0.Ethnic.background)4003	0.165444	1.018865	0.162	0.871006	
as.factor(X21000.0.0.Ethnic.background)5	-0.159283	0.339526	-0.469	0.638974	
as.factor(X21000.0.0.Ethnic.background)55551	-0.514877	1.011270	-0.509	0.610655	
as.factor(X21000.0.0.Ethnic.background)6	-0.429484	0.227911	-1.884	0.059506	.
as.factor(X20160.0.0.Ever.smoked)0	-0.040852	0.033872	-1.206	0.227788	
as.factor(X20160.0.0.Ever.smoked)6	-0.056449	0.273705	-0.206	0.836602	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)1	0.134294	0.034443	3.899	9.66e-05	***
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55551	0.267595	0.325784	0.821	0.411426	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55553	0.917047	0.652269	1.406	0.159744	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)1	-0.080065	0.096019	-0.834	0.404364	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)55553	-0.937597	1.279251	-0.733	0.463603	
as.factor(X2834.0.0.Bilateral.oophorectomy..both.ovaries.removed.)55555	-0.150068	0.185712	-0.808	0.419052	
as.factor(Town_5)2	-0.032261	0.050637	-0.637	0.524049	
as.factor(Town_5)3	-0.061359	0.051077	-1.201	0.229631	
as.factor(Town_5)4	0.046773	0.050267	0.930	0.352118	
as.factor(Town_5)5	-0.136609	0.054176	-2.522	0.011684	*
as.factor(Town_5)6	-0.747296	0.713617	-1.047	0.295010	
as.factor(X2754.0.0.Age.at.first.live.birth)0	-0.087277	0.511115	-0.171	0.864414	
as.factor(X2754.0.0.Age.at.first.live.birth)1	0.021905	0.057760	0.379	0.704512	
as.factor(X2754.0.0.Age.at.first.live.birth)3	0.132916	0.061547	2.160	0.030806	*
as.factor(X2714.0.0.Age.when.periods.started..menarche.)0	-0.061649	0.102872	-0.599	0.548991	
as.factor(X2714.0.0.Age.when.periods.started..menarche.)1	0.144595	0.073637	1.964	0.049573	*
as.factor(X2714.0.0.Age.when.periods.started..menarche.)3	0.069493	0.071392	0.973	0.330350	
as.factor(X2714.0.0.Age.when.periods.started..menarche.)4	1.182376	0.742074	1.593	0.111084	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)0	-0.059324	0.069229	-0.857	0.391493	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)1	-0.436781	0.275660	-1.584	0.113082	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)2	-0.247528	0.085450	-2.897	0.003770	**
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)3	-0.221125	0.060073	-3.681	0.000232	***
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)4	-0.121733	0.039958	-3.047	0.002315	**
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)6	0.054013	0.059701	0.905	0.365619	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)7	0.395314	0.365667	1.081	0.279663	
as.factor(X3581.0.0.Age.at.menopause..last.menstrual.period.)8	0.014732	1.022838	0.014	0.988508	
as.factor(exercise_5)1	0.047057	0.051568	0.913	0.361486	
as.factor(exercise_5)2	0.050785	0.053093	0.957	0.338806	
as.factor(exercise_5)3	0.009742	0.053081	0.184	0.854383	
as.factor(exercise_5)4	-0.096555	0.054667	-1.766	0.077355	.
as.factor(exercise_5)5	-0.115902	0.055445	-2.090	0.036581	*
as.factor(Categorical.age.5.y)2	-0.337578	0.310180	-1.088	0.276450	
as.factor(Categorical.age.5.y)3	-0.598941	0.136788	-4.379	1.19e-05	***
as.factor(Categorical.age.5.y)4	-0.390223	0.059466	-6.562	5.31e-11	***
as.factor(Categorical.age.5.y)5	-0.153766	0.042582	-3.611	0.000305	***
as.factor(Categorical.age.5.y)7	-0.053327	0.043054	-1.239	0.215492	
as.factor(Categorical.age.5.y)8	0.107161	0.202289	0.530	0.596292	

Supplementary table S2.8: Logistic regression covariates in the BMI model predicting post-menopausal breast cancer

a

	BMI	WBFM	WBFFM	Standing Height
BMI	1	0.92	0.62	-0.14
WBFM		1	0.64	0.14
WBFFM			1	0.49
Standing Height				1

b

	BMI	WBFM	WBFFM	Standing Height
BMI	1	0.89	0.61	-0.06
WBFM		1	0.53	0.12
WBFFM			1	0.62
Standing Height				1

c

	BMI	WBFM	WBFFM	Standing Height
BMI	1	0.93	0.61	-0.14
WBFM		1	0.63	0.13
WBFFM			1	0.50
Standing Height				1

d

	BMI	WBFM	WBFFM	Standing Height
BMI	1	0.90	0.61	-0.06
WBFM		1	0.53	0.12
WBFFM			1	0.62
Standing Height				1

Supplementary table S2.9: Spearman's rank correlation coefficients of BMI, WBFM, WBFFM and Standing Height in the post-menopausal analysis cohort (a), the prostate cancer analysis cohort (b), the female colorectal cancer analysis cohort (c) and the male colorectal cancer analysis cohort (d). P values for all correlations are <0.001.

	Estimate	Std. Error	Z	p	
(Intercept)	-6.942e+00	2.099e-01	-33.064	< 2e-16	***
as.factor(Whole_body_fat_free_mass_male_quantile)2	1.082e-01	3.982e-02	2.717	0.006580	**
as.factor(Whole_body_fat_free_mass_male_quantile)3	9.998e-02	4.086e-02	2.447	0.014400	*
as.factor(Whole_body_fat_free_mass_male_quantile)4	6.720e-02	4.230e-02	1.588	0.112181	
as.factor(Whole_body_fat_free_mass_male_quantile)5	1.442e-02	4.582e-02	0.315	0.753039	
as.factor(X20160.0.0.Ever.smoked)0	5.552e-02	2.890e-02	1.921	0.054745	.
as.factor(X20160.0.0.Ever.smoked)8888888	3.382e-01	1.852e-01	1.826	0.067857	.
as.factor(X21000.0.0.Ethnic.background)1	3.547e-01	3.275e-01	1.083	0.278723	
as.factor(X21000.0.0.Ethnic.background)1002	8.351e-02	8.321e-02	1.004	0.315547	
as.factor(X21000.0.0.Ethnic.background)1003	-5.167e-02	9.198e-02	-0.562	0.574277	
as.factor(X21000.0.0.Ethnic.background)2001	2.065e-01	5.126e-01	0.403	0.687058	
as.factor(X21000.0.0.Ethnic.background)2002	9.210e-01	4.708e-01	1.956	0.050455	.
as.factor(X21000.0.0.Ethnic.background)2003	-4.125e-01	5.077e-01	-0.813	0.416485	
as.factor(X21000.0.0.Ethnic.background)2004	2.971e-01	3.439e-01	0.864	0.387557	
as.factor(X21000.0.0.Ethnic.background)3001	-6.282e-01	1.872e-01	-3.356	0.000789	***
as.factor(X21000.0.0.Ethnic.background)3002	-3.915e-01	3.265e-01	-1.199	0.230602	
as.factor(X21000.0.0.Ethnic.background)3003	4.376e-01	7.279e-01	0.601	0.547730	
as.factor(X21000.0.0.Ethnic.background)3004	-1.831e-01	2.742e-01	-0.668	0.504186	
as.factor(X21000.0.0.Ethnic.background)4001	1.050e+00	1.245e-01	8.434	< 2e-16	***
as.factor(X21000.0.0.Ethnic.background)4002	8.272e-01	1.566e-01	5.283	1.27e-07	***
as.factor(X21000.0.0.Ethnic.background)5	-4.120e-01	3.847e-01	-1.071	0.284217	
as.factor(X21000.0.0.Ethnic.background)55551	6.021e-01	5.216e-01	1.154	0.248354	
as.factor(X21000.0.0.Ethnic.background)55553	-7.654e-02	2.245e-01	-0.341	0.733147	
as.factor(X21000.0.0.Ethnic.background)6	-3.739e-02	1.881e-01	-0.199	0.842464	
as.factor(Town_5)2	1.420e-02	4.006e-02	0.354	0.723033	
as.factor(Town_5)3	-5.692e-05	4.057e-02	-0.001	0.998881	
as.factor(Town_5)4	-5.590e-02	4.231e-02	-1.321	0.186462	
as.factor(Town_5)5	-1.169e-01	4.524e-02	-2.584	0.009770	**
as.factor(Town_5)6	-9.499e-02	4.192e-01	-0.227	0.820755	
as.factor(X1369.0.0.Beef.intake)0	-5.156e-02	6.493e-02	-0.794	0.427152	
as.factor(X1369.0.0.Beef.intake)2	-1.740e-03	3.213e-02	-0.054	0.956811	
as.factor(X1369.0.0.Beef.intake)3	-2.350e-02	4.512e-02	-0.521	0.602538	
as.factor(X1369.0.0.Beef.intake)4	4.455e-02	2.772e-01	0.161	0.872293	
as.factor(X1369.0.0.Beef.intake)5	-7.495e-01	7.274e-01	-1.030	0.302824	
as.factor(X1369.0.0.Beef.intake)55551	-4.483e-01	2.856e-01	-1.570	0.116483	
as.factor(X1369.0.0.Beef.intake)55553	-1.046e-01	1.069e+00	-0.098	0.922074	
as.factor(X1379.0.0.Lamb.mutton.intake)0	-1.026e-01	4.809e-02	-2.134	0.032839	*
as.factor(X1379.0.0.Lamb.mutton.intake)2	2.472e-02	3.356e-02	0.737	0.461338	
as.factor(X1379.0.0.Lamb.mutton.intake)3	4.842e-02	7.489e-02	0.647	0.517910	
as.factor(X1379.0.0.Lamb.mutton.intake)4	-6.309e-01	7.225e-01	-0.873	0.382577	
as.factor(X1379.0.0.Lamb.mutton.intake)5	-7.393e-01	1.026e+00	-0.720	0.471257	
as.factor(X1379.0.0.Lamb.mutton.intake)55551	1.797e-01	1.772e-01	1.015	0.310325	
as.factor(X1379.0.0.Lamb.mutton.intake)55553	-6.903e-01	1.065e+00	-0.648	0.516799	
as.factor(X1329.0.0.Oily.fish.intake)0	-6.407e-02	5.267e-02	-1.216	0.223863	
as.factor(X1329.0.0.Oily.fish.intake)1	2.757e-04	3.198e-02	0.009	0.993123	
as.factor(X1329.0.0.Oily.fish.intake)3	3.855e-03	3.767e-02	0.102	0.918489	
as.factor(X1329.0.0.Oily.fish.intake)4	-1.509e-02	1.421e-01	-0.106	0.915441	
as.factor(X1329.0.0.Oily.fish.intake)5	-1.032e-02	2.442e-01	-0.042	0.966281	
as.factor(X1329.0.0.Oily.fish.intake)55551	-1.560e-01	1.944e-01	-0.802	0.422270	
as.factor(X1329.0.0.Oily.fish.intake)55553	-3.315e-01	1.030e+00	-0.322	0.747484	
as.factor(exercise_5)2	2.921e-02	4.712e-02	0.620	0.535276	
as.factor(exercise_5)3	8.214e-02	4.699e-02	1.748	0.080463	.
as.factor(exercise_5)4	9.586e-02	4.645e-02	2.064	0.039029	*
as.factor(exercise_5)5	1.162e-02	4.763e-02	0.244	0.807332	
as.factor(exercise_5)6	3.709e-02	4.796e-02	0.773	0.439348	
as.factor(X20003.Saw.palmetto)1	4.079e-01	1.922e-01	2.122	0.033835	*
as.factor(X20003.Aspirin)1	8.926e-03	3.637e-02	0.245	0.806143	
as.factor(X20003.Statins)1	-3.058e-02	3.528e-02	-0.867	0.386051	
as.factor(X20003.Metformin)1	-1.328e-01	1.067e-01	-1.245	0.213250	
as.factor(X20003.Testosterone.supplement)1	3.727e-01	2.884e-01	1.292	0.196194	
as.factor(X20003.Anabolic.steroids)1	2.423e-01	5.177e-01	0.468	0.639826	
as.factor(X20003.5.alpha.reductase)1	1.690e-01	1.046e-01	1.616	0.106176	
as.factor(Fhx.breast.cancer..all.)1	1.299e-01	4.152e-02	3.129	0.001756	**
as.factor(X20002.Touchscreen...Prostatitis)1	-3.171e-01	4.171e-01	-0.760	0.447100	
as.factor(X6179.Touchscreen...Selenium)1	-4.828e-02	9.179e-02	-0.526	0.598864	
as.factor(X6179.Touchscreen...Selenium)55551	-4.304e-01	5.108e-01	-0.843	0.399470	
as.factor(X6155.Touchscreen.Vitamin.E.supplementation)1	2.107e-01	8.851e-02	2.380	0.017300	*
as.factor(X6155.Touchscreen...Vitamin.D)1	-2.333e-02	8.815e-02	-0.265	0.791223	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	1.657e-02	3.607e-02	0.459	0.645913	
as.factor(X1558.0.0.Alcohol.intake.frequency.)3	1.663e-02	3.763e-02	0.442	0.658624	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	5.977e-03	5.461e-02	0.109	0.912842	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	3.861e-02	5.720e-02	0.675	0.499659	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	-1.371e+00	1.010e+00	-1.358	0.174493	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-5.658e-02	6.593e-02	-0.858	0.390772	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	-2.561e-01	8.067e-02	-3.175	0.001498	**
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	-3.771e-01	3.081e-01	-1.224	0.221027	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	5.304e-01	7.422e-01	0.715	0.474848	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)6	2.627e+00	1.311e+00	2.003	0.045194	*
as.factor(Fhx.prostate.cancer..all.)1	6.015e-01	3.952e-02	15.219	< 2e-16	***
as.factor(Categorical.age.5.y)3	1.525e+00	2.174e-01	7.017	2.26e-12	***
as.factor(Categorical.age.5.y)4	2.589e+00	2.060e-01	12.569	< 2e-16	***
as.factor(Categorical.age.5.y)5	3.213e+00	2.032e-01	15.811	< 2e-16	***
as.factor(Categorical.age.5.y)6	3.874e+00	2.019e-01	19.188	< 2e-16	***
as.factor(Categorical.age.5.y)7	4.115e+00	2.023e-01	20.341	< 2e-16	***
as.factor(Categorical.age.5.y)8	4.337e+00	2.399e-01	18.074	< 2e-16	***

Supplementary table S2.10: Logistic regression covariates in the WBFFM model predicting prostate cancer

	Estimate	Std. Error	Z	p	
(Intercept)	-6.846919	0.209260	-32.720	< 2e-16	***
as.factor(Whole_body_fat_mass_male_quantile)2	0.003510	0.042907	0.082	0.934805	
as.factor(Whole_body_fat_mass_male_quantile)3	0.024148	0.042222	0.572	0.567368	
as.factor(Whole_body_fat_mass_male_quantile)4	-0.044694	0.043135	-1.036	0.300141	
as.factor(Whole_body_fat_mass_male_quantile)5	-0.183715	0.045411	-4.046	5.22e-05	***
as.factor(X20160.0.0.Ever.smoked)0	0.048799	0.028939	1.686	0.091742	.
as.factor(X20160.0.0.Ever.smoked)8888888	0.347472	0.185237	1.876	0.060679	.
as.factor(X21000.0.0.Ethnic.background)1	0.358229	0.327575	1.094	0.274140	
as.factor(X21000.0.0.Ethnic.background)1002	0.084161	0.083219	1.011	0.311862	
as.factor(X21000.0.0.Ethnic.background)1003	-0.054534	0.091985	-0.593	0.553276	
as.factor(X21000.0.0.Ethnic.background)2001	0.197760	0.512721	0.386	0.699714	
as.factor(X21000.0.0.Ethnic.background)2002	0.915815	0.470732	1.946	0.051713	.
as.factor(X21000.0.0.Ethnic.background)2003	-0.423505	0.507685	-0.834	0.404175	
as.factor(X21000.0.0.Ethnic.background)2004	0.287913	0.343849	0.837	0.402410	
as.factor(X21000.0.0.Ethnic.background)3001	-0.667582	0.186799	-3.574	0.000352	***
as.factor(X21000.0.0.Ethnic.background)3002	-0.420432	0.326490	-1.288	0.197838	
as.factor(X21000.0.0.Ethnic.background)3003	0.364441	0.727835	0.501	0.616569	
as.factor(X21000.0.0.Ethnic.background)3004	-0.244506	0.273905	-0.893	0.372036	
as.factor(X21000.0.0.Ethnic.background)4001	1.041503	0.124533	8.363	< 2e-16	***
as.factor(X21000.0.0.Ethnic.background)4002	0.816323	0.156575	5.214	1.85e-07	***
as.factor(X21000.0.0.Ethnic.background)5	-0.478026	0.384685	-1.243	0.213999	
as.factor(X21000.0.0.Ethnic.background)55551	0.576828	0.521998	1.105	0.269143	
as.factor(X21000.0.0.Ethnic.background)55553	-0.082992	0.224554	-0.370	0.711690	
as.factor(X21000.0.0.Ethnic.background)6	-0.057653	0.188059	-0.307	0.759171	
as.factor(Town_5)2	0.014823	0.040066	0.370	0.711412	
as.factor(Town_5)3	0.002328	0.040577	0.057	0.954242	
as.factor(Town_5)4	-0.051638	0.042328	-1.220	0.222478	
as.factor(Town_5)5	-0.112844	0.045239	-2.494	0.012618	*
as.factor(Town_5)6	-0.095924	0.419446	-0.229	0.819108	
as.factor(X1369.0.0.Beef.intake)0	-0.061604	0.064999	-0.948	0.343246	
as.factor(X1369.0.0.Beef.intake)2	0.003519	0.032152	0.109	0.912847	
as.factor(X1369.0.0.Beef.intake)3	-0.013704	0.045184	-0.303	0.761665	
as.factor(X1369.0.0.Beef.intake)4	0.055031	0.277186	0.199	0.842626	
as.factor(X1369.0.0.Beef.intake)5	-0.750483	0.727310	-1.032	0.302137	
as.factor(X1369.0.0.Beef.intake)55551	-0.449003	0.285721	-1.571	0.116073	
as.factor(X1369.0.0.Beef.intake)55553	-0.113350	1.068781	-0.106	0.915539	
as.factor(X1379.0.0.Lamb.mutton.intake)0	-0.105876	0.048097	-2.201	0.027715	*
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.027098	0.033554	0.808	0.419327	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.050969	0.074905	0.680	0.496218	
as.factor(X1379.0.0.Lamb.mutton.intake)4	-0.638622	0.722581	-0.884	0.376800	
as.factor(X1379.0.0.Lamb.mutton.intake)5	-0.737193	1.026598	-0.718	0.472700	
as.factor(X1379.0.0.Lamb.mutton.intake)55551	0.184375	0.177170	1.041	0.298029	
as.factor(X1379.0.0.Lamb.mutton.intake)55553	-0.690185	1.063360	-0.649	0.516299	
as.factor(X1329.0.0.Oily.fish.intake)0	-0.060134	0.052693	-1.141	0.253780	
as.factor(X1329.0.0.Oily.fish.intake)1	0.002120	0.031992	0.066	0.947154	
as.factor(X1329.0.0.Oily.fish.intake)3	0.003420	0.037678	0.091	0.927666	
as.factor(X1329.0.0.Oily.fish.intake)4	-0.015792	0.142094	-0.111	0.911506	
as.factor(X1329.0.0.Oily.fish.intake)5	-0.003863	0.244206	-0.016	0.987378	
as.factor(X1329.0.0.Oily.fish.intake)55551	-0.149507	0.194439	-0.769	0.441943	
as.factor(X1329.0.0.Oily.fish.intake)55553	-0.369072	1.029923	-0.358	0.720082	
as.factor(exercise_5)2	0.018217	0.047171	0.386	0.699356	
as.factor(exercise_5)3	0.066925	0.047105	1.421	0.155390	
as.factor(exercise_5)4	0.077338	0.046622	1.659	0.097155	.
as.factor(exercise_5)5	-0.009655	0.047863	-0.202	0.840136	
as.factor(exercise_5)6	0.026874	0.047995	0.560	0.575533	
as.factor(X20003.Saw.palmetto)1	0.397001	0.192243	2.065	0.038913	*
as.factor(X20003.Aspirin)1	0.016440	0.036419	0.451	0.651694	
as.factor(X20003.Statins)1	-0.018300	0.035497	-0.516	0.606177	
as.factor(X20003.Metformin)1	-0.118860	0.106718	-1.114	0.265372	
as.factor(X20003.Testosterone.supplement)1	0.383856	0.288376	1.331	0.183157	
as.factor(X20003.Anabolic.steroids)1	0.251818	0.517845	0.486	0.626768	
as.factor(X20003.5.alpha.reductase)1	0.168668	0.104608	1.612	0.106877	
as.factor(Fhx.breast.cancer..all.)1	0.129128	0.041526	3.110	0.001873	**
as.factor(X20002.Touchscreen...Prostatitis)1	-0.317283	0.417208	-0.760	0.446961	
as.factor(X6179.Touchscreen...Selenium)1	-0.049809	0.091806	-0.543	0.587446	
as.factor(X6179.Touchscreen...Selenium)55551	-0.432852	0.511127	-0.847	0.397075	
as.factor(X6155.Touchscreen.Vitamin.E.supplementation)1	0.208103	0.088504	2.351	0.018706	*
as.factor(X6155.Touchscreen...Vitamin.D)1	-0.027125	0.088134	-0.308	0.758259	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.013162	0.036075	0.365	0.715225	
as.factor(X1558.0.0.Alcohol.intake.frequency.)3	0.020715	0.037642	0.550	0.582102	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	0.011342	0.054620	0.208	0.835505	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	0.042222	0.057228	0.738	0.460648	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	-1.370302	1.009739	-1.357	0.174754	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.056073	0.065944	-0.850	0.395149	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	-0.235393	0.080736	-2.916	0.003550	**
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	-0.357137	0.308231	-1.159	0.246592	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	0.524335	0.743209	0.706	0.480498	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)6	2.586562	1.320100	1.959	0.050070	.
as.factor(Fhx.prostate.cancer..all.)1	0.603019	0.039524	15.257	< 2e-16	***
as.factor(Categorical.age.5.y)3	1.528919	0.217365	7.034	2.01e-12	***
as.factor(Categorical.age.5.y)4	2.596170	0.205959	12.605	< 2e-16	***
as.factor(Categorical.age.5.y)5	3.221345	0.203137	15.858	< 2e-16	***
as.factor(Categorical.age.5.y)6	3.881375	0.201767	19.237	< 2e-16	***
as.factor(Categorical.age.5.y)7	4.119677	0.202085	20.386	< 2e-16	***
as.factor(Categorical.age.5.y)8	4.339009	0.239672	18.104	< 2e-16	***

Supplementary table S2.11: Logistic regression covariates in the WBFM model predicting prostate cancer

	Estimate	Std. Error	Z	p	
(Intercept)	-6.851095	0.209269	-32.738	< 2e-16	***
as.factor(BMI_quantile)2	0.012821	0.041818	0.307	0.759162	
as.factor(BMI_quantile)3	0.007119	0.041940	0.170	0.865212	
as.factor(BMI_quantile)4	-0.019434	0.042535	-0.457	0.647747	
as.factor(BMI_quantile)5	-0.180517	0.045597	-3.959	7.53e-05	***
as.factor(X20160.0.0.Ever.smoked)0	0.050736	0.028924	1.754	0.079412	.
as.factor(X20160.0.0.Ever.smoked)8888888	0.345844	0.185268	1.867	0.061941	.
as.factor(X21000.0.0.Ethnic.background)1	0.361864	0.327588	1.105	0.269319	
as.factor(X21000.0.0.Ethnic.background)1002	0.085336	0.083210	1.026	0.305105	
as.factor(X21000.0.0.Ethnic.background)1003	-0.050827	0.091984	-0.553	0.580564	
as.factor(X21000.0.0.Ethnic.background)2001	0.203018	0.512693	0.396	0.692117	
as.factor(X21000.0.0.Ethnic.background)2002	0.925428	0.470805	1.966	0.049342	*
as.factor(X21000.0.0.Ethnic.background)2003	-0.418871	0.507653	-0.825	0.409308	
as.factor(X21000.0.0.Ethnic.background)2004	0.296610	0.343819	0.863	0.388306	
as.factor(X21000.0.0.Ethnic.background)3001	-0.670015	0.186782	-3.587	0.000334	***
as.factor(X21000.0.0.Ethnic.background)3002	-0.420128	0.326471	-1.287	0.198138	
as.factor(X21000.0.0.Ethnic.background)3003	0.358460	0.727801	0.493	0.622348	
as.factor(X21000.0.0.Ethnic.background)3004	-0.241672	0.273903	-0.882	0.377599	
as.factor(X21000.0.0.Ethnic.background)4001	1.047362	0.124559	8.409	< 2e-16	***
as.factor(X21000.0.0.Ethnic.background)4002	0.819383	0.156566	5.233	1.66e-07	***
as.factor(X21000.0.0.Ethnic.background)5	-0.467160	0.384573	-1.215	0.224461	
as.factor(X21000.0.0.Ethnic.background)55551	0.582829	0.522019	1.116	0.264212	
as.factor(X21000.0.0.Ethnic.background)55553	-0.083777	0.224543	-0.373	0.709074	
as.factor(X21000.0.0.Ethnic.background)6	-0.053211	0.188046	-0.283	0.777199	
as.factor(Town_5)2	0.015070	0.040066	0.376	0.706810	
as.factor(Town_5)3	0.003093	0.040580	0.076	0.939236	
as.factor(Town_5)4	-0.051070	0.042333	-1.206	0.227667	
as.factor(Town_5)5	-0.112092	0.045248	-2.477	0.013238	*
as.factor(Town_5)6	-0.099110	0.419368	-0.236	0.813175	
as.factor(X1369.0.0.Beef.intake)0	-0.060720	0.065008	-0.934	0.350285	
as.factor(X1369.0.0.Beef.intake)2	0.003275	0.032154	0.102	0.918879	
as.factor(X1369.0.0.Beef.intake)3	-0.014620	0.045187	-0.324	0.746282	
as.factor(X1369.0.0.Beef.intake)4	0.055389	0.277197	0.200	0.841623	
as.factor(X1369.0.0.Beef.intake)5	-0.739639	0.727403	-1.017	0.309239	
as.factor(X1369.0.0.Beef.intake)55551	-0.444790	0.285644	-1.557	0.119436	
as.factor(X1369.0.0.Beef.intake)55553	-0.099187	1.071525	-0.093	0.926248	
as.factor(X1379.0.0.Lamb.mutton.intake)0	-0.105204	0.048097	-2.187	0.028719	*
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.027407	0.033554	0.817	0.414043	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.052924	0.074907	0.707	0.479857	
as.factor(X1379.0.0.Lamb.mutton.intake)4	-0.628854	0.722329	-0.871	0.383977	
as.factor(X1379.0.0.Lamb.mutton.intake)5	-0.732446	1.027067	-0.713	0.475757	
as.factor(X1379.0.0.Lamb.mutton.intake)55551	0.183504	0.177152	1.036	0.300270	
as.factor(X1379.0.0.Lamb.mutton.intake)55553	-0.668175	1.065823	-0.627	0.530719	
as.factor(X1329.0.0.Oily.fish.intake)0	-0.061290	0.052684	-1.163	0.244684	
as.factor(X1329.0.0.Oily.fish.intake)1	0.001637	0.031990	0.051	0.959189	
as.factor(X1329.0.0.Oily.fish.intake)3	0.004118	0.037676	0.109	0.912962	
as.factor(X1329.0.0.Oily.fish.intake)4	-0.016385	0.142091	-0.115	0.908199	
as.factor(X1329.0.0.Oily.fish.intake)5	-0.002792	0.244203	-0.011	0.990879	
as.factor(X1329.0.0.Oily.fish.intake)55551	-0.149919	0.194456	-0.771	0.440728	
as.factor(X1329.0.0.Oily.fish.intake)55553	-0.363613	1.030700	-0.353	0.724252	
as.factor(exercise_5)2	0.019309	0.047167	0.409	0.682263	
as.factor(exercise_5)3	0.069809	0.047069	1.483	0.138048	
as.factor(exercise_5)4	0.082009	0.046546	1.762	0.078088	.
as.factor(exercise_5)5	-0.002689	0.047721	-0.056	0.955064	
as.factor(exercise_5)6	0.030348	0.047968	0.633	0.526949	
as.factor(X20003.Saw.palmetto)1	0.394580	0.192224	2.053	0.040101	*
as.factor(X20003.Aspirin)1	0.017327	0.036427	0.476	0.634323	
as.factor(X20003.Statins)1	-0.019194	0.035507	-0.541	0.588805	
as.factor(X20003.Metformin)1	-0.117024	0.106723	-1.097	0.272851	
as.factor(X20003.Testosterone.supplement)1	0.388457	0.288366	1.347	0.177948	
as.factor(X20003.Anabolic.steroids)1	0.265150	0.517729	0.512	0.608552	
as.factor(X20003.5.alpha.reductase)1	0.169799	0.104605	1.623	0.104538	
as.factor(Fhx.breast.cancer..all.)1	0.128531	0.041525	3.095	0.001966	**
as.factor(X20002.Touchscreen...Prostatitis)1	-0.323714	0.417198	-0.776	0.437794	
as.factor(X6179.Touchscreen...Selenium)1	-0.050551	0.091784	-0.551	0.581794	
as.factor(X6179.Touchscreen...Selenium)55551	-0.432084	0.511127	-0.845	0.397912	
as.factor(X6155.Touchscreen.Vitamin.E.supplementation)1	0.210649	0.088508	2.380	0.017312	*
as.factor(X6155.Touchscreen...Vitamin.D)1	-0.027264	0.088144	-0.309	0.757085	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.012220	0.036082	0.339	0.734861	
as.factor(X1558.0.0.Alcohol.intake.frequency.)3	0.021413	0.037648	0.569	0.569522	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	0.012044	0.054622	0.220	0.825484	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	0.044689	0.057237	0.781	0.434937	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	-1.374003	1.009908	-1.361	0.173664	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.054221	0.065942	-0.822	0.410937	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	-0.233922	0.080760	-2.897	0.003773	**
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	-0.352878	0.308241	-1.145	0.252288	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	0.516795	0.743269	0.695	0.486868	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)6	2.600826	1.327380	1.959	0.050070	.
as.factor(Fhx.prostate.cancer..all.)1	0.603077	0.039523	15.259	< 2e-16	***
as.factor(Categorical.age.5.y)3	1.527671	0.217364	7.028	2.09e-12	***
as.factor(Categorical.age.5.y)4	2.594125	0.205955	12.596	< 2e-16	***
as.factor(Categorical.age.5.y)5	3.216234	0.203128	15.834	< 2e-16	***
as.factor(Categorical.age.5.y)6	3.874336	0.201754	19.203	< 2e-16	***
as.factor(Categorical.age.5.y)7	4.110400	0.202073	20.341	< 2e-16	***
as.factor(Categorical.age.5.y)8	4.326775	0.239662	18.054	< 2e-16	***

Supplementary table S2.12: Logistic regression covariates in the BMI model predicting prostate cancer

	Estimate	Std. Error	Z	p	
(Intercept)	-6.4614685	0.2320403	-27.846	< 2e-16	***
as.factor(Whole_body_fat_free_mass_female_quantile)2	0.0005705	0.0848282	0.007	0.994634	
as.factor(Whole_body_fat_free_mass_female_quantile)3	0.0856030	0.0834404	1.026	0.304930	
as.factor(Whole_body_fat_free_mass_female_quantile)4	0.1634374	0.0839967	1.946	0.051683	.
as.factor(Whole_body_fat_free_mass_female_quantile)5	0.1683494	0.0869245	1.937	0.052778	.
as.factor(Town_5)2	-0.0456859	0.0824845	-0.554	0.579666	
as.factor(Town_5)3	-0.0756867	0.0837162	-0.904	0.365950	
as.factor(Town_5)4	-0.0049993	0.0834054	-0.060	0.952204	
as.factor(Town_5)5	-0.0564751	0.0873070	-0.647	0.517725	
as.factor(Town_5)8888888	0.3117698	0.7133648	0.437	0.662081	
as.factor(X21000.0.0.Ethnic.background)1	0.8245280	0.5849217	1.410	0.158647	
as.factor(X21000.0.0.Ethnic.background)1002	0.2357491	0.1528997	1.542	0.123109	
as.factor(X21000.0.0.Ethnic.background)1003	-0.2262588	0.1638377	-1.381	0.167281	
as.factor(X21000.0.0.Ethnic.background)2001	0.7168000	0.5829265	1.230	0.218825	
as.factor(X21000.0.0.Ethnic.background)2002	-0.3067425	1.0043575	-0.305	0.760053	
as.factor(X21000.0.0.Ethnic.background)2003	-0.7502520	1.0025059	-0.748	0.454233	
as.factor(X21000.0.0.Ethnic.background)2004	0.4067867	0.5041726	0.807	0.419759	
as.factor(X21000.0.0.Ethnic.background)3001	-0.3615019	0.3281527	-1.102	0.270624	
as.factor(X21000.0.0.Ethnic.background)3002	-0.0595392	0.5925846	-0.100	0.919968	
as.factor(X21000.0.0.Ethnic.background)3004	0.0056418	0.5062331	0.011	0.991108	
as.factor(X21000.0.0.Ethnic.background)4001	-0.2800635	0.3223045	-0.869	0.384880	
as.factor(X21000.0.0.Ethnic.background)4002	-0.3919791	0.4546227	-0.862	0.388573	
as.factor(X21000.0.0.Ethnic.background)5	0.1481771	0.4536898	0.327	0.743967	
as.factor(X21000.0.0.Ethnic.background)55553	-0.1507583	0.5820044	-0.259	0.795610	
as.factor(X21000.0.0.Ethnic.background)6	-0.0696858	0.3068724	-0.227	0.820358	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.0834532	0.0827620	1.008	0.313286	
as.factor(X1558.0.0.Alcohol.intake.frequency.)2	-0.1255293	0.0831749	-1.509	0.131242	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	0.0107460	0.0921192	0.117	0.907135	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.0147838	0.0879614	-0.168	0.866527	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	0.6316973	0.7216157	0.875	0.381360	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	0.1457270	0.0995128	1.464	0.143083	
as.factor(X20160.0.0.Ever.smoked)0	-0.0543638	0.0553801	-0.982	0.326273	
as.factor(X20160.0.0.Ever.smoked)8888888	-0.6765640	0.5821711	-1.162	0.245179	
as.factor(Fhx.bowel.cancer..all.)1	0.1652379	0.0781710	2.114	0.034532	*
as.factor(X1369.0.0.Beef.intake)0	-0.0166108	0.1008150	-0.165	0.869129	
as.factor(X1369.0.0.Beef.intake)2	0.0813779	0.0655116	1.242	0.214166	
as.factor(X1369.0.0.Beef.intake)3	0.0781592	0.0924891	0.845	0.398075	
as.factor(X1369.0.0.Beef.intake)4	-0.3660465	1.0052354	-0.364	0.715753	
as.factor(X1369.0.0.Beef.intake)55551	0.1182895	0.4925208	0.240	0.810197	
as.factor(X1379.0.0.Lamb.mutton.intake)0	0.0300313	0.0822687	0.365	0.715082	
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.0644119	0.0699279	0.921	0.356988	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.0843294	0.1725871	0.489	0.625111	
as.factor(X1379.0.0.Lamb.mutton.intake)5	0.9599165	1.0160562	0.945	0.344788	
as.factor(X1379.0.0.Lamb.mutton.intake)55551	-0.3297506	0.4490051	-0.734	0.462703	
as.factor(exercise_5)1	0.0319841	0.0852727	0.375	0.707601	
as.factor(exercise_5)2	0.0027810	0.0858788	0.032	0.974167	
as.factor(exercise_5)3	-0.0722803	0.0878128	-0.823	0.410441	
as.factor(exercise_5)4	-0.1501540	0.0893132	-1.681	0.092723	.
as.factor(exercise_5)5	-0.1762407	0.0890501	-1.979	0.047803	*
as.factor(X20003.Statins)1	0.0189059	0.0835490	0.226	0.820979	
as.factor(X20003.Metformin)1	0.2300190	0.2484158	0.926	0.354475	
as.factor(X20003.Aspirin)1	0.0187625	0.0879868	0.213	0.831138	
as.factor(X6155.Touchscreen...Vitamin.D)1	0.0259403	0.1155874	0.224	0.822429	
as.factor(X6155.Touchscreen...Vitamin.D)55551	-0.1813348	0.4777414	-0.380	0.704267	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	-0.0777220	0.1990414	-0.390	0.696180	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	0.0749568	0.5818803	0.129	0.897501	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	1.4201324	0.8113007	1.750	0.080043	.
as.factor(X2724.0.0.Had.menopause)0	-0.0515045	0.1277100	-0.403	0.686733	
as.factor(X2724.0.0.Had.menopause)2	-0.0617461	0.0845802	-0.730	0.465372	
as.factor(X2724.0.0.Had.menopause)3	0.1128994	0.1592265	0.709	0.478294	
as.factor(X2724.0.0.Had.menopause)55553	-1.4215672	1.1223779	-1.267	0.205310	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)1	0.0289871	0.0584814	0.496	0.620132	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55551	-0.1743087	0.5818504	-0.300	0.764500	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55553	0.7949982	0.8135228	0.977	0.328456	
as.factor(Categorical.age.5.y)3	0.7096003	0.1938746	3.660	0.000252	***
as.factor(Categorical.age.5.y)4	1.0884434	0.1975708	5.509	3.61e-08	***
as.factor(Categorical.age.5.y)5	1.3310127	0.2093237	6.359	2.04e-10	***
as.factor(Categorical.age.5.y)6	1.6322822	0.2085011	7.829	4.93e-15	***
as.factor(Categorical.age.5.y)7	2.0032566	0.2092560	9.573	< 2e-16	***
as.factor(Categorical.age.5.y)8	2.5359648	0.3187269	7.957	1.77e-15	***

Supplementary table S2.13: Logistic regression covariates in the WBFFM model predicting colorectal cancer in women

	Estimate	Std. Error	Z	p	
(Intercept)	-6.424101	0.231772	27.717	< 2e-16	***
as.factor(Whole_body_fat_mass_female_quantile)2	0.041276	0.089493	0.461	0.644643	
as.factor(Whole_body_fat_mass_female_quantile)3	0.157679	0.087358	1.805	0.071079	.
as.factor(Whole_body_fat_mass_female_quantile)4	0.019462	0.089678	0.217	0.828195	
as.factor(Whole_body_fat_mass_female_quantile)5	0.078177	0.091361	0.856	0.392171	
as.factor(Town_5)2	-0.045663	0.082487	-0.554	0.579869	
as.factor(Town_5)3	-0.075264	0.083730	-0.899	0.368712	
as.factor(Town_5)4	-0.002487	0.083440	-0.030	0.976226	
as.factor(Town_5)5	-0.053232	0.087412	-0.609	0.542541	
as.factor(Town_5)8888888	0.302353	0.713371	0.424	0.671685	
as.factor(X21000.0.0.Ethnic.background)1	0.833946	0.584911	1.426	0.153936	
as.factor(X21000.0.0.Ethnic.background)1002	0.225158	0.152857	1.473	0.140751	
as.factor(X21000.0.0.Ethnic.background)1003	-0.217816	0.163873	-1.329	0.183791	
as.factor(X21000.0.0.Ethnic.background)2001	0.723310	0.582934	1.241	0.214676	
as.factor(X21000.0.0.Ethnic.background)2002	-0.305656	1.004343	-0.304	0.760874	
as.factor(X21000.0.0.Ethnic.background)2003	-0.765009	1.002455	-0.763	0.445383	
as.factor(X21000.0.0.Ethnic.background)2004	0.395469	0.504121	0.784	0.432763	
as.factor(X21000.0.0.Ethnic.background)3001	-0.418091	0.327430	-1.277	0.201642	
as.factor(X21000.0.0.Ethnic.background)3002	-0.097735	0.592504	-0.165	0.868981	
as.factor(X21000.0.0.Ethnic.background)3004	-0.045432	0.505768	-0.090	0.928424	
as.factor(X21000.0.0.Ethnic.background)4001	-0.267326	0.322340	-0.829	0.406919	
as.factor(X21000.0.0.Ethnic.background)4002	-0.376471	0.454771	-0.828	0.407770	
as.factor(X21000.0.0.Ethnic.background)5	0.110615	0.453850	0.244	0.807442	
as.factor(X21000.0.0.Ethnic.background)55553	-0.150188	0.581963	-0.258	0.796351	
as.factor(X21000.0.0.Ethnic.background)6	-0.091563	0.306747	-0.298	0.765323	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.080653	0.082952	0.972	0.330913	
as.factor(X1558.0.0.Alcohol.intake.frequency.)2	-0.128763	0.083224	-1.547	0.121816	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	0.015454	0.092140	0.168	0.866803	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.008010	0.088029	-0.091	0.927497	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	0.630352	0.721485	0.874	0.382289	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	0.151061	0.099559	1.517	0.129193	
as.factor(X20160.0.0.Ever.smoked)0	-0.059053	0.055365	-1.067	0.286140	
as.factor(X20160.0.0.Ever.smoked)8888888	-0.681030	0.582092	-1.170	0.242013	
as.factor(Fhx.bowel.cancer..all.)1	0.166362	0.078169	2.128	0.033318	*
as.factor(X1369.0.0.Beef.intake)0	-0.017095	0.100870	-0.169	0.865423	
as.factor(X1369.0.0.Beef.intake)2	0.079844	0.065533	1.218	0.223082	
as.factor(X1369.0.0.Beef.intake)3	0.078238	0.092538	0.845	0.397850	
as.factor(X1369.0.0.Beef.intake)4	-0.373271	1.005238	-0.371	0.710395	
as.factor(X1369.0.0.Beef.intake)55551	0.119775	0.492174	0.243	0.807727	
as.factor(X1379.0.0.Lamb.mutton.intake)0	0.027088	0.082222	0.329	0.741815	
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.067204	0.069921	0.961	0.336485	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.087957	0.172580	0.510	0.610291	
as.factor(X1379.0.0.Lamb.mutton.intake)5	0.963406	1.015951	0.948	0.342987	
as.factor(X1379.0.0.Lamb.mutton.intake)55551	-0.328215	0.448730	-0.731	0.464516	
as.factor(exercise_5)1	0.035062	0.085341	0.411	0.681188	
as.factor(exercise_5)2	0.003124	0.085883	0.036	0.970985	
as.factor(exercise_5)3	-0.073899	0.087862	-0.841	0.400301	
as.factor(exercise_5)4	-0.150081	0.089443	-1.678	0.093355	.
as.factor(exercise_5)5	-0.176076	0.089286	-1.972	0.048604	*
as.factor(X20003.Statins)1	0.021079	0.083809	0.252	0.801422	
as.factor(X20003.Metformin)1	0.239374	0.248585	0.963	0.335575	
as.factor(X6155.Touchscreen...Vitamin.D)1	0.022086	0.115650	0.191	0.848546	
as.factor(X6155.Touchscreen...Vitamin.D)55551	-0.188381	0.478001	-0.394	0.693506	
as.factor(X20003.Aspirin)1	0.023215	0.088050	0.264	0.792046	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	-0.053922	0.199101	-0.271	0.786523	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	0.086186	0.581872	0.148	0.882249	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	1.433538	0.812155	1.765	0.077546	.
as.factor(X2724.0.0.Had.menopause)0	-0.039919	0.127589	-0.313	0.754376	
as.factor(X2724.0.0.Had.menopause)2	-0.056638	0.084644	-0.669	0.503413	
as.factor(X2724.0.0.Had.menopause)3	0.121246	0.159176	0.762	0.446234	
as.factor(X2724.0.0.Had.menopause)55553	-1.409300	1.123995	-1.254	0.209903	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)1	0.025856	0.058483	0.442	0.658403	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55551	-0.166839	0.581874	-0.287	0.774322	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55553	0.797140	0.814073	0.979	0.327481	
as.factor(Categorical.age.5.y)3	0.705544	0.193882	3.639	0.000274	***
as.factor(Categorical.age.5.y)4	1.079898	0.197592	5.465	4.62e-08	***
as.factor(Categorical.age.5.y)5	1.315358	0.209313	6.284	3.30e-10	***
as.factor(Categorical.age.5.y)6	1.608728	0.208443	7.718	1.18e-14	***
as.factor(Categorical.age.5.y)7	1.970996	0.209045	9.429	< 2e-16	***
as.factor(Categorical.age.5.y)8	2.494535	0.318549	7.831	4.84e-15	***

Supplementary table S2.14: Logistic regression covariates in the WBFM model predicting colorectal cancer in women

	Estimate	Std. Error	Z	p	
(Intercept)	-6.4390828	0.2315718	27.806	< 2e-16	***
as.factor(BMI_quantile)2	0.1157832	0.0888572	1.303	0.19257	
as.factor(BMI_quantile)3	0.1825858	0.0869597	2.100	0.03576	*
as.factor(BMI_quantile)4	-0.0101914	0.0906344	-0.112	0.91047	
as.factor(BMI_quantile)5	0.0664872	0.0921054	0.722	0.47038	
as.factor(Town_5)2	-0.0446758	0.0824905	-0.542	0.58810	
as.factor(Town_5)3	-0.0733547	0.0837420	-0.876	0.38105	
as.factor(Town_5)4	0.0004611	0.0834666	0.006	0.99559	
as.factor(Town_5)5	-0.0484186	0.0875045	-0.553	0.58004	
as.factor(Town_5)8888888	0.3031694	0.7133863	0.425	0.67086	
as.factor(X21000.0.0.Ethnic.background)1	0.8362028	0.5849207	1.430	0.15283	
as.factor(X21000.0.0.Ethnic.background)1002	0.2248524	0.1528536	1.471	0.14128	
as.factor(X21000.0.0.Ethnic.background)1003	-0.2189576	0.1638632	-1.336	0.18148	
as.factor(X21000.0.0.Ethnic.background)2001	0.7276834	0.5829389	1.248	0.21192	
as.factor(X21000.0.0.Ethnic.background)2002	-0.2990227	1.0043337	-0.298	0.76591	
as.factor(X21000.0.0.Ethnic.background)2003	-0.7729594	1.0024618	-0.771	0.44067	
as.factor(X21000.0.0.Ethnic.background)2004	0.3952274	0.5041199	0.784	0.43304	
as.factor(X21000.0.0.Ethnic.background)3001	-0.4206273	0.3274170	-1.285	0.19890	
as.factor(X21000.0.0.Ethnic.background)3002	-0.0900705	0.5924300	-0.152	0.87916	
as.factor(X21000.0.0.Ethnic.background)3004	-0.0570985	0.5057779	-0.113	0.91012	
as.factor(X21000.0.0.Ethnic.background)4001	-0.2614109	0.3223835	-0.811	0.41744	
as.factor(X21000.0.0.Ethnic.background)4002	-0.3687846	0.4548729	-0.811	0.41751	
as.factor(X21000.0.0.Ethnic.background)5	0.0926472	0.4533460	0.204	0.83807	
as.factor(X21000.0.0.Ethnic.background)55553	-0.1530001	0.5820096	-0.263	0.79264	
as.factor(X21000.0.0.Ethnic.background)6	-0.0993683	0.3067122	-0.324	0.74595	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.0785305	0.0830329	0.946	0.34426	
as.factor(X1558.0.0.Alcohol.intake.frequency.)2	-0.1307740	0.0832497	-1.571	0.11621	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	0.0182633	0.0921503	0.198	0.84290	
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.0030909	0.0880873	-0.035	0.97201	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	0.6272784	0.7216964	0.869	0.38475	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	0.1564309	0.0996197	1.570	0.11635	
as.factor(X20160.0.0.Ever.smoked)0	-0.0597203	0.0553554	-1.079	0.28065	
as.factor(X20160.0.0.Ever.smoked)8888888	-0.6793754	0.5820638	-1.167	0.24314	
as.factor(Fhx.bowel.cancer..all.)1	0.1653090	0.0781705	2.115	0.03445	*
as.factor(X1369.0.0.Beef.intake)0	-0.0191145	0.1008862	-0.189	0.84973	
as.factor(X1369.0.0.Beef.intake)2	0.0815573	0.0655332	1.245	0.21331	
as.factor(X1369.0.0.Beef.intake)3	0.0804671	0.0925494	0.869	0.38460	
as.factor(X1369.0.0.Beef.intake)4	-0.3678528	1.0052353	-0.366	0.71441	
as.factor(X1369.0.0.Beef.intake)55551	0.1213131	0.4918418	0.247	0.80518	
as.factor(X1379.0.0.Lamb.mutton.intake)0	0.0264519	0.0822298	0.322	0.74769	
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.0668848	0.0699148	0.957	0.33874	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.0866489	0.1725899	0.502	0.61563	
as.factor(X1379.0.0.Lamb.mutton.intake)5	0.9396902	1.0161788	0.925	0.35511	
as.factor(X1379.0.0.Lamb.mutton.intake)55551	-0.3325630	0.4484450	-0.742	0.45833	
as.factor(exercise_5)1	0.0392644	0.0853001	0.460	0.64529	
as.factor(exercise_5)2	0.0024487	0.0858897	0.029	0.97726	
as.factor(exercise_5)3	-0.0753203	0.0878783	-0.857	0.39139	
as.factor(exercise_5)4	-0.1539663	0.0894542	-1.721	0.08522	.
as.factor(exercise_5)5	-0.1817546	0.0892490	-2.036	0.04170	*
as.factor(X20003.Statins)1	0.0254298	0.0839459	0.303	0.76194	
as.factor(X20003.Metformin)1	0.2415488	0.2486460	0.971	0.33132	
as.factor(X6155.Touchscreen...Vitamin.D)1	0.0199632	0.1156657	0.173	0.86297	
as.factor(X6155.Touchscreen...Vitamin.D)55551	-0.1860091	0.4780709	-0.389	0.69722	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	-0.0471897	0.1992971	-0.237	0.81283	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	0.0882652	0.5819409	0.152	0.87944	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	1.4685774	0.8135209	1.805	0.07104	.
as.factor(X20003.Aspirin)1	0.0268459	0.0880793	0.305	0.76052	
as.factor(X2724.0.0.Had.menopause)0	-0.0420294	0.1275758	-0.329	0.74182	
as.factor(X2724.0.0.Had.menopause)2	-0.0529363	0.0846437	-0.625	0.53171	
as.factor(X2724.0.0.Had.menopause)3	0.1211865	0.1591766	0.761	0.44646	
as.factor(X2724.0.0.Had.menopause)55553	-1.4203736	1.1270155	-1.260	0.20756	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)1	0.0247944	0.0584885	0.424	0.67163	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55551	-0.1587643	0.5819191	-0.273	0.78498	
as.factor(X2814.0.0.Ever.used.hormone.replacement.therapy..HRT.)55553	0.7830858	0.8152571	0.961	0.33678	
as.factor(Categorical.age.5.y)3	0.7043635	0.1938867	3.633	0.00028	***
as.factor(Categorical.age.5.y)4	1.0797663	0.1976036	5.464	4.65e-08	***
as.factor(Categorical.age.5.y)5	1.3150215	0.2093213	6.282	3.34e-10	***
as.factor(Categorical.age.5.y)6	1.6091661	0.2084621	7.719	1.17e-14	***
as.factor(Categorical.age.5.y)7	1.9709943	0.2090564	9.428	< 2e-16	***
as.factor(Categorical.age.5.y)8	2.4950274	0.3185832	7.832	4.82e-15	***

Supplementary table S2.15: Logistic regression covariates in the BMI model predicting colorectal cancer in women

	Estimate	Std. Error	Z	p	
(Intercept)	-6.490676	0.196066	33.105	< 2e-16	***
as.factor(Whole_body_fat_free_mass_male_quantile)2	0.098773	0.070525	1.401	0.161352	
as.factor(Whole_body_fat_free_mass_male_quantile)3	0.180281	0.070512	2.557	0.010566	*
as.factor(Whole_body_fat_free_mass_male_quantile)4	0.092462	0.073537	1.257	0.208626	
as.factor(Whole_body_fat_free_mass_male_quantile)5	0.230811	0.074688	3.090	0.001999	**
as.factor(Town_5)2	0.026046	0.068320	0.381	0.703029	
as.factor(Town_5)3	-0.070186	0.070573	-0.995	0.319973	
as.factor(Town_5)4	-0.051160	0.072014	-0.710	0.477443	
as.factor(Town_5)5	0.033154	0.073600	0.450	0.652379	
as.factor(Town_5)8888888	-0.730215	1.004629	-0.727	0.467318	
as.factor(X21000.0.0.Ethnic.background)1	0.138076	0.584529	0.236	0.813264	
as.factor(X21000.0.0.Ethnic.background)1002	0.420805	0.116823	3.602	0.000316	***
as.factor(X21000.0.0.Ethnic.background)1003	-0.126643	0.157623	-0.803	0.421711	
as.factor(X21000.0.0.Ethnic.background)2	2.121108	1.059040	2.003	0.045192	*
as.factor(X21000.0.0.Ethnic.background)2001	-0.303936	1.005913	-0.302	0.762538	
as.factor(X21000.0.0.Ethnic.background)2002	0.919112	0.718651	1.279	0.200918	
as.factor(X21000.0.0.Ethnic.background)2003	-0.822104	1.003611	-0.819	0.412703	
as.factor(X21000.0.0.Ethnic.background)2004	1.075195	0.363142	2.961	0.003068	**
as.factor(X21000.0.0.Ethnic.background)3001	-0.435493	0.289138	-1.506	0.132022	
as.factor(X21000.0.0.Ethnic.background)3002	-1.052373	0.717548	-1.467	0.142479	
as.factor(X21000.0.0.Ethnic.background)3004	-0.742182	0.582270	-1.275	0.202438	
as.factor(X21000.0.0.Ethnic.background)4001	-0.427396	0.382418	-1.118	0.263731	
as.factor(X21000.0.0.Ethnic.background)4002	-0.776374	0.506004	-1.534	0.124950	
as.factor(X21000.0.0.Ethnic.background)5	-0.532994	0.711073	-0.750	0.453518	
as.factor(X21000.0.0.Ethnic.background)55553	0.339947	0.306934	1.108	0.268052	
as.factor(X21000.0.0.Ethnic.background)6	-0.660374	0.412015	-1.603	0.108981	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.100869	0.060070	1.679	0.093116	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)3	0.011069	0.063343	0.175	0.861274	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	-0.159168	0.096629	-1.647	0.099517	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.120704	0.101721	-1.187	0.235376	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	0.238477	0.722428	0.330	0.741321	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.326477	0.122942	-2.656	0.007918	**
as.factor(X20160.0.0.Ever.smoked)0	-0.103367	0.050614	-2.042	0.041126	*
as.factor(X20160.0.0.Ever.smoked)8888888	0.023439	0.340404	0.069	0.945104	
as.factor(Fhx.bowel.cancer..all.)1	0.297136	0.061855	4.804	1.56e-06	***
as.factor(X1369.0.0.Beef.intake)0	-0.098646	0.114863	-0.859	0.390444	
as.factor(X1369.0.0.Beef.intake)2	0.022389	0.054694	0.409	0.682287	
as.factor(X1369.0.0.Beef.intake)3	0.144152	0.072331	1.993	0.046267	*
as.factor(X1369.0.0.Beef.intake)4	-0.365460	0.583042	-0.627	0.530780	
as.factor(X1369.0.0.Beef.intake)5	-0.033622	0.811455	-0.041	0.966950	
as.factor(X1369.0.0.Beef.intake)55551	-0.010709	0.448280	-0.024	0.980942	
as.factor(X1379.0.0.Lamb.mutton.intake)0	-0.076767	0.080263	-0.956	0.338849	
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.023434	0.056597	0.414	0.678836	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.091350	0.121737	0.750	0.453021	
as.factor(X1379.0.0.Lamb.mutton.intake)5	1.343161	0.671697	2.000	0.045538	*
as.factor(X1379.0.0.Lamb.mutton.intake)55551	-0.399559	0.365568	-1.093	0.274403	
as.factor(X1379.0.0.Lamb.mutton.intake)55553	1.006766	1.023258	0.984	0.325173	
as.factor(exercise_5)1	0.049640	0.078090	0.636	0.524989	
as.factor(exercise_5)2	-0.032743	0.078947	-0.415	0.678330	
as.factor(exercise_5)3	-0.036677	0.079926	-0.459	0.646316	
as.factor(exercise_5)4	-0.003627	0.078490	-0.046	0.963138	
as.factor(exercise_5)5	-0.056459	0.079815	-0.707	0.479332	
as.factor(X20003.Statins)1	0.072455	0.059557	1.217	0.223768	
as.factor(X20003.Metformin)1	0.260236	0.147439	1.765	0.077557	.
as.factor(X6155.Touchscreen...Vitamin.D)1	-0.153425	0.152188	-1.008	0.313393	
as.factor(X6155.Touchscreen...Vitamin.D)55551	0.317075	0.237687	1.334	0.182203	
as.factor(X20003.Aspirin)1	-0.103513	0.062084	-1.667	0.095452	.
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	0.059922	0.121968	0.491	0.623219	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	-0.339734	0.504507	-0.673	0.500695	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	-0.020398	1.024108	-0.020	0.984109	
as.factor(Categorical.age.5.y)3	0.552590	0.202199	2.733	0.006278	**
as.factor(Categorical.age.5.y)4	1.399819	0.181606	7.708	1.28e-14	***
as.factor(Categorical.age.5.y)5	1.751906	0.176540	9.924	< 2e-16	***
as.factor(Categorical.age.5.y)6	2.141598	0.173255	12.361	< 2e-16	***
as.factor(Categorical.age.5.y)7	2.574910	0.173725	14.822	< 2e-16	***
as.factor(Categorical.age.5.y)8	2.781041	0.279303	9.957	< 2e-16	***

Supplementary table S2.16: Logistic regression covariates in the WBFFM model predicting colorectal cancer in men

	Estimate	Std. Error	Z	p	
(Intercept)	-6.472089	0.195067	-33.179	< 2e-16	***
as.factor(Whole_body_fat_mass_male_quantile)2	0.024654	0.080018	0.308	0.758000	
as.factor(Whole_body_fat_mass_male_quantile)3	0.216727	0.076700	2.826	0.004719	**
as.factor(Whole_body_fat_mass_male_quantile)4	0.174165	0.076869	2.266	0.023468	*
as.factor(Whole_body_fat_mass_male_quantile)5	0.305702	0.076514	3.995	6.46e-05	***
as.factor(Town_5)2	0.023768	0.068322	0.348	0.727926	
as.factor(Town_5)3	-0.077337	0.070581	-1.096	0.273198	
as.factor(Town_5)4	-0.061025	0.072041	-0.847	0.396945	
as.factor(Town_5)5	0.014654	0.073561	0.199	0.842099	
as.factor(Town_5)8888888	-0.752659	1.004639	-0.749	0.453747	
as.factor(X21000.0.0.Ethnic.background)1	0.132423	0.584536	0.227	0.820778	
as.factor(X21000.0.0.Ethnic.background)1002	0.411904	0.116825	3.526	0.000422	***
as.factor(X21000.0.0.Ethnic.background)1003	-0.116366	0.157627	-0.738	0.460370	
as.factor(X21000.0.0.Ethnic.background)2	2.084555	1.059809	1.967	0.049193	*
as.factor(X21000.0.0.Ethnic.background)2001	-0.286667	1.005964	-0.285	0.775669	
as.factor(X21000.0.0.Ethnic.background)2002	0.909791	0.718582	1.266	0.205480	
as.factor(X21000.0.0.Ethnic.background)2003	-0.825036	1.003563	-0.822	0.411016	
as.factor(X21000.0.0.Ethnic.background)2004	1.083455	0.363186	2.983	0.002853	**
as.factor(X21000.0.0.Ethnic.background)3001	-0.482609	0.288487	-1.673	0.094348	.
as.factor(X21000.0.0.Ethnic.background)3002	-1.084836	0.717479	-1.512	0.130531	
as.factor(X21000.0.0.Ethnic.background)3004	-0.771299	0.581846	-1.326	0.184970	
as.factor(X21000.0.0.Ethnic.background)4001	-0.414710	0.382411	-1.084	0.278161	
as.factor(X21000.0.0.Ethnic.background)4002	-0.792947	0.505899	-1.567	0.117021	
as.factor(X21000.0.0.Ethnic.background)5	-0.504997	0.711048	-0.710	0.477571	
as.factor(X21000.0.0.Ethnic.background)55553	0.333895	0.306953	1.088	0.276695	
as.factor(X21000.0.0.Ethnic.background)6	-0.671194	0.411911	-1.629	0.103214	
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.105116	0.060080	1.750	0.080186	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)3	0.006681	0.063367	0.105	0.916029	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	-0.163454	0.096659	-1.691	0.090830	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.126020	0.101760	-1.238	0.215566	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	0.222181	0.722165	0.308	0.758342	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.332024	0.122960	-2.700	0.006928	**
as.factor(X20160.0.0.Ever.smoked)0	-0.094616	0.050676	-1.867	0.061893	.
as.factor(X20160.0.0.Ever.smoked)8888888	0.016135	0.340307	0.047	0.962184	
as.factor(Fhx.bowel.cancer..all.)1	0.295485	0.061857	4.777	1.78e-06	***
as.factor(X1369.0.0.Beef.intake)0	-0.084101	0.114955	-0.732	0.464413	
as.factor(X1369.0.0.Beef.intake)2	0.013088	0.054746	0.239	0.811056	
as.factor(X1369.0.0.Beef.intake)3	0.127526	0.072448	1.760	0.078369	.
as.factor(X1369.0.0.Beef.intake)4	-0.391047	0.583112	-0.671	0.502462	
as.factor(X1369.0.0.Beef.intake)5	-0.063531	0.815403	-0.078	0.937897	
as.factor(X1369.0.0.Beef.intake)55551	-0.024366	0.447977	-0.054	0.956624	
as.factor(X1379.0.0.Lamb.mutton.intake)0	-0.079605	0.080186	-0.993	0.320831	
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.021852	0.056602	0.386	0.699448	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.090449	0.121723	0.743	0.457439	
as.factor(X1379.0.0.Lamb.mutton.intake)5	1.360562	0.674735	2.016	0.043754	*
as.factor(X1379.0.0.Lamb.mutton.intake)55551	-0.411706	0.365291	-1.127	0.259716	
as.factor(X1379.0.0.Lamb.mutton.intake)55553	1.019728	1.023377	0.996	0.319039	
as.factor(exercise_5)1	0.039799	0.078151	0.509	0.610574	
as.factor(exercise_5)2	-0.028450	0.078939	-0.360	0.718540	
as.factor(exercise_5)3	-0.024036	0.079954	-0.301	0.763706	
as.factor(exercise_5)4	0.014980	0.078569	0.191	0.848794	
as.factor(exercise_5)5	-0.033759	0.079968	-0.422	0.672909	
as.factor(X20003.Statins)1	0.049930	0.059742	0.836	0.403290	
as.factor(X20003.Metformin)1	0.250057	0.147419	1.696	0.089842	.
as.factor(X6155.Touchscreen...Vitamin.D)1	-0.147377	0.152207	-0.968	0.332912	
as.factor(X6155.Touchscreen...Vitamin.D)55551	0.305155	0.237739	1.284	0.199293	
as.factor(X20003.Aspirin)1	-0.109302	0.062057	-1.761	0.078182	.
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	0.049670	0.122033	0.407	0.683994	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	-0.357602	0.504499	-0.709	0.478433	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	-0.037304	1.023904	-0.036	0.970937	
as.factor(Categorical.age.5.y)3	0.539977	0.202197	2.671	0.007573	**
as.factor(Categorical.age.5.y)4	1.374010	0.181580	7.567	3.82e-14	***
as.factor(Categorical.age.5.y)5	1.715759	0.176419	9.725	< 2e-16	***
as.factor(Categorical.age.5.y)6	2.092674	0.172929	12.101	< 2e-16	***
as.factor(Categorical.age.5.y)7	2.514578	0.172998	14.535	< 2e-16	***
as.factor(Categorical.age.5.y)8	2.708603	0.278682	9.719	< 2e-16	***

Supplementary table S2.17: Logistic regression covariates in the WBFM model predicting colorectal cancer in men

	Estimate	Std. Error	Z	p	
(Intercept)	-6.476670	0.195202	33.179	< 2e-16	***
as.factor(BMI_quantile)2	-0.016789	0.078865	-0.213	0.831418	
as.factor(BMI_quantile)3	0.215022	0.074939	2.869	0.004114	**
as.factor(BMI_quantile)4	0.219916	0.075125	2.927	0.003419	**
as.factor(BMI_quantile)5	0.277798	0.076087	3.651	0.000261	***
as.factor(Town_5)2	0.021976	0.068324	0.322	0.747724	
as.factor(Town_5)3	-0.079177	0.070592	-1.122	0.262027	
as.factor(Town_5)4	-0.062497	0.072055	-0.867	0.385750	
as.factor(Town_5)5	0.013634	0.073585	0.185	0.853012	
as.factor(Town_5)8888888	-0.747878	1.004621	-0.744	0.456612	
as.factor(X21000.0.0.Ethnic.background)1	0.129261	0.584534	0.221	0.824987	
as.factor(X21000.0.0.Ethnic.background)1002	0.413417	0.116825	3.539	0.000402	***
as.factor(X21000.0.0.Ethnic.background)1003	-0.125819	0.157625	-0.798	0.424743	
as.factor(X21000.0.0.Ethnic.background)2	2.077119	1.059873	1.960	0.050021	.
as.factor(X21000.0.0.Ethnic.background)2001	-0.298884	1.005907	-0.297	0.766369	
as.factor(X21000.0.0.Ethnic.background)2002	0.897125	0.718628	1.248	0.211890	
as.factor(X21000.0.0.Ethnic.background)2003	-0.834013	1.003591	-0.831	0.405958	
as.factor(X21000.0.0.Ethnic.background)2004	1.073457	0.363240	2.955	0.003124	**
as.factor(X21000.0.0.Ethnic.background)3001	-0.469361	0.288520	-1.627	0.103781	
as.factor(X21000.0.0.Ethnic.background)3002	-1.084258	0.717481	-1.511	0.130737	
as.factor(X21000.0.0.Ethnic.background)3004	-0.763918	0.581845	-1.313	0.189209	
as.factor(X21000.0.0.Ethnic.background)4001	-0.429048	0.382396	-1.122	0.261863	
as.factor(X21000.0.0.Ethnic.background)4002	-0.798536	0.505799	-1.579	0.114391	
as.factor(X21000.0.0.Ethnic.background)5	-0.529161	0.710844	-0.744	0.456627	
as.factor(X21000.0.0.Ethnic.background)55553	0.339885	0.306943	1.107	0.268154	
as.factor(X21000.0.0.Ethnic.background)6	-0.680284	0.411892	-1.652	0.098615	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)1	0.107098	0.060094	1.782	0.074722	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)3	0.005166	0.063379	0.082	0.935038	
as.factor(X1558.0.0.Alcohol.intake.frequency.)4	-0.164294	0.096667	-1.700	0.089208	.
as.factor(X1558.0.0.Alcohol.intake.frequency.)5	-0.129203	0.101790	-1.269	0.204331	
as.factor(X1558.0.0.Alcohol.intake.frequency.)55553	0.216052	0.722391	0.299	0.764880	
as.factor(X1558.0.0.Alcohol.intake.frequency.)6	-0.334096	0.122966	-2.717	0.006588	**
as.factor(X20160.0.0.Ever.smoked)0	-0.096440	0.050653	-1.904	0.056917	.
as.factor(X20160.0.0.Ever.smoked)8888888	0.020191	0.340298	0.059	0.952688	
as.factor(Fhx.bowel.cancer..all.)1	0.295912	0.061860	4.784	1.72e-06	***
as.factor(X1369.0.0.Beef.intake)0	-0.083293	0.114935	-0.725	0.468637	
as.factor(X1369.0.0.Beef.intake)2	0.011765	0.054757	0.215	0.829886	
as.factor(X1369.0.0.Beef.intake)3	0.126808	0.072458	1.750	0.080103	.
as.factor(X1369.0.0.Beef.intake)4	-0.387440	0.583123	-0.664	0.506420	
as.factor(X1369.0.0.Beef.intake)5	-0.061772	0.814783	-0.076	0.939568	
as.factor(X1369.0.0.Beef.intake)55551	-0.022021	0.447919	-0.049	0.960789	
as.factor(X1379.0.0.Lamb.mutton.intake)0	-0.078217	0.080161	-0.976	0.329193	
as.factor(X1379.0.0.Lamb.mutton.intake)2	0.022354	0.056611	0.395	0.692935	
as.factor(X1379.0.0.Lamb.mutton.intake)3	0.092828	0.121736	0.763	0.445740	
as.factor(X1379.0.0.Lamb.mutton.intake)5	1.352268	0.674247	2.006	0.044899	*
as.factor(X1379.0.0.Lamb.mutton.intake)55551	-0.414024	0.365222	-1.134	0.256953	
as.factor(X1379.0.0.Lamb.mutton.intake)55553	0.991677	1.023346	0.969	0.332518	
as.factor(exercise_5)1	0.046774	0.078098	0.599	0.549227	
as.factor(exercise_5)2	-0.022170	0.078949	-0.281	0.778847	
as.factor(exercise_5)3	-0.021139	0.079971	-0.264	0.791521	
as.factor(exercise_5)4	0.013982	0.078556	0.178	0.858730	
as.factor(exercise_5)5	-0.041463	0.079882	-0.519	0.603723	
as.factor(X20003.Statins)1	0.047044	0.059759	0.787	0.431152	
as.factor(X20003.Metformin)1	0.251925	0.147442	1.709	0.087518	.
as.factor(X6155.Touchscreen...Vitamin.D)1	-0.144909	0.152206	-0.952	0.341065	
as.factor(X6155.Touchscreen...Vitamin.D)55551	0.305243	0.237720	1.284	0.199127	
as.factor(X20003.Aspirin)1	-0.111213	0.062077	-1.792	0.073210	.
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)1	0.047709	0.122072	0.391	0.695925	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55551	-0.364426	0.504544	-0.722	0.470118	
as.factor(X2443.0.0.Diabetes.diagnosed.by.doctor)55553	-0.033170	1.024141	-0.032	0.974163	
as.factor(Categorical.age.5.y)3	0.542020	0.202193	2.681	0.007347	**
as.factor(Categorical.age.5.y)4	1.379797	0.181559	7.600	2.97e-14	***
as.factor(Categorical.age.5.y)5	1.726680	0.176383	9.789	< 2e-16	***
as.factor(Categorical.age.5.y)6	2.107539	0.172881	12.191	< 2e-16	***
as.factor(Categorical.age.5.y)7	2.534517	0.172961	14.654	< 2e-16	***
as.factor(Categorical.age.5.y)8	2.734780	0.278668	9.814	< 2e-16	***

Supplementary table S2.18: Logistic regression covariates in the BMI model predicting colorectal cancer in men

Cancer eSNPs	eGenes	eQTL source	Cancer Causing Gene (CCG) from COSMIC Cancer Gene Census	Summarised GWAS catalog disease trait	GWAS Pubmed ID
rs2069757	AFF4	PancanQTL	oncogene, fusion	Hodgkin's lymphoma	24920014
rs6596100	AFF4	PancanQTL	oncogene, fusion	Breast cancer	29059683
rs6964587	AKAP9	NESDA.NTR	fusion	Breast cancer	29059683
rs10623258	AKT1	PancanQTL	oncogene	Breast cancer	29059683
rs4767364	ALDH2	NESDA.NTR	fusion	Oral cavity	21437268
rs3184504	ALDH2	PancanQTL	fusion	Endometrial cancer	26621817
rs3184504	ALDH2	PancanQTL	fusion	Colorectal cancer	26151821
rs1014971	APOBEC3B	NESDA.NTR	oncogene, TSG	Bladder cancer	24163127
rs1014971	APOBEC3B	NESDA.NTR	oncogene, TSG	Bladder cancer	20972438
rs17599629	ARNT	NESDA.NTR	oncogene, TSG, fusion	Prostate cancer	25217961
rs7412746	ARNT	GTEX; NESDA.NTR; PancanQTL	oncogene, TSG, fusion	Melanoma	21983785
rs7412746	ARNT	GTEX; NESDA.NTR; PancanQTL	oncogene, TSG, fusion	Melanoma	28212542
rs11169552	ATF1	NESDA.NTR	oncogene, fusion	Colorectal cancer	20972440
rs34245511	ATF1	NESDA.NTR	oncogene, fusion	Colorectal cancer	24737748
rs11374964	ATM	PancanQTL	TSG	Breast cancer	29058716
rs1830298	CASP8	GTEX; NESDA.NTR; PancanQTL	TSG	Breast cancer	29059683
rs13014235	CASP8	NESDA.NTR; PancanQTL	TSG	Non-melanomatous skin cancer	25855136
rs3769821	CASP8	NESDA.NTR; PancanQTL	TSG	Breast cancer	29059683
rs7558911	CASP8	NESDA.NTR	TSG	Chronic lymphocytic leukemia	28165464
rs77688320	CASP8	NESDA.NTR; PancanQTL	TSG	Breast cancer	29059683
rs13016963	CASP8	GTEX; NESDA.NTR; PancanQTL	TSG	Melanoma	21983787
rs13016963	CASP8	GTEX; NESDA.NTR; PancanQTL	TSG	Melanoma	28212542
rs13015798	CASP8	NESDA.NTR	TSG	Chronic lymphocytic leukemia	26956414
rs2080303	CASP8	PancanQTL	TSG	Non-melanomatous skin cancer	27539887
rs2714486	CASP8	NESDA.NTR; PancanQTL	TSG	Breast cancer	29059683

rs3769823	CASP8	NESDA.NTR; PancanQTL	TSG	Non-melanomatous skin cancer	28177523
rs3769825	CASP8	NESDA.NTR	TSG	Chronic lymphocytic leukemia	23770605
rs498136	CCND1	NESDA.NTR	oncogene, fusion	Melanoma	28212542
rs498136	CCND1	NESDA.NTR	oncogene, fusion	Melanoma	26237428
rs8102137	CCNE1	GTEX; PancanQTL	oncogene	Bladder cancer	20972438
rs8102137	CCNE1	GTEX; PancanQTL	oncogene	Bladder cancer	24163127
rs2290400	CDK12	NESDA.NTR	TSG	Acute lymphoblastic leukemia	29348612
rs1321311	CDKN1A	NESDA.NTR	oncogene, TSG	Colorectal cancer	22634755
rs17694493	CDKN2A	PancanQTL	TSG	Prostate cancer	25217961
rs6005807	CHEK2	PancanQTL	TSG	Ovarian cancer	28346442
rs7006527	COX6C	NESDA.NTR	fusion	Non-melanomatous skin cancer	24403052
rs7006527	COX6C	NESDA.NTR	fusion	Non-melanomatous skin cancer	25855136
rs141115006	COX6C	NESDA.NTR	fusion	Non-melanomatous skin cancer	27539887
rs12803321	DDX6	NESDA.NTR	oncogene, fusion	Glioma	26424050
rs12803321	DDX6	NESDA.NTR	oncogene, fusion	Glioma	28346443
rs4808801	ELL	NESDA.NTR; PancanQTL	TSG, fusion	Breast cancer	23535729
rs4808801	ELL	NESDA.NTR; PancanQTL	TSG, fusion	Breast cancer	29059683
rs7258465	ELL	NESDA.NTR; PancanQTL	TSG, fusion	Breast cancer	29059683
rs172032	ELL	GTEX; NESDA.NTR	TSG, fusion	Breast cancer	29059683
rs4808801	ELL	NESDA.NTR; PancanQTL	TSG, fusion	Breast cancer	25751625
rs7529522	FAM46C	NESDA.NTR	TSG	Breast cancer	29059683
rs258322	FANCA	PancanQTL	TSG	Melanoma	19578364
rs258322	FANCA	PancanQTL	TSG	Melanoma	28212542
rs258322	FANCA	PancanQTL	TSG	Melanoma	24980573
rs4785763	FANCA	NESDA.NTR	TSG	Melanoma	19578364
rs1805007	FANCA	PancanQTL	TSG	Non-melanomatous skin cancer	23548203
rs8063761	FANCA	NESDA.NTR	TSG	Non-melanomatous skin cancer	26908436
rs4785763	FANCA	NESDA.NTR	TSG	Melanoma	28212542
rs1805007	FANCA	PancanQTL	TSG	Non-melanomatous skin cancer	21700618
rs258322	FANCA	PancanQTL	TSG	Melanoma	21983787

rs1805007	FANCA	PancanQTL	TSG	Melanoma	28212542
rs1805007	FANCA	PancanQTL	TSG	Non-melanomatous skin cancer	27539887
rs1805007	FANCA	PancanQTL	TSG	Non-melanomatous skin cancer	27424798
rs1800682	FAS	GTEX; NESDA.NTR; PancanQTL	TSG	Chronic lymphocytic leukemia	24292274
rs4406737	FAS	GTEX; NESDA.NTR	TSG	Chronic lymphocytic leukemia	26956414
rs6586163	FAS	NESDA.NTR; PancanQTL	TSG	Chronic lymphocytic leukemia	28165464
rs4406737	FAS	GTEX; NESDA.NTR	TSG	Chronic lymphocytic leukemia	23770605
rs71658797	FUBP1	PancanQTL	oncogene	Lung cancer	28604730
rs17391694	FUBP1	GTEX	oncogene	Lung cancer	28604730
rs71658797	FUBP1	PancanQTL	oncogene	Lung cancer	28604730
rs7953249	HNF1A	PancanQTL	TSG	Ovarian cancer	28346442
rs1110701	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	25310577
rs11978267	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	19684603
rs11978267	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	22076464
rs4132601	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	19684604
rs11978267	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	29348612
rs2167364	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	22076464
rs11980379	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	23996088
rs6964969	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	23512250
rs4132601	IKZF1	NESDA.NTR	TSG, fusion	Acute lymphoblastic leukemia	22076464
rs12203592	IRF4	GTEX	oncogene, TSG, fusion	Non-melanomatous skin cancer	27539887
rs12203592	IRF4	GTEX	oncogene, TSG, fusion	Non-melanomatous skin cancer	23548203
rs12203592	IRF4	GTEX	oncogene, TSG, fusion	Non-melanomatous skin cancer	26829030
rs12203592	IRF4	GTEX	oncogene, TSG, fusion	Non-melanomatous skin cancer	27424798
rs12752552	JAK1	NESDA.NTR; PancanQTL	oncogene, TSG	Glioma	28346443
rs898518	LEF1	NESDA.NTR	oncogene, TSG	Chronic lymphocytic leukemia	23770605

rs7690934	LEF1	NESDA.NTR	oncogene, TSG	Chronic lymphocytic leukemia	28165464
rs11071896	MAP2K1	NESDA.NTR	oncogene	Testicular germ cell tumor	28604728
rs60180747	MAP2K1	NESDA.NTR	oncogene	Testicular germ cell tumor	28604732
rs889312	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	25751625
rs889312	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	17529967
rs889312	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	20453838
rs1862626	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	27197191
rs889312	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	24493630
rs62355902	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	29059683
rs62355901	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	29059683
rs889312	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	23535729
rs6890270	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	29059683
rs16886181	MAP3K1	NESDA.NTR	oncogene, TSG	Breast cancer	24493630
rs4245739	MDM4	NESDA.NTR; PancanQTL	oncogene	Breast cancer	23535733
rs4245739	MDM4	NESDA.NTR; PancanQTL	oncogene	Breast cancer	29058716
rs4245739	MDM4	NESDA.NTR; PancanQTL	oncogene	Prostate cancer	23535732
rs4245739	MDM4	NESDA.NTR; PancanQTL	oncogene	Breast cancer	27117709
rs4252707	MDM4	NESDA.NTR	oncogene	Glioma	28346443
rs6920364	MLLT4	NESDA.NTR	oncogene, fusion	Lung cancer	28604730
rs760077	MUC1	GTEX	fusion	Gastric cancer	26098866
rs1707302	MUTYH	PancanQTL	TSG	Breast cancer	29059683
rs1447295	MYC	PancanQTL	oncogene, fusion	Prostate cancer	19767754
rs1447295	MYC	PancanQTL	oncogene, fusion	Prostate cancer	17401363
rs1447295	MYC	PancanQTL	oncogene, fusion	Prostate cancer	17401366
rs1447295	MYC	PancanQTL	oncogene, fusion	Prostate cancer	24753544
rs7404843	MYH11	NESDA.NTR	fusion	Testicular germ cell tumor	28604728
rs79480871	NCOA1	NESDA.NTR	fusion	Diffuse large B cell lymphoma	25261932
rs10993994	NCOA4	GTEX; NESDA.NTR	TSG, fusion	Prostate cancer	26034056
rs10993994	NCOA4	GTEX; NESDA.NTR	TSG, fusion	Prostate cancer	18264096
rs10993994	NCOA4	GTEX; NESDA.NTR	TSG, fusion	Prostate cancer	18264097
rs10763567	NCOA4	NESDA.NTR; PancanQTL	TSG, fusion	Prostate cancer	27262462
rs10993994	NCOA4	GTEX; NESDA.NTR	TSG, fusion	Prostate cancer	24753544

rs10993994	NCOA4	GTEX; NESDA.NTR	TSG, fusion	Prostate cancer	25939597
rs3123078	NCOA4	NESDA.NTR; PancanQTL	TSG, fusion	Prostate cancer	19767753
rs7034162	NFIB	GTEX	fusion	Osteosarcoma	26084801
rs368187	NKX2-1	PancanQTL	oncogene, TSG	Thyroid cancer	28195142
rs57994353	NOTCH1	NESDA.NTR	oncogene, TSG, fusion	Non-melanomatous skin cancer	27424798
rs4236709	NRG1	NESDA.NTR; PancanQTL	TSG, fusion	Lung cancer	28604730
rs2439302	NRG1	NESDA.NTR; PancanQTL	TSG, fusion	Thyroid cancer	22267200
rs2466076	NRG1	NESDA.NTR; PancanQTL	TSG, fusion	Thyroid cancer	28195142
rs3850699	NT5C2	NESDA.NTR	oncogene	Prostate cancer	23535732
rs928674	NUP214	NESDA.NTR	fusion	Oral cavity cancer	27749845
rs3802842	POU2AF1	NESDA.NTR	oncogene, fusion	Colorectal cancer	18372901
rs6589219	POU2AF1	NESDA.NTR	oncogene, fusion	Colorectal cancer	28960316
rs3802842	POU2AF1	NESDA.NTR	oncogene, fusion	Colorectal cancer	26151821
rs2594714	PRKACA	PancanQTL	oncogene	Breast cancer	29059683
rs1432295	REL	NESDA.NTR	oncogene	Hodgkin's lymphoma	21037568
rs4888262	RFWD3	NESDA.NTR; PancanQTL	TSG	Testicular germ cell tumor	23666239
rs4888262	RFWD3	NESDA.NTR; PancanQTL	TSG	Testicular germ cell tumor	28604732
rs4888262	RFWD3	NESDA.NTR; PancanQTL	TSG	Testicular germ cell tumor	28604728
rs7193541	RFWD3	GTEX; NESDA.NTR; PancanQTL	TSG	Multiple myeloma	27363682
rs9905704	RNF43	NESDA.NTR	TSG	Testicular germ cell tumor	28604728
rs9905704	RNF43	NESDA.NTR	TSG	Testicular germ cell tumor	23666240
rs9905704	RNF43	NESDA.NTR	TSG	Testicular germ cell tumor	28604732
rs9905704	RNF43	NESDA.NTR	TSG	Testicular germ cell tumor	23666239
rs7229639	SMAD2	PancanQTL	TSG	Colorectal cancer	24836286
rs2289261	SMAD3	PancanQTL	TSG	Thyroid cancer	28195142
rs56062135	SMAD3	PancanQTL	TSG	Thyroid cancer	28195142
rs34245511	SMARCD1	NESDA.NTR	TSG	Colorectal cancer	24737748
rs401681	TERT	GTEX	oncogene, TSG	Non-melanomatous skin cancer	19151717
rs401681	TERT	GTEX	oncogene, TSG	Bladder cancer	24163127
rs401681	TERT	GTEX	oncogene, TSG	Lung cancer	18978787

rs401681	TERT	GTEX	oncogene, TSG	Non-melanomatous skin cancer	24403052
rs401681	TERT	GTEX	oncogene, TSG	Non-melanomatous skin cancer	25855136
rs2853676	TERT	GTEX	oncogene, TSG	Glioma	19578367
rs401681	TERT	GTEX	oncogene, TSG	Melanoma	21983787
rs17838698	TGFBR2	NESDA.NTR	TSG	Breast cancer	29059683
rs78378222	TP53	NESDA.NTR	oncogene, TSG, fusion	Non-melanomatous skin cancer	27539887
rs78378222	TP53	NESDA.NTR	oncogene, TSG, fusion	Non-melanomatous skin cancer	24403052
rs78378222	TP53	NESDA.NTR	oncogene, TSG, fusion	Non-melanomatous skin cancer	25855136
rs78378222	TP53	NESDA.NTR	oncogene, TSG, fusion	Glioma	28346443
rs78378222	TP53	NESDA.NTR	oncogene, TSG, fusion	Non-melanomatous skin cancer	21946351
rs35850753	TP53	NESDA.NTR	oncogene, TSG, fusion	Glioma	26424050
rs8753	TP53	NESDA.NTR	oncogene, TSG, fusion	Glioma	26424050
rs67340775	TRIM27	PancanQTL	oncogene, fusion	Lung cancer	28604730

Supplementary table S3.1: Cancer eSNPs identified from mining the Cancer Gene Census, the GWAS catalog and eQTL databases (PancanQTL, NESDA.NTR, GTEX).

Cancer eSNPs	eGenes	Present in UK Biobank	Passed QC
rs1014971	APOBEC3B	Yes	Yes
rs10623258	AKT1	Yes	Yes
rs10763567	NCOA4	Yes	Yes
rs10993994	NCOA4	Yes	Yes
rs11071896	MAP2K1	Yes	Yes
rs1110701	IKZF1	Yes	Yes
rs11169552	ATF1	Yes	Yes
rs11374964	ATM	Yes	Yes
rs11978267	IKZF1	Yes	Yes
rs11980379	IKZF1	Yes	Yes
rs12203592	IRF4	Yes	Yes
rs12752552	JAK1	Yes	Yes
rs12803321	DDX6	Yes	Yes
rs13014235	CASP8	Yes	Yes
rs13015798	CASP8	Yes	Yes
rs13016963	CASP8	Yes	Yes
rs1321311	CDKN1A	Yes	Yes
rs141115006	COX6C	No	NA

rs1432295	REL	Yes	Yes
rs1447295	MYC	Yes	Yes
rs16886181	MAP3K1	Yes	Yes
rs1707302	MUTYH	Yes	Yes
rs172032	ELL	Yes	Yes
rs17391694	FUBP1	Yes	Yes
rs17599629	ARNT	Yes	Yes
rs17694493	CDKN2A	Yes	Yes
rs17838698	TGFBR2	Yes	Yes
rs1800682	FAS	Yes	Yes
rs1805007	FANCA	Yes	Yes
rs1830298	CASP8	Yes	Yes
rs1862626	MAP3K1	Yes	Yes
rs2069757	AFF4	Yes	Yes
rs2080303	CASP8	Yes	Yes
rs2167364	IKZF1	Yes	Yes
rs2289261	SMAD3	Yes	Yes
rs2290400	CDK12	Yes	Yes
rs2439302	NRG1	Yes	Yes
rs2466076	NRG1	Yes	Yes

rs258322	FANCA	Yes	Yes
rs2594714	PRKACA	Yes	Yes
rs2714486	CASP8	Yes	Yes
rs2853676	TERT	Yes	Yes
rs3123078	NCOA4	Yes	Yes
rs3184504	ALDH2	Yes	Yes
rs34245511	ATF1	Yes	Yes
rs34245511	SMARCD1	Yes	Yes
rs35850753	TP53	Yes	Yes
rs368187	NKX2-1	Yes	Yes
rs3769821	CASP8	Yes	Yes
rs3769823	CASP8	Yes	Yes
rs3769825	CASP8	Yes	Yes
rs3802842	POU2AF1	Yes	Yes
rs3850699	NT5C2	Yes	Yes
rs401681	TERT	Yes	Yes
rs4132601	IKZF1	Yes	Yes
rs4236709	NRG1	Yes	Yes
rs4245739	MDM4	Yes	Yes
rs4252707	MDM4	Yes	Yes

rs4406737	FAS	Yes	Yes
rs4767364	ALDH2	Yes	Yes
rs4785763	FANCA	Yes	Yes
rs4808801	ELL	Yes	Yes
rs4888262	RFWD3	Yes	Yes
rs498136	CCND1	Yes	Yes
rs56062135	SMAD3	Yes	Yes
rs57994353	NOTCH1	Yes	Yes
rs6005807	CHEK2	Yes	Yes
rs60180747	MAP2K1	Yes	Yes
rs62355901	MAP3K1	Yes	Yes
rs62355902	MAP3K1	Yes	Yes
rs6586163	FAS	Yes	Yes
rs6589219	POU2AF1	Yes	Yes
rs6596100	AFF4	Yes	Yes
rs67340775	TRIM27	Yes	Yes
rs6890270	MAP3K1	Yes	Yes
rs6920364	MLLT4	Yes	Yes
rs6964587	AKAP9	Yes	Yes
rs6964969	IKZF1	Yes	Yes

rs7006527	COX6C	Yes	Yes
rs7034162	NFIB	Yes	Yes
rs71658797	FUBP1	Yes	Yes
rs7193541	RFWD3	Yes	Yes
rs7229639	SMAD2	Yes	Yes
rs7258465	ELL	Yes	Yes
rs7404843	MYH11	Yes	Yes
rs7412746	ARNT	Yes	Yes
rs7529522	FAM46C	Yes	Yes
rs7558911	CASP8	Yes	Yes
rs760077	MUC1	Yes	Yes
rs7690934	LEF1	Yes	Yes
rs77688320	CASP8	Yes	Yes
rs78378222	TP53	Yes	Yes
rs79480871	NCOA1	Yes	Yes
rs7953249	HNF1A	Yes	Yes
rs8063761	FANCA	Yes	Yes
rs8102137	CCNE1	Yes	Yes
rs8753	TP53	Yes	Yes
rs889312	MAP3K1	Yes	Yes

rs898518	LEF1	Yes	Yes
rs928674	NUP214	Yes	Yes
rs9905704	RNF43	Yes	Yes

Supplementary table S3.2: Cancer eSNPs presence in the UK Biobank and whether passed each SNP quality control (QC).

Supplementary table S3.3: Cancer eSNP associations with anthropometric traits in the UK Biobank. Rows highlighted in yellow indicate the three lead cancer eSNPs which have pleiotropic associations with anthropometric traits and cancer in the UK Biobank (online only)

Supplementary table S3.4: Cancer eSNP associations with cancer types in the UK Biobank. Rows highlighted in yellow indicate the three lead cancer eSNPs which have pleiotropic associations with anthropometric traits and cancer in the UK Biobank (online only)

eSNP	eGene	Summarised GWAS catalog disease trait	Pubmed ID	GWAS catalog study p value of association	GWAS catalog odds ratio or beta value	GWAS catalog odds ratio confidence interval (if present)
rs78378222	TP53	Non-melanomatous skin cancer	27539887	2E-10	1.41	Not recorded in GWAS catalog
rs78378222	TP53	Non-melanomatous skin cancer	24403052	4E-22	2.24	1.90-2.64
rs78378222	TP53	Non-melanomatous skin cancer	25855136	1E-20	2.07	Not recorded in GWAS catalog
rs78378222	TP53	Glioma	28346443	5E-27	2.73	2.27-3.28
rs78378222	TP53	Non-melanomatous skin cancer	21946351	2E-20	2.16	1.83-2.54
rs78378222	TP53	Glioma	28346443	5E-29	2.63	2.22-3.11
rs78378222	TP53	Glioma	28346443	9E-38	2.53	2.19-2.91
rs1805007	FANCA	Non-melanomatous skin cancer	21700618	4E-17	1.55	1.45-1.66
rs1805007	FANCA	Melanoma	28212542	4E-17	1.5	1.41-1.59
rs1805007	FANCA	Non-melanomatous skin cancer	27539887	3E-63	1.4	Not recorded in GWAS catalog
rs1805007	FANCA	Non-melanomatous skin cancer	27424798	9E-39	1.46	Not recorded in GWAS catalog
rs1805007	FANCA	Non-melanomatous skin cancer	23548203	3E-10	0.34	0.24-0.44 unit increase
rs889312	MAP3K1	Breast cancer	25751625	1E-39	1.12	1.1-1.15
rs889312	MAP3K1	Breast cancer	17529967	7E-20	1.13	1.10-1.16
rs889312	MAP3K1	Breast cancer	20453838	5.E-09	1.22	1.14-1.30

Supplementary table S3.5: Summary of the cancer eSNPs associations in the UK Biobank and in GWAS catalog

	Standing Height	Sitting Height	Seated Height	Breast Cancer
Phenotype (adjusted-p)	5.91E-14	3.25E-09	0.0000657	0.000000225
Phenotype (beta)	-0.01500	-0.01399	-0.01179	0.10097
Sex (beta)	-1.43674	-1.26032	-1.15247	5.15663
Age at recruitment (beta)	-0.13758	-0.19098	-0.16400	0.39247
Genetic Principal Component 1 (beta)	-0.01123	-0.02121	-0.01985	-0.03260
Genetic Principal Component 2 (beta)	0.01805	0.01638	0.01874	-0.00215
Genetic Principal Component 3 (beta)	-0.00689	-0.01486	-0.01162	0.02108
Genetic Principal Component 4 (beta)	0.00802	0.00408	0.01204	-0.00925
Genetic Principal Component 5 (beta)	-0.07004	-0.04383	-0.01592	0.01303
Genetic Principal Component 6 (beta)	-0.00498	-0.00298	-0.00378	-0.00043
Genetic Principal Component 7 (beta)	-0.02606	-0.01680	-0.02108	-0.00497
Genetic Principal Component 8 (beta)	0.00647	0.00580	0.00735	-0.00910
Genetic Principal Component 9 (beta)	0.00702	0.00269	0.00497	-0.01434
Genetic Principal Component 10 (beta)	0.00292	0.00497	0.00790	0.01934
Genetic Principal Component 11 (beta)	-0.02675	-0.02349	-0.03450	-0.00411
Genetic Principal Component 12 (beta)	0.02509	0.02045	0.03191	0.00472
Genetic Principal Component 13 (beta)	0.00119	0.00092	0.00259	0.02537
Genetic Principal Component 14 (beta)	-0.01422	-0.00351	0.02552	0.03213
Genetic Principal Component 15 (beta)	-0.00203	-0.00273	-0.00031	0.02050
Genetic Principal Component 16 (beta)	-0.00666	-0.00510	0.01194	0.00659
Genetic Principal Component 17 (beta)	0.00293	0.00041	0.00112	0.00055
Genetic Principal Component 18 (beta)	-0.00299	-0.00076	-0.01707	-0.00965
Genetic Principal Component 19 (beta)	0.00144	0.00032	-0.00374	-0.00466
Genetic Principal Component 20 (beta)	-0.00827	-0.00664	0.00070	-0.01501

Supplementary table S3.6: PLINK analysis of lead MAP3K1 cancer eSNP (rs889312)

	Standing Height	Melanoma	Non-melanomatous skin cancer
Phenotype (adjusted-p)	7.43E-15	1.18E-36	1.64E-58
Phenotype (beta)	-0.02286	0.48122	0.29054
Sex (beta)	-1.43675	0.20279	-0.19010
Age at recruitment (beta)	-0.13758	0.27661	0.69601
Genetic Principal Component 1 (beta)	-0.01140	-0.06021	-0.08380
Genetic Principal Component 2 (beta)	0.01820	0.05465	0.03900
Genetic Principal Component 3 (beta)	-0.00696	-0.08207	-0.04078
Genetic Principal Component 4 (beta)	0.00834	-0.05715	0.00720
Genetic Principal Component 5 (beta)	-0.06989	-0.03217	0.08130
Genetic Principal Component 6 (beta)	-0.00515	0.00432	0.00257
Genetic Principal Component 7 (beta)	-0.02612	-0.03544	-0.01440
Genetic Principal Component 8 (beta)	0.00632	-0.01878	-0.03837
Genetic Principal Component 9 (beta)	0.00718	-0.02169	0.01792
Genetic Principal Component 10 (beta)	0.00306	0.06202	0.00682
Genetic Principal Component 11 (beta)	-0.02656	-0.08199	0.00145
Genetic Principal Component 12 (beta)	0.02513	-0.00340	0.02291
Genetic Principal Component 13 (beta)	0.00119	-0.02480	-0.00338
Genetic Principal Component 14 (beta)	-0.01394	-0.01355	0.01681
Genetic Principal Component 15 (beta)	-0.00209	-0.01827	-0.01082
Genetic Principal Component 16 (beta)	-0.00668	0.01386	0.00339
Genetic Principal Component 17 (beta)	0.00282	0.00604	-0.01024
Genetic Principal Component 18 (beta)	-0.00291	-0.01575	0.01793
Genetic Principal Component 19 (beta)	0.00144	0.03056	-0.00924
Genetic Principal Component 20 (beta)	-0.00819	-0.03665	-0.00929

Supplementary table S3.7: PLINK analysis of lead FANCA cancer eSNP (rs1805007)

	Standing height	Sitting height	Seated height	Weight	Basal metabolic rate	Whole body fat-free mass	Trunk fat-free mass	Arm fat-free mass left	Arm fat-free mass right	Leg fat-free mass left	Leg fat-free mass right	Brain	Non-melanomatous skin cancer
Phenotype (adjusted-p)	1.74E-20	9.63E-29	5.80E-19	3.45E-12	8.71E-28	6.54E-33	7.93E-36	1.82E-33	3.24E-28	1.05E-22	5.92E-26	0.000000139	9.16E-14
Phenotype (beta)	0.07397	0.09563	0.08391	0.07647	0.07650	0.07333	0.07555	0.07676	0.06739	0.06842	0.07207	1.17069	0.39388
Sex (beta)	-1.43664	-1.26019	-1.15236	-0.93149	-1.57671	-1.67901	-1.68909	-1.65569	-1.69024	-1.59133	-1.59644	-0.40494	-0.18942
Age at recruitment (beta)	-0.13757	-0.19097	-0.16399	-0.03201	-0.11593	-0.10303	-0.08333	-0.08038	-0.09266	-0.12980	-0.14006	0.20876	0.69533
Genetic Principal Component 1 (beta)	-0.01124	-0.02121	-0.01984	0.00249	-0.00065	-0.00131	-0.00179	0.00001	-0.00134	-0.00075	-0.00089	-0.22848	-0.08584
Genetic Principal Component 2 (beta)	0.01812	0.01645	0.01880	0.01092	0.01095	0.01066	0.01165	0.00947	0.00917	0.00941	0.00936	-0.05807	0.04015
Genetic Principal Component 3 (beta)	-0.00687	-0.01483	-0.01160	0.00320	-0.00092	-0.00165	-0.00251	-0.00046	-0.00038	-0.00101	-0.00071	-0.04168	-0.04213
Genetic Principal Component 4 (beta)	0.00795	0.00398	0.01195	0.02139	0.00560	0.00197	0.00053	0.00542	0.00419	0.00358	0.00226	-0.13149	0.00998
Genetic Principal Component 5 (beta)	-0.06989	-0.04362	-0.01573	-0.01369	-0.03032	-0.03322	-0.03730	-0.02542	-0.02707	-0.02682	-0.03064	0.05250	0.08521
Genetic Principal Component 6 (beta)	-0.00498	-0.00296	-0.00377	0.00007	0.00158	0.00177	0.00160	0.00206	0.00227	0.00157	0.00198	-0.02187	0.00103
Genetic Principal Component 7 (beta)	-0.02609	-0.01684	-0.02111	-0.00233	-0.00956	-0.01090	-0.01283	-0.00774	-0.00830	-0.00842	-0.00924	0.02111	-0.01486
Genetic Principal Component 8 (beta)	0.00655	0.00586	0.00740	0.00801	0.00954	0.00958	0.00985	0.00964	0.00832	0.00895	0.00913	0.05473	-0.04206
Genetic Principal Component 9 (beta)	0.00713	0.00281	0.00507	-0.00412	-0.00511	-0.00517	-0.00681	-0.00844	-0.00652	-0.00197	-0.00130	-0.03610	0.01971
Genetic Principal Component 10 (beta)	0.00290	0.00494	0.00788	0.00270	0.00171	0.00160	0.00141	0.00169	0.00181	0.00147	0.00168	0.02506	0.00854
Genetic Principal Component 11 (beta)	-0.02664	-0.02336	-0.03438	0.00031	-0.01054	-0.01272	-0.01531	-0.00980	-0.00941	-0.00860	-0.01007	0.07317	0.00399
Genetic Principal Component 12 (beta)	0.02509	0.02044	0.03190	0.01212	0.01027	0.00957	0.01067	0.00911	0.00921	0.00802	0.00785	0.05598	0.02240
Genetic Principal Component 13 (beta)	0.00122	0.00095	0.00262	0.00304	0.00150	0.00107	0.00106	0.00174	0.00167	0.00114	0.00071	-0.01785	-0.00377
Genetic Principal Component 14 (beta)	-0.01419	-0.00349	0.02553	-0.00430	-0.00850	-0.00926	-0.01029	-0.00858	-0.00763	-0.00713	-0.00859	0.03186	0.02001
Genetic Principal Component 15 (beta)	-0.00212	-0.00283	-0.00039	-0.00081	-0.00124	-0.00126	-0.00103	-0.00077	-0.00099	-0.00126	-0.00171	-0.03294	-0.01090
Genetic Principal Component 16 (beta)	-0.00674	-0.00519	0.01187	-0.01012	-0.00645	-0.00535	-0.00494	-0.00565	-0.00422	-0.00624	-0.00598	-0.01527	0.00390
Genetic Principal Component 17 (beta)	0.00286	0.00034	0.00106	-0.00095	-0.00032	0.00002	0.00032	-0.00064	-0.00057	-0.00031	-0.00024	-0.09038	-0.01146
Genetic Principal Component 18 (beta)	-0.00297	-0.00075	-0.01706	0.00137	0.00126	0.00104	0.00123	0.00278	0.00208	-0.00002	0.00030	-0.03965	0.01835
Genetic Principal Component 19 (beta)	0.00142	0.00030	-0.00375	-0.00186	0.00041	0.00098	0.00130	-0.00056	0.00035	0.00081	0.00132	-0.02947	-0.00888
Genetic Principal Component 20 (beta)	-0.00822	-0.00657	0.00076	0.01137	0.00270	0.00067	-0.00030	0.00425	0.00288	0.00117	-0.00019	-0.04569	-0.00797

Supplementary table S3.8: PLINK analysis of lead TP53 cancer eSNP (rs78378222)

Supplementary table S4.1: GWAS results for Body Mass Index (BMI). (Online only)

Supplementary table S4.2: GWAS results for Whole Body Fat-Free Mass (WBFFM). (Online only)

Supplementary table S4.3: GWAS results for Whole Body Fat Mass (WBFM). (Online only)

Supplementary table S4.4: GWAS results for Colorectal cancer. (Online only)

Supplementary table S4.5: GWAS results for Prostate cancer. (Online only)

Supplementary table S4.6: GWAS results for Post-menopausal cancer. (Online only)

Supplementary table S4.7: SNPs with positive pleiotropic associations with Prostate cancer and anthropometric traits. (Online only)

Supplementary table S4.8: SNPs with negative pleiotropic associations with Prostate cancer and anthropometric traits. (Online only)

Supplementary table S4.9: SNPs with mixed pleiotropic associations with Prostate cancer and anthropometric traits [negative associations with Prostate cancer and positive with anthropometric traits]. (Online only)

Supplementary table S4.10: SNPs with mixed pleiotropic associations with Prostate cancer and anthropometric traits [positive associations with Prostate cancer and negative with anthropometric traits]. (Online only)

Supplementary table S4.11: SNPs with positive pleiotropic associations with Post-menopausal breast cancer and anthropometric traits. (Online only)

Supplementary table S4.12: SNPs with negative pleiotropic associations with Post-menopausal breast cancer and anthropometric traits. (Online only)

Supplementary table S4.13: SNPs with mixed pleiotropic associations with Post-menopausal breast cancer and anthropometric traits [negative associations with Post-menopausal breast cancer and positive with anthropometric traits]. (Online only)

Supplementary table S4.14: SNPs with mixed pleiotropic associations with Post-menopausal breast cancer and anthropometric traits [positive associations with Post-menopausal breast cancer and negative with anthropometric traits]. (Online only)

Baseline Demographics	Number of patients
Male	5
Age	51 (34-69)
BAME	0
Diabetes	2
Cardiovascular (HTN)	1
Respiratory (asthma/COPD)	1
Chronic Kidney Disease	1
Active cancer	0
Immunosuppressed	0

Supplementary table S5.1: Demographic information workers in the Asymptomatic patient cohort. N.b. BAME stands for the Black, Asian and Minority Ethnic group.

	Overall	Post-COVID-19 Syndrome	Non-Post-COVID-19 Syndrome
Number of staff	42	21	17
Male (%)	14	5	29
Mean age (years)	43	43	44
BAME (%)	36	48	24
Comorbidities present (%)	31	33	24
Pulmonary disease	2	2	0
Cardiovascular disease	4	1	0
Active oncological disease	1	0	1
Endocrine disease	7	4	2
Taking immunosuppressants	1	0	1
Staff on regular medication	10	5	4
Mean number of symptoms	6	7	5
Asymptomatic	3	0	3
Multidisciplinary Team Role			
Administrators	8	6	2
Dietician	1	1	0
Healthcare assistants	3	2	0
Housekeeping staff	3	2	1
Medical doctors	3	1	2
Nurses	12	4	6
Occupational therapists	1	0	1
Pharmacists	2	2	0
Phlebotomist	1	0	1
Physiotherapists	7	3	3

Supplementary table S5.2: Demographic information workers in the Post-COVID-19 Syndrome cohort. N.b. Four staff were lost to follow up and BAME stands for the Black, Asian and Minority Ethnic group.

References

1. Ward, Z. J. *et al.* Projected U.S. State-Level Prevalence of Adult Obesity and Severe Obesity. *New England Journal of Medicine* **381**, 2440–2450 (2019).
2. Smittenaar, C. R., Petersen, K. A., Stewart, K. & Moitt, N. Cancer incidence and mortality projections in the UK until 2035. *Br. J. Cancer* **115**, 1147–1155 (2016).
3. Cancer Intelligence Team. When could overweight and obesity overtake smoking as the biggest cause of cancer in the UK? Cancer Research UK (2018). Available online at: https://www.cancerresearchuk.org/sites/default/files/obesity_tobacco_cross_over_report_final.pdf Accessed: 7/12/20
4. Steele, C. B. *et al.* Vital Signs: Trends in Incidence of Cancers Associated with Overweight and Obesity — United States, 2005–2014. *MMWR. Morbidity and Mortality Weekly Report* **66**, 1052–1058 (2017).
5. Adela Hruby, F. B. H. The Epidemiology of Obesity: A Big Picture. *Pharmacoeconomics* **33**, 673 (2015).
6. Lauby-Secretan, B. *et al.* Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine* **375**, 794–798 (2016).
7. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organ. Tech. Rep. Ser.* **894**, i–xii, 1–253 (2000).
8. Carey, V. J. *et al.* Body fat distribution and risk of non-insulin-dependent diabetes mellitus in women. The Nurses' Health Study. *Am. J. Epidemiol.* **145**, 614–619 (1997).
9. Cassano, P. A., Segal, M. R., Vokonas, P. S. & Weiss, S. T. Body fat distribution, blood pressure, and hypertension. A prospective cohort study of men in the normative aging study. *Ann. Epidemiol.* **1**, 33–48 (1990).
10. Snijder, M. B. *et al.* Trunk fat and leg fat have independent and opposite associations with fasting and postload glucose levels: the Hoorn study. *Diabetes Care* **27**, 372–377 (2004).
11. Calle, E. E., Rodriguez, C., Walker-Thurmond, K. & Thun, M. J. Overweight, obesity,

and mortality from cancer in a prospectively studied cohort of U.S. adults. *New England Journal of Medicine* **348**, (2003).

12. Schauer, D. P. *et al.* Bariatric Surgery and the Risk of Cancer in a Large Multisite Cohort. *Annals of Surgery* **269**, 95–101 (2019).
13. Pearson-Stuttard, J. *et al.* Worldwide burden of cancer attributable to diabetes and high body-mass index: a comparative risk assessment. *Lancet Diabetes Endocrinol* **6**, e6–e15 (2018).
14. Petrelli, F. *et al.* Association of Obesity With Survival Outcomes in Patients With Cancer: A Systematic Review and Meta-analysis. *JAMA Netw Open* **4**, e213520–e213520 (2021).
15. Garofalo, C. & Surmacz, E. Leptin and cancer. *Journal of Cellular Physiology* **207**, 12–22 (2006).
16. Ouchi, N., Parker, J. L., Lugus, J. J. & Walsh, K. Adipokines in inflammation and metabolic disease. *Nature Reviews Immunology* **11**, 85–97 (2011).
17. Tumminia, A. *et al.* Adipose Tissue, Obesity and Adiponectin: Role in Endocrine Cancer Risk. *Int. J. Mol. Sci.* **20**, (2019).
18. Cleary, M. P. & Grossmann, M. E. Minireview: Obesity and breast cancer: the estrogen connection. *Endocrinology* **150**, 2537–2542 (2009).
19. Nathalie Fuentes, P. S. Estrogen receptor signaling mechanisms. *Adv. Protein Chem. Struct. Biol.* **116**, 135 (2019).
20. Vendrell, J. *et al.* Resistin, Adiponectin, Ghrelin, Leptin, and Proinflammatory Cytokines: Relationships in Obesity. *Obesity Research* **12**, 962–971 (2004).
21. Ellulu, M. S., Patimah, I., Khaza'ai, H., Rahmat, A. & Abed, Y. Obesity and inflammation: the linking mechanism and the complications. *Arch. Med. Sci.* **13**, 851 (2017).
22. Gunter, M. J. *et al.* Insulin, insulin-like growth factor-I, and risk of breast cancer in postmenopausal women. *J. Natl. Cancer Inst.* **101**, 48-60 (2009).
23. Christopoulos, P. F., Msaouel, P. & Koutsilieris, M. The role of the insulin-like growth

factor-1 system in breast cancer. *Mol. Cancer* **14**, 43 (2015).

24. Kaaks, R. *et al.* Insulin-like growth factor I and risk of breast cancer by age and hormone receptor status-A prospective study within the EPIC cohort. *Int. J. Cancer* **134**, 2683–2690 (2014).
25. Renehan, A. G., Zwahlen, M. & Egger, M. Adiposity and cancer risk: new mechanistic insights from epidemiology. *Nat. Rev. Cancer* **15**, 484-98 (2015).
26. Insulin-Like Growth Factors are Mitogenic for Human Keratinocytes and a Squamous Cell Carcinoma. *J. Invest. Dermatol.* **96**, 104–110 (1991).
27. Lewitt, M. S., Dent, M. S. & Hall, K. The Insulin-Like Growth Factor System in Obesity, Insulin Resistance and Type 2 Diabetes Mellitus. *J. Clin. Med. Res.* **3**, 1561–1574 (2014).
28. Yau, S. W. *et al.* Serum IGFBP-2 levels are associated with reduced insulin sensitivity in obese children. *Clin. Obes.* **8**, 184–190 (2018).
29. Alessi, M. C. *et al.* Plasminogen activator inhibitor 1, transforming growth factor-beta1, and BMI are closely associated in human adipose tissue during morbid obesity. *Diabetes* **49**, 1374–1380 (2000).
30. Osório-Costa, F. & Carvalheira, J. B. C. TNF-α in obesity-associated colon cancer. *Transl. Gastrointest. Cancer* **2**, 17993–17193 (2013).
31. Stone, T. W., McPherson, M. & Gail Darlington, L. Obesity and Cancer: Existing and New Hypotheses for a Causal Connection. *EBioMedicine* vol. **30**, 14–28 (2018).
32. Ling, C. & Rönn, T. Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metabolism* **29**, 1028–1044 (2019).
33. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
34. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
35. Gong, J., Chehrazi-Raffle, A., Reddi, S. & Salgia, R. Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration

trials and future considerations. *J Immunother Cancer* **6**, 8 (2018).

36. Kishida, K. *et al.* Relationships between circulating adiponectin levels and fat distribution in obese subjects. *J. Atheroscler. Thromb.* **18**, 592–595 (2011).
37. Lönnqvist, F. *et al.* Leptin secretion from adipose tissue in women. Relationship to plasma levels and gene expression. *J. Clin. Invest.* **99**, 2398–2404 (1997).
38. Silha, J. V. & Murphy, L. J. Serum resistin (FIZZ3) protein is increased in obese humans. *The Journal of Clinical Endocrinology and Metabolism* **89**, 1977–8 (2004).
39. Asghari, A. & Umetani, M. Obesity and Cancer: 27-Hydroxycholesterol, the Missing Link. *Int. J. Mol. Sci.* **21**, 4822 (2020).
40. Hashimoto, N. & Hara, H. Dietary Amino Acids Promote Pancreatic Protease Synthesis at the Translation Stage in Rats. *J. Nutr.* **133**, 3052–3057 (2003).
41. Ni, Y. *et al.* Circulating Unsaturated Fatty Acids Delineate the Metabolic Status of Obese Individuals. *EBioMedicine* **2**, 1513–1522 (2015).
42. John, H. J. The relationship of obesity to carbohydrate metabolism. The American Journal Of The Medical Sciences 173, 184–194 (1927).
43. Favennec, M. *et al.* The kynurenine pathway is activated in human obesity and shifted toward kynurenine monooxygenase activation. *Obesity* **23**, 2066–2074 (2015).
44. Filippatos, T. D. *et al.* Effects of increased body weight and short-term weight loss on serum PCSK9 levels – a prospective pilot study. *Archives of Medical Science - Atherosclerotic Diseases* **2**, 46–51 (2017).
45. Karbownik-Lewinska, M. *et al.* Direct contribution of obesity to oxidative damage to macromolecules. *Neuro Endocrinol. Lett.* **33**, 453–461 (2012).
46. Bankoglu, E. E. *et al.* Impact of weight loss induced by gastric bypass or caloric restriction on oxidative stress and genomic damage in obese Zucker rats. *Free Radical Biology and Medicine* **94**, 208–217 (2016).
47. Schmidt, F. M. *et al.* Inflammatory cytokines in general and central obesity and modulating effects of physical activity. *PLoS One* **10**, e0121971 (2015).

48. Moschen, A. R. *et al.* Adipose and Liver Expression of Interleukin (IL)-1 Family Members in Morbid Obesity and Effects of Weight Loss. *Molecular Medicine* **17**, 840–845 (2011).
49. Straczkowski, M. *et al.* Plasma Interleukin-8 Concentrations Are Increased in Obese Subjects and Related to Fat Mass and Tumor Necrosis Factor-α System. *The Journal of Clinical Endocrinology & Metabolism* **87**, 4602–4606 (2002).
50. Blüher, M. *et al.* Association of interleukin-6, C-reactive protein, interleukin-10 and adiponectin plasma concentrations with measures of obesity, insulin sensitivity and glucose metabolism. *Exp. Clin. Endocrinol. Diabetes* **113**, 534–537 (2005).
51. Ahmad, R., Thomas, R., Kochumon, S. & Sindhu, S. Increased adipose tissue expression of IL-18R and its ligand IL-18 associates with inflammation and insulin resistance in obesity. *Immun Inflamm Dis* **5**, 318–335 (2017).
52. Catalán, V. *et al.* Proinflammatory cytokines in obesity: impact of type 2 diabetes mellitus and gastric bypass. *Obes. Surg.* **17**, 1464–1474 (2007).
53. Olszanecka-Glinianowicz, M., Zahorska-Markiewicz, B., Janowska, J. & Zurakowski, A. Serum concentrations of nitric oxide, tumor necrosis factor (TNF)-alpha and TNF soluble receptors in women with overweight and obesity. *Metabolism* **53**, (2004).
54. Yadav, H. *et al.* Protection from obesity and diabetes by blockade of TGF-β/Smad3 signaling. *Cell Metab.* **14**, 67–79 (2011).
55. Lynch, L. *et al.* Invariant NKT cells and CD1d(+) cells amass in human omentum and are depleted in patients with cancer and obesity. *Eur. J. Immunol.* **39**, 1893–1901 (2009).
56. Frasca, D. *et al.* Obesity decreases B cell responses in young and elderly individuals. *Obesity* **24**, 615 (2016).
57. Nishimura, S. *et al.* CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nat. Med.* **15**, 914–920 (2009).
58. Wu, D. *et al.* Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. *Science* **332**, 243–247 (2011).

59. Lumeng, C. N., Bodzin, J. L. & Saltiel, A. R. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. *J. Clin. Invest.* **117**, 175–184 (2007).
60. D'Abbondanza, M. *et al.* Increased plasmatic NETs by-products in patients in severe obesity. *Sci. Rep.* **9**, 14678 (2019).
61. Wang, Z. *et al.* Paradoxical effects of obesity on T cell function during tumor progression and PD-1 checkpoint blockade. *Nature Medicine* **25**, 141–151 (2019).
62. Strissel, K. J. *et al.* T-cell recruitment and Th1 polarization in adipose tissue during diet-induced obesity in C57BL/6 mice. *Obesity* **18**, 1918–1925 (2010).
63. Winer, S. *et al.* Normalization of obesity-associated insulin resistance through immunotherapy. *Nature Medicine* **15**, 921–929 (2009).
64. Feuerer, M. *et al.* Lean, but not obese, fat is enriched for a unique population of regulatory T cells that affect metabolic parameters. *Nature Medicine* **15**, 930–939 (2009).
65. Eriksson, P. *et al.* Adipose tissue secretion of plasminogen activator inhibitor-1 in non-obese and obese individuals. *Diabetologia* **41**, 65–71 (1998).
66. Samad, F., Pandey, M. & Loskutoff, D. J. Regulation of tissue factor gene expression in obesity. *Blood* **98**, 3353–3358 (2001).
67. Chung, K. M. *et al.* Endocrine-Exocrine Signaling Drives Obesity-Associated Pancreatic Ductal Adenocarcinoma. *Cell* **181**, (2020).
68. Kim, M. K., Reaven, G. M., Chen, Y.-D. I., Kim, E. & Kim, S. H. Hyperinsulinemia in individuals with obesity: Role of insulin clearance. *Obesity* **23**, 2430–2434 (2015).
69. Lukanova, A. *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.* **150**, 161–171 (2004).
70. Gómez-Ambrosi, J. *et al.* FGF19 and FGF21 serum concentrations in human obesity and type 2 diabetes behave differently after diet- or surgically-induced weight loss. *Clin. Nutr.* **36**, (2017).

71. Zhang, X. *et al.* Serum FGF21 Levels Are Increased in Obesity and Are Independently Associated With the Metabolic Syndrome in Humans. *Diabetes* **57**, 1246–1253 (2008).
72. Associations of serum fibroblast growth factor 23 levels with obesity and visceral fat accumulation. *Clin. Nutr.* **37**, 223–228 (2018).
73. Bell, L. N. *et al.* Adipose tissue production of hepatocyte growth factor contributes to elevated serum HGF in obesity. *Am. J. Physiol. Endocrinol. Metab.* **291**, E843–8 (2006).
74. Baxter, R. C. IGF binding proteins in cancer: mechanistic and clinical insights. *Nat. Rev. Cancer* **14**, 329–341 (2014).
75. Pollak, M. The insulin and insulin-like growth factor receptor family in neoplasia: an update. *Nat. Rev. Cancer* **12**, (2012).
76. Renehan, A. G., Frystyk, J. & Flyvbjerg, A. Obesity and cancer risk: the role of the insulin–IGF axis. *Trends in Endocrinology & Metabolism* **17**, 328–336 (2006).
77. Holly, J. M. P., Biernacka, K. & Perks, C. M. The Neglected Insulin: IGF-II, a Metabolic Regulator with Implications for Diabetes, Obesity, and Cancer. *Cells* **8**, 1207 (2019).
78. Mohanraj, L. *et al.* IGFBP-3 inhibits cytokine-induced insulin resistance and early manifestations of atherosclerosis. *PLoS One* **8**, e55084 (2013).
79. Gram, I. T. *et al.* Body mass index, waist circumference and waist-hip ratio and serum levels of IGF-I and IGFBP-3 in European women. *Int. J. Obes.* **30**, 1623–1631 (2006).
80. Pervanidou, P. *et al.* Increased placental growth factor (PlGF) concentrations in children and adolescents with obesity and the metabolic syndrome. *Hormones* **13**, 369–374 (2014).
81. Silha, J. V., Krsek, M., Sucharda, P. & Murphy, L. J. Angiogenic factors are elevated in overweight and obese individuals. *Int. J. Obes.* **29**, 1308–1314 (2005).
82. Pasarica, M. *et al.* Adipose tissue collagen VI in obesity. *J. Clin. Endocrinol. Metab.* **94**, 5155–5162 (2009).
83. Juárez-Cruz, J. C. *et al.* Leptin induces cell migration and invasion in a FAK-Src-dependent manner in breast cancer cells. Endocr Connect. **11**, 1539-1552 (2019).

84. Zhang, L. *et al.* The role of AEBP1 in sex-specific diet-induced obesity. *Mol. Med.* **11**, 39–47 (2005).
85. Hao, J. *et al.* Circulating Adipose Fatty Acid Binding Protein Is a New Link Underlying Obesity-Associated Breast/Mammary Tumor Development. *Cell Metab.* **28**, 689–705.e5 (2018).
86. Tabata, M. *et al.* Angiopoietin-like protein 2 promotes chronic adipose tissue inflammation and obesity-related systemic insulin resistance. *Cell Metab.* **10**, 178–188 (2009).
87. Wang, D. *et al.* Colonic Lysine Homocysteinylation Induced by High-Fat Diet Suppresses DNA Damage Repair. *Cell Reports* **25**, 398–412.e6 (2018).
88. Arner, E. *et al.* Ceruloplasmin Is a Novel Adipokine Which Is Overexpressed in Adipose Tissue of Obese Subjects and in Obesity-Associated Cancer Cells. *PLoS ONE* **9,** e80274 (2014).
89. Blüher, M. *et al.* Fas andFasLExpression in Human Adipose Tissue Is Related to Obesity, Insulin Resistance, and Type 2 Diabetes. *The Journal of Clinical Endocrinology & Metabolism* **99**, E36–E44 (2014).
90. Lee, Y. S. *et al.* Increased adipocyte O2 consumption triggers HIF-1α, causing inflammation and insulin resistance in obesity. *Cell* **157**, 1339–1352 (2014).
91. García-Alonso, V. *et al.* Prostaglandin E2 Exerts Multiple Regulatory Actions on Human Obese Adipose Tissue Remodeling, Inflammation, Adaptive Thermogenesis and Lipolysis. *PLoS One* **11**, e0153751 (2016).
92. Cipolletta, D. *et al.* PPAR-γ is a major driver of the accumulation and phenotype of adipose tissue Treg cells. *Nature* **486**, 549–553 (2012).
93. Castaño, C., Kalko, S., Novials, A. & Párrizas, M. Obesity-associated exosomal miRNAs modulate glucose and lipid metabolism in mice. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12158–12163 (2018).
94. Sun, B. *et al.* Adipose tissue-secreted miR-27a promotes liver cancer by targeting

FOXO1 in obese individuals. *OncoTargets and Therapy* **735**, 735–744 (2015).

95. Ren, H. *et al.* Leptin upregulates telomerase activity and transcription of human telomerase reverse transcriptase in MCF-7 breast cancer cells. *Biochem. Biophys. Res. Commun.* **394**, 59–63 (2010).
96. Zwezdaryk, K., Sullivan, D. & Saifudeen, Z. The p53/Adipose-Tissue/Cancer Nexus. *Front. Endocrinol.* **9**, 457 (2018).
97. Toro, A. R. *et al.* Leptin Is an Anti-Apoptotic Effector in Placental Cells Involving p53 Downregulation. *PLoS One* **9**, e99187 (2014).
98. Kung, C.-P. *et al.* The P72R Polymorphism of p53 Predisposes to Obesity and Metabolic Dysfunction. *Cell Rep.* **14**, 2413–2425 (2016).
99. Di Giovannantonio, M. *et al.* Heritable genetic variants in key cancer genes link cancer risk with anthropometric traits. *J. Med. Genet.* **6**, 392-399 (2021).
100. Pópulo, H., Lopes, J. M. & Soares, P. The mTOR Signalling Pathway in Human Cancer. *International Journal of Molecular Sciences* **13**, 1886–1918 (2012).
101. Grahame Hardie, D., Ross, F. A. & Hawley, S. A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat. Rev. Mol. Cell Biol.* **13**, 251–262 (2012).
102. Chen, C., Chang, Y.-C., Liu, C.-L., Chang, K.-J. & Guo, I.-C. Leptin-induced growth of human ZR-75-1 breast cancer cells is associated with up-regulation of cyclin D1 and c-Myc and down-regulation of tumor suppressor p53 and p21WAF1/CIP1. *Breast Cancer Res. Treat.* **98**, 121–132 (2006).
103. Chen, C., Chang, Y.-C., Liu, C.-L., Chang, K.-J. & Guo, I.-C. Leptin-induced growth of human ZR-75-1 breast cancer cells is associated with up-regulation of cyclin D1 and c-Myc and down-regulation of tumor suppressor p53 and p21WAF1/CIP1. *Breast Cancer Res. Treat.* **98**, 121–132 (2006).
104. Foster, P. A. Oestrogen and colorectal cancer: mechanisms and controversies. *Int. J. Colorectal Dis.* **28**, 737–749 (2013).
105. Rodriguez, A. C., Blanchard, Z., Maurer, K. A. & Gertz, J. Estrogen Signaling in

Endometrial Cancer: a Key Oncogenic Pathway with Several Open Questions. *Horm. Cancer* **10**, 51 (2019).

106. Rossouw, J. E. *et al.* Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* **288**, (2002).
107. Sasso, C. V. *et al.* Estradiol and progesterone regulate proliferation and apoptosis in colon cancer. *Endocrine Connections* **8**, 217–229 (2019).
108. Konduri, S. & Schwarz, R. E. Estrogen Receptor β/α Ratio Predicts Response of Pancreatic Cancer Cells to Estrogens and Phytoestrogens. *Journal of Surgical Research* **140**, 55–66 (2007).
109. Sheetal Parida, D. S. The Microbiome–Estrogen Connection and Breast Cancer Risk. *Cells* **8**, (2019).
110. Han, L. *et al.* High Glucose Promotes Pancreatic Cancer Cell Proliferation via the Induction of EGF Expression and Transactivation of EGFR. *PLoS ONE* vol. 6 e27074 (2011).
111. Lopez, R. *et al.* Hyperglycemia enhances the proliferation of non-tumorigenic and malignant mammary epithelial cells through increased leptin/IGF1R signaling and activation of AKT/mTOR. *PLoS One* **8**, e79708 (2013).
112. Li, W. *et al.* Effects of hyperglycemia on the progression of tumor diseases. *J. Exp. Clin. Cancer Res.* **38**, 327 (2019).
113. Park, J., Sarode, V. R., Euhus, D., Kittler, R. & Scherer, P. E. Neuregulin 1-HER axis as a key mediator of hyperglycemic memory effects in breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21058 (2012).
114. Shen, J. *et al.* Oncogenic mutations and dysregulated pathways in obesity-associated hepatocellular carcinoma. *Oncogene* **35**, 6271–6280 (2016).
115. Gallagher, E. J. & LeRoith, D. Hyperinsulinaemia in cancer. *Nat. Rev. Cancer* **20**, 629–644 (2020).

116. D'Esposito, V. *et al.* Adipocyte-released insulin-like growth factor-1 is regulated by glucose and fatty acids and controls breast cancer cell growth in vitro. *Diabetologia* **55**, 2811–2822 (2012).
117. Ling, S. *et al.* Association of Type 2 Diabetes With Cancer: A Meta-analysis With Bias Analysis for Unmeasured Confounding in 151 Cohorts Comprising 32 Million People. *Diabetes Care* **43**, 2313–2322 (2020).
118. Sanaki, Y., Nagata, R., Kizawa, D., Léopold, P. & Igaki, T. Hyperinsulinemia Drives Epithelial Tumorigenesis by Abrogating Cell Competition. *Dev. Cell* **53**, 379–389.e5 (2020).
119. Kapil, S. *et al.* The cell polarity protein Scrib functions as a tumor suppressor in liver cancer. *Oncotarget* **8**, 26515 (2017).
120. Nam, S. Y. *et al.* Effect of obesity on total and free insulin-like growth factor (IGF)-1, and their relationship to IGF-binding protein (BP)-1, IGFBP-2, IGFBP-3, insulin, and growth hormone. *Int. J. Obes. Relat. Metab. Disord.* **21**, 355-359 (1997).
121. Frystyk, J., Skjaerbaek, C., Vestbo, E., Fisker, S. & Orskov, H. Circulating levels of free insulin-like growth factors in obese subjects: the impact of type 2 diabetes. *Diabetes. Metab. Res. Rev.* **15**, 314-322 (1999).
122. Kang, H. J. *et al.* BRCA1 negatively regulates IGF-1 expression through an estrogen-responsive element-like site. *Cell Death Dis.* **3**, e336 (2012).
123. Fagan, D. H. & Yee, D. Crosstalk Between IGF1R and Estrogen Receptor Signaling in Breast Cancer. *J. Mammary Gland Biol. Neoplasia* **13**, 423–429 (2008).
124. Morrione, A. *et al.* Insulin-like growth factor II stimulates cell proliferation through the insulin receptor. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 3777 (1997).
125. IGF2 signaling and regulation in cancer. *Cytokine Growth Factor Rev.* **26**, 371–377 (2015).
126. Porta, C. *et al.* Circulating interleukin-6 as a tumor marker for hepatocellular carcinoma. *Annals of Oncology* **19**, 353–358 (2008).

127. Park, E. J. *et al.* Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. *Cell* **140**, 197 (2010).
128. Xu, H., Uysal, K. T., Becherer, J. D., Arner, P. & Hotamisligil, G. S. Altered tumor necrosis factor-alpha (TNF-alpha) processing in adipocytes and increased expression of transmembrane TNF-alpha in obesity. *Diabetes* **51**, 1876–1883 (2002).
129. Hao, J. *et al.* Circulating Adipose Fatty Acid Binding Protein Is a New Link Underlying Obesity-Associated Breast/Mammary Tumor Development. *Cell Metab.* **28**, 689–705.e5 (2018).
130. Suvannasankha, A. *et al.* FGF23 is elevated in multiple myeloma and increases heparanase expression by tumor cells. *Oncotarget* **6**, 19647–19660 (2015).
131. Kennedy, B. E. *et al.* NAD salvage pathway in cancer metabolism and therapy. *Pharmacological Research* **114**, 274–283 (2016).
132. Venkateswaran, N. & Conacci-Sorrell, M. Kynurenine: an oncometabolite in colon cancer. *Cell stress & chaperones* **4**, 24–26 (2020).
133. Heng, B. *et al.* Understanding the role of the kynurenine pathway in human breast cancer immunobiology. *Oncotarget* **7**, 6506–6520 (2016).
134. Opitz, C. A. *et al.* An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* **478**, 197–203 (2011).
135. Kadowaki, T. & Yamauchi, T. Adiponectin and adiponectin receptors. *Endocr. Rev.* **26**, 439–451 (2005).
136. Ziemke, F. & Mantzoros, C. S. Adiponectin in insulin resistance: lessons from translational research. *Am. J. Clin. Nutr.* **91**, 258S–261S (2010).
137. Kishida, K. *et al.* Relationships between circulating adiponectin levels and fat distribution in obese subjects. *J. Atheroscler. Thromb.* **18**, 592–595 (2011).
138. Chang, E. *et al.* Restoration of adiponectin expression via the ERK pathway in TNFα-treated 3T3-L1 adipocytes. *Mol. Med. Rep.* **10**, 905–910 (2014).
139. Cust, A. E. *et al.* Plasma adiponectin levels and endometrial cancer risk in pre- and

postmenopausal women. *J. Clin. Endocrinol. Metab.* **92**, 255-263 (2007).

140. Macis, D., Guerrieri-Gonzaga, A. & Gandini, S. Circulating adiponectin and breast cancer risk: a systematic review and meta-analysis. *Int. J. Epidemiol.* **43**, 1226–1236 (2014).

141. Brochu-Gaudreau, K. *et al.* Adiponectin action from head to toe. *Endocrine* **37**, 11–32 (2009).

142. Saxena, N. K. *et al.* Adiponectin modulates C-jun N-terminal kinase and mammalian target of rapamycin and inhibits hepatocellular carcinoma. *Gastroenterology* **139**, 1762–73, 1773.e1–5 (2010).

143. Arditi, J., Venihaki, M., Karalis, K. & Chrousos, G. Antiproliferative Effect of Adiponectin on MCF7 Breast Cancer Cells: A Potential Hormonal Link between Obesity and Cancer. *Hormone and Metabolic Research* **39**, 9–13 (2007).

144. Dos Santos, E. *et al.* Adiponectin mediates an antiproliferative response in human MDA-MB 231 breast cancer cells. *Oncol. Rep.* **20**, 971–977 (2008).

145. Cong, L. *et al.* Human adiponectin inhibits cell growth and induces apoptosis in human endometrial carcinoma cells, HEC-1-A and RL95 2. *Endocr. Relat. Cancer* **14**, 713-720 (2007).

146. Park, J., Euhus, D. M. & Scherer, P. E. Paracrine and endocrine effects of adipose tissue on cancer development and progression. *Endocr. Rev.* **32**, 550–570 (2011).

147. Dagon, Y. *et al.* p70S6 Kinase Phosphorylates AMPK on Serine 491 to Mediate Leptin's Effect on Food Intake. *Cell Metabolism* **16**, 104–112 (2012).

148. Mauro, L. *et al.* Uncoupling effects of estrogen receptor α on LKB1/AMPK interaction upon adiponectin exposure in breast cancer. *FASEB J.* **32**, 4343–4355 (2018).

149. Yoon, M. J. *et al.* Adiponectin Increases Fatty Acid Oxidation in Skeletal Muscle Cells by Sequential Activation of AMP-Activated Protein Kinase, p38 Mitogen-Activated Protein Kinase, and Peroxisome Proliferator–Activated Receptor α. *Diabetes* **55**, 2562–2570 (2006).

150. Hoxhaj, G. & Manning, B. D. The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. *Nat. Rev. Cancer* **20**, 74–88 (2020).
151. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
152. Majdalawieh, A. F., Massri, M. & Ro, H.-S. AEBP1 is a Novel Oncogene: Mechanisms of Action and Signaling Pathways. *J. Oncol.* **2020**, 8097872 (2020).
153. Akhurst, R. J. & Derynck, R. TGF-β signaling in cancer — a double-edged sword. *Trends in Cell Biology* **11**, S44–S51 (2001).
154. Samad, F., Yamamoto, K., Pandey, M. & Loskutoff, D. J. Elevated expression of transforming growth factor-beta in adipose tissue from obese mice. *Mol. Med.* **3**, 37–48 (1997).
155. Seoane, J. & Gomis, R. R. TGF-β Family Signaling in Tumor Suppression and Cancer Progression. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).
156. Mishra, A. K., Parish, C. R., Wong, M.-L., Licinio, J. & Blackburn, A. C. Leptin signals via TGFB1 to promote metastatic potential and stemness in breast cancer. *PLoS One* **12**, e0178454 (2017).
157. Forrest, C. M., McNair, K., Vincenten, M. C. J., Gail Darlington, L. & Stone, T. W. Selective depletion of tumour suppressors Deleted in Colorectal Cancer (DCC) and neogenin by environmental and endogenous serine proteases: linking diet and cancer. *BMC Cancer* **16**, 772 (2016).
158. Sun, X. *et al.* Proprotein convertase subtilisin/kexin type 9 deficiency reduces melanoma metastasis in liver. *Neoplasia* **14**, 1122–1131 (2012).
159. Perera, C. N., Chin, H. G., Duru, N. & Camarillo, I. G. Leptin-regulated gene expression in MCF-7 breast cancer cells: mechanistic insights into leptin-regulated mammary tumor growth and progression. *Journal of Endocrinology* **199**, 499 (2008).
160. Jardé, T., Perrier, S., Vasson, M.-P. & Caldefie-Chézet, F. Molecular mechanisms of leptin and adiponectin in breast cancer. *European Journal of Cancer* **47**, 33–43 (2011).

161. Nepal, S. *et al.* Autophagy induction by leptin contributes to suppression of apoptosis in cancer cells and xenograft model: Involvement of p53/FoxO3A axis. *Oncotarget* **6**, 7166–7181 (2015).

162. Ptak, A., Kolaczkowska, E. & Gregoraszczuk, E. L. Leptin stimulation of cell cycle and inhibition of apoptosis gene and protein expression in OVCAR-3 ovarian cancer cells. *Endocrine* **43**, 394–403 (2013).

163. Leze, E. *et al.* Leptin Regulates Proliferation and Apoptosis in Human Prostate. *The Scientific World Journal* **2012**, 1–6 (2012).

164. Russo, V. C., Metaxas, S., Kobayashi, K., Harris, M. & Werther, G. A. Antiapoptotic effects of leptin in human neuroblastoma cells. *Endocrinology* **145**, 4103–4112 (2004).

165. Hilmi, C. *et al.* IGF1 Promotes Resistance to Apoptosis in Melanoma Cells through an Increased Expression of BCL2, BCL-X(L), and Survivin. *Journal of Investigative Dermatology* **128**, 1499–1505 (2008).

166. Leri, A. *et al.* Insulin-Like Growth Factor-1 Induces Mdm2 and Down-Regulates p53, Attenuating the Myocyte Renin-Angiotensin System and Stretch-Mediated Apoptosis. *The American Journal of Pathology* **154**, 567–580 (1999).

167. Park, J. & Scherer, P. E. Adipocyte-derived endotrophin promotes malignant tumor progression. *J. Clin. Invest.* **122**, 4243–4256 (2012).

168. Cinti, S. *et al.* Adipocyte death defines macrophage localization and function in adipose tissue of obese mice and humans. *Journal of Lipid Research* **46**, 2347–2355 (2005).

169. Wang, L., Zhao, R.-P., Song, X.-Y. & Wu, W.-F. Targeting ERβ in Macrophage Reduces Crown-like Structures in Adipose Tissue by Inhibiting Osteopontin and HIF-1α. *Sci. Rep.* **9**, 1–11 (2019).

170. Wei, L. H. *et al.* The anti-apoptotic role of interleukin-6 in human cervical cancer is mediated by up-regulation of Mcl-1 through a PI 3-K/Akt pathway. *Oncogene* **20**, 5799-5809 (2001).

171. Lin, M. T., Juan, C. Y., Chang, K. J., Chen, W. J. & Kuo, M. L. IL-6 inhibits apoptosis

and retains oxidative DNA lesions in human gastric cancer AGS cells through up-regulation of anti-apoptotic gene mcl-1. *Carcinogenesis* **22**, 1947-1953 (2001).

172. Chen, R.-H., Chang, M.-C., Su, Y.-H., Tsai, Y.-T. & Kuo, M.-L. Interleukin-6 Inhibits Transforming Growth Factor-β-induced Apoptosis through the Phosphatidylinositol 3-Kinase/Akt and Signal Transducers and Activators of Transcription 3 Pathways. *Journal of Biological Chemistry* **274**, 23013–23019 (1999).
173. Trayhurn, P. Hypoxia and Adipose Tissue Function and Dysfunction in Obesity. *Physiological Reviews* **93**, 1–21 (2013).
174. Harris, A. L. Hypoxia — a key regulatory factor in tumour growth. *Nature Reviews Cancer* **2**, 38–47 (2002).
175. Harris, B. H. L., Barberis, A., West, C. M. L. & Buffa, F. M. Gene Expression Signatures as Biomarkers of Tumour Hypoxia. *Clinical Oncology* **27**, 547–560 (2015).
176. Forsythe, J. A. *et al.* Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Molecular and Cellular Biology* **16**, 4604–4613 (1996).
177. Pidgeon, G. P., Barr, M. P., Harmey, J. H., Foley, D. A. & Bouchier-Hayes, D. J. Vascular endothelial growth factor (VEGF) upregulates BCL-2 and inhibits apoptosis in human and murine mammary adenocarcinoma cells. *British Journal of Cancer* **85**, 273–278 (2001).
178. Hardie, D. G. AMPK and autophagy get connected. *EMBO J.* **30**, 634–635 (2011).
179. Dann, S. G., Selvaraj, A. & Thomas, G. mTOR Complex1–S6K1 signaling: at the crossroads of obesity, diabetes and cancer. *Trends in Molecular Medicine* **13**, 252–259 (2007).
180. Wen, H. *et al.* Fatty acid–induced NLRP3-ASC inflammasome activation interferes with insulin signaling. *Nature Immunology* **12**, 408–415 (2011).
181. Li, S. *et al.* Sirtuin 3 acts as a negative regulator of autophagy dictating hepatocyte susceptibility to lipotoxicity. *Hepatology* **66**, 936–952 (2017).
182. Yang, L., Li, P., Fu, S., Calay, E. S. & Hotamisligil, G. S. Defective hepatic autophagy in

obesity promotes ER stress and causes insulin resistance. *Cell Metab.* **11**, 467–478 (2010).

183. Chung, S. J. *et al.* ADIPOQ/adiponectin induces cytotoxic autophagy in breast cancer cells through STK11/LKB1-mediated activation of the AMPK-ULK1 axis. *Autophagy* **13,** 1386–1403 (2017).

184. Jansen, H. J. *et al.* Autophagy activity is up-regulated in adipose tissue of obese individuals and modulates proinflammatory cytokine expression. *Endocrinology* **153**, 5866–5874 (2012).

185. Kovsan, J. *et al.* Altered autophagy in human adipose tissues in obesity. *J. Clin. Endocrinol. Metab.* **96**, E268–77 (2011).

186. Nuñez, C. E. *et al.* Defective regulation of adipose tissue autophagy in obesity. *Int. J. Obes.* **37**, 1473–1480 (2013).

187. Bergsbaken, T., Fink, S. L. & Cookson, B. T. Pyroptosis: host cell death and inflammation. *Nat. Rev. Microbiol.* **7**, (2009).

188. Giordano, A. *et al.* Obese adipocytes show ultrastructural features of stressed cells and die of pyroptosis. *Journal of Lipid Research* **54**, 2423–2436 (2013).

189. Wang, L. *et al.* Metformin induces human esophageal carcinoma cell pyroptosis by targeting the miR-497/PELP1 axis. *Cancer Lett.* **450**, 22–31 (2019).

190. Hangel, C. Leptin-mediated downregulation of glutathione peroxidase 4 and iron overload contributing to podocyte ferroptosis in diabetic nephropathy. (2019). doi:10.11588/heidok.00027328.

191. Viswanathan, V. S. *et al.* Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. *Nature* **547**, 453–457 (2017).

192. Dixon, S. J. & Stockwell, B. R. The Hallmarks of Ferroptosis. *Annual Review of Cancer Biology* **3**, 35–54 (2019).

193. Fallarino, F. *et al.* T cell apoptosis by tryptophan catabolism. *Cell Death & Differentiation* **9**, 1069–1077 (2002).

194. Song, H. *et al.* l-Kynurenine-induced apoptosis in human NK cells is mediated by reactive oxygen species. *International Immunopharmacology* **11**, 932–938 (2011).
195. Bishnupuri, K. S. *et al.* IDO1 and Kynurenine Pathway Metabolites Activate PI3K-Akt Signaling in the Neoplastic Colon Epithelium to Promote Cancer Cell Proliferation and Inhibit Apoptosis. *Cancer Research* **79**, 1138–1150 (2019).
196. Au Yeung, C. L. *et al.* Exosomal transfer of stroma-derived miR21 confers paclitaxel resistance in ovarian cancer cells through targeting APAF1. *Nat. Commun.* **7**, (2016).
197. Akincilar, S. C., Unal, B. & Tergaonkar, V. Reactivation of telomerase in cancer. *Cellular and Molecular Life Sciences* **73**, 1659–1670 (2016).
198. Rahmati-Yamchi, M., Zarghami, N., Rahbani, M. & Montazeri, A. Plasma Leptin, hTERT Gene Expression, and Anthropometric Measures in Obese and Non-Obese Women with Breast Cancer. *Breast Cancer* **5**, 27–35 (2011).
199. Stefanou, N., Papanikolaou, V., Furukawa, Y., Nakamura, Y. & Tsezou, A. Leptin as a critical regulator of hepatocellular carcinoma development through modulation of human telomerase reverse transcriptase. *BMC Cancer* **10**, 442 (2010).
200. Mohammadi, M., Hedayati, M., Zarghami, N. & Ghaemmaghami, S. Resistin effect on telomerase gene expression in gastric cancer cell line AGS. *Acta Endocrinol.* **12**, 145-149 (2016).
201. Wetterau, L. A., Francis, M. J., Ma, L. & Cohen, P. Insulin-like growth factor I stimulates telomerase activity in prostate cancer cells. *J. Clin. Endocrinol. Metab.* **88**, 3354-3359 (2003).
202. Matsumoto, R. et al. Accelerated Telomere Shortening in Acromegaly; IGF-I Induces Telomere Shortening and Cellular Senescence. PLoS One 10, e0140189 (2015).
203. Kanaya, T. *et al.* Adenoviral expression of p53 represses telomerase activity through down-regulation of human telomerase reverse transcriptase transcription. *Clin. Cancer Res.* **6**, 1239–1247 (2000).
204. Gielen, M. *et al.* Body mass index is negatively associated with telomere length: a

collaborative cross-sectional meta-analysis of 87 observational studies. *Am. J. Clin. Nutr.* **108**, 453–475 (2018).

205. Mangge, H. *et al.* Telomere shortening associates with elevated insulin and nuchal fat accumulation. *Sci. Rep.* **10**, 1–9 (2020).

206. Włodarczyk, M. & Nowicka, G. Obesity, DNA Damage, and Development of Obesity-Related Diseases. *Int. J. Mol. Sci.* **20**, (2019).

207. Li, J., Yu, X., Pan, W. & Unger, R. H. Gene expression profile of rat adipose tissue at the onset of high-fat-diet obesity. *American Journal of Physiology-Endocrinology and Metabolism* **282,** E1334–E1341 (2002).

208. Cao, Y. Angiogenesis modulates adipogenesis and obesity. *J. Clin. Invest.* **117**, 2362–2368 (2007).

209. Incio, J. *et al.* PlGF/VEGFR-1 Signaling Promotes Macrophage Polarization and Accelerated Tumor Progression in Obesity. *Clin. Cancer Res.* **22**, 2993–3004 (2016).

210. Falco, S. D. & De Falco, S. Structure and Function of Placental Growth Factor. *Trends in Cardiovascular Medicine* **12**, 241–246 (2002).

211. Favia, A. *et al.* VEGF-induced neoangiogenesis is mediated by NAADP and two-pore channel-2-dependent Ca2+ signaling. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4706–15 (2014).

212. Cao, R., Brakenhielm, E., Wahlestedt, C., Thyberg, J. & Cao, Y. Leptin induces vascular permeability and synergistically stimulates angiogenesis with FGF-2 and VEGF. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6390–6395 (2001).

213. Yang, W.-H. *et al.* Leptin promotes VEGF-C production and induces lymphangiogenesis by suppressing miR-27b in human chondrosarcoma cells. *Sci. Rep.* **6**, 28647 (2016).

214. Yang, W.-H. *et al.* Leptin increases VEGF expression and enhances angiogenesis in human chondrosarcoma cells. *Biochim. Biophys. Acta* **1840**, 3483–3493 (2014).

215. Lanier, V. *et al.* Leptin-induced transphosphorylation of vascular endothelial growth factor receptor increases Notch and stimulates endothelial cell angiogenic

transformation. *Int. J. Biochem. Cell Biol.* **79**, 139–150 (2016).

216. Gonzalez, R. R. *et al.* Leptin signaling promotes the growth of mammary tumors and increases the expression of vascular endothelial growth factor (VEGF) and its receptor type two (VEGF-R2). *J. Biol. Chem.* **281**, 26320–26328 (2006).

217. Carmeliet, P. *et al.* Synergism between vascular endothelial growth factor and placental growth factor contributes to angiogenesis and plasma extravasation in pathological conditions. *Nat. Med.* **7**, 575–583 (2001).

218. Lijnen, H. R. *et al.* Impaired adipose tissue development in mice with inactivation of placental growth factor function. *Diabetes* **55**, 2698–2704 (2006).

219. Schölch, S. *et al.* Serum PlGF and EGF are independent prognostic markers in non-metastatic colorectal cancer. *Sci. Rep.* **9**, 10921 (2019).

220. Kopp, C. W. *et al.* Weight loss reduces tissue factor in morbidly obese patients. *Obes. Res.* **11**, 950–956 (2003).

221. Zhang, Y. *et al.* Tissue factor controls the balance of angiogenic and antiangiogenic properties of tumor cells in mice. *J. Clin. Invest.* **94**, 1320–1327 (1994).

222. Khorana, A. A. *et al.* Tissue Factor Expression, Angiogenesis, and Thrombosis in Pancreatic Cancer. *Clinical Cancer Research* **13**, 2870–2875 (2007).

223. van den Berg, Y. W. *et al.* Alternatively spliced tissue factor induces angiogenesis through integrin ligation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19497–19502 (2009).

224. Toiyama, Y. *et al.* Elevated serum angiopoietin-like protein 2 correlates with the metastatic properties of colorectal cancer: a serum biomarker for early diagnosis and recurrence. *Clin. Cancer Res.* **20**, 6175–6186 (2014).

225. Wang, X., Hu, Z., Wang, Z., Cui, Y. & Cui, X. Angiopoietin-like protein 2 is an important facilitator of tumor proliferation, metastasis, angiogenesis and glycolysis in osteosarcoma. *Am. J. Transl. Res.* **11**, 6341 (2019).

226. Richardson, M. R. *et al.* Angiopoietin-like protein 2 regulates endothelial colony forming cell vasculogenesis. *Angiogenesis* **17**, 675–683 (2014).

227. Guerrero, P. A. & McCarty, J. H. TGF-β Activation and Signaling in Angiogenesis. *Physiologic and Pathologic Angiogenesis - Signaling Mechanisms and Targeted Therapy* (2017) doi:10.5772/66405.

228. Hong, H. *et al.* TNF-alpha promotes lymphangiogenesis and lymphatic metastasis of gallbladder cancer through the ERK1/2/AP-1/VEGF-D pathway. *BMC Cancer* **16**, 240 (2016).

229. Bråkenhielm, E. *et al.* Adiponectin-induced antiangiogenesis and antitumor activity involve caspase-mediated endothelial cell apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2476–2481 (2004).

230. Man, K. *et al.* Suppression of liver tumor growth and metastasis by adiponectin in nude mice through inhibition of tumor angiogenesis and downregulation of Rho kinase/IFN-inducible protein 10/matrix metalloproteinase 9 signaling. *Clin. Cancer Res.* **16**, 967–977 (2010).

231. Wang, H. *et al.* The Function of the HGF/c-Met Axis in Hepatocellular Carcinoma. *Frontiers in Cell and Developmental Biology* **8**, (2020).

232. Umbrello, M., Dyson, A., Feelisch, M. & Singer, M. The Key Role of Nitric Oxide in Hypoxia: Hypoxic Vasodilation and Energy Supply–Demand Matching. *Antioxidants & Redox Signaling* **19**, 1690–1710 (2013).

233. Ziche, M. & Morbidelli, L. Nitric oxide and angiogenesis. *J. Neurooncol.* **50**, (2000).

234. Gallo, O. *et al.* Role of nitric oxide in angiogenesis and tumor progression in head and neck cancer. *J. Natl. Cancer Inst.* **90**, 587–596 (1998).

235. Gao, Y. *et al.* Nitric oxide synthase inhibitors 1400W and L-NIO inhibit angiogenesis pathway of colorectal cancer. *Nitric Oxide* **83**, 33–39 (2019).

236. Sessa, W. C. Nitric oxide synthase inhibition and cancer. *The Lancet Oncology* **8**, 88–89 (2007).

237. Choi, J. W. *et al.* Increases in nitric oxide concentrations correlate strongly with body fat in obese humans. *Clin. Chem.* **47**, 1106–1109 (2001).

238. Gruber, H.-J. *et al.* Obesity reduces the bioavailability of nitric oxide in juveniles. *Int. J. Obes.* **32**, 826–831 (2008).

239. Holguin, F. Arginine and nitric oxide pathways in obesity-associated asthma. *J. Allergy* **2013**, 714595 (2013).

240. Cignarelli, M. *et al.* Relationship of obesity and body fat distribution with ceruloplasmin serum levels. *Int. J. Obes. Relat. Metab. Disord.* **20**, (1996).

241. Roman, A. C., Carvajal-Gonzalez, J. M., Rico-Leo, E. M. & Fernandez-Salguero, P. M. Dioxin Receptor Deficiency Impairs Angiogenesis by a Mechanism Involving VEGF-A Depletion in the Endothelium and Transforming Growth Factor-β Overexpression in the Stroma. *Journal of Biological Chemistry* **284**, 25135–25148 (2009).

242. Maisonpierre, P. C. *et al.* Angiopoietin-2, a natural antagonist for Tie2 that disrupts in vivo angiogenesis. *Science* **277**, 55–60 (1997).

243. Ye, J. Adipose Tissue Vascularization: Its Role in Chronic Inflammation. *Curr. Diab. Rep.* **11**, 203 (2011).

244. Hosogai, N. *et al.* Adipose Tissue Hypoxia in Obesity and Its Impact on Adipocytokine Dysregulation. *Diabetes* **56**, 901–911 (2007).

245. Gobble, R. M., Groesch, K. A., Chang, M., Torry, R. J. & Torry, D. S. Differential regulation of human PlGF gene expression in trophoblast and nontrophoblast cells by oxygen tension. *Placenta* **30**, 869–875 (2009).

246. Mukhopadhyay, C. K., Mazumder, B. & Fox, P. L. Role of Hypoxia-inducible Factor-1 in Transcriptional Activation of Ceruloplasmin by Iron Deficiency. *J. Biol. Chem.* **275**, 21048–21054 (2000).

247. Kuett, A. *et al.* IL-8 as mediator in the microenvironment-leukaemia network in acute myeloid leukaemia. *Sci. Rep.* **5**, 18411 (2015).

248. Kubo, H. *et al.* Regulation and clinical significance of the hypoxia-induced expression of ANGPTL4 in gastric cancer. *Oncol. Lett.* **11**, 1026–1034 (2016).

249. Langley, R. R. & Fidler, I. J. The seed and soil hypothesis revisited--the role of tumor-

stroma interactions in metastasis to different organs. *Int. J. Cancer* **128**, 2527–2535 (2011).

250. Choi, S. S. *et al.* Leptin promotes the myofibroblastic phenotype in hepatic stellate cells by activating the hedgehog pathway. *J. Biol. Chem.* **285**, 36551–36560 (2010).
251. Wang, L. *et al.* Activation of IL-8 via PI3K/Akt-dependent pathway is involved in leptin-mediated epithelial-mesenchymal transition in human breast cancer cells. *Cancer Biol. Ther.* **16**, 1220–1230 (2015).
252. Ray, A. & Cleary, M. P. The potential role of leptin in tumor invasion and metastasis. *Cytokine Growth Factor Rev.* **38**, 80–97 (2017).
253. Strong, A. L. *et al.* Leptin produced by obese adipose stromal/stem cells enhances proliferation and metastasis of estrogen receptor positive breast cancers. *Breast Cancer Res.* **17**, 112 (2015).
254. Avtanski, D. *et al.* Resistin induces breast cancer cells epithelial to mesenchymal transition (EMT) and stemness through both adenylyl cyclase-associated protein 1 (CAP1)-dependent and CAP1-independent mechanisms. *Cytokine* **120**, 155–164 (2019).
255. Saxena, N. K. & Sharma, D. Metastasis suppression by adiponectin: LKB1 rises up to the challenge. *Cell Adh. Migr.* **4**, 358–362 (2010).
256. Lu, C.-C. *et al.* Insulin induction instigates cell proliferation and metastasis in human colorectal cancer cells. *Int. J. Oncol.* **50**, 736–744 (2017).
257. Zhang, H. *et al.* Inhibition of cancer cell proliferation and metastasis by insulin receptor downregulation. *Oncogene* **29**, 2517–2527 (2010).
258. Chiu, H.-Y. *et al.* Effects of hyperinsulinemia on acquired resistance to epidermal growth factor receptor-tyrosine kinase inhibitor via the PI3K/AKT pathway in non-small cell lung cancer cells in vitro. *Oncol. Lett.* **20**, (2020).
259. Wen, Y.-A. *et al.* Adipocytes activate mitochondrial fatty acid oxidation and autophagy to promote tumor growth in colon cancer. *Cell Death Dis.* **8**, e2593 (2017).

260. D'Esposito, V. *et al.* Adipose microenvironment promotes triple negative breast cancer cell invasiveness and dissemination by producing CCL5. *Oncotarget* **7**, 24495–24509 (2016).

261. Walter, M., Liang, S., Ghosh, S., Hornsby, P. J. & Li, R. Interleukin 6 secreted from adipose stromal cells promotes migration and invasion of breast cancer cells. *Oncogene* **28**, 2745–2755 (2009).

262. Welte, G. *et al.* Interleukin-8 derived from local tissue-resident stromal cells promotes tumor cell invasion. *Mol. Carcinog.* **51**, 861–868 (2012).

263. Hillers, L. E., D'Amato, J. V., Chamberlin, T., Paderta, G. & Arendt, L. M. Obesity-Activated Adipose-Derived Stromal Cells Promote Breast Cancer Growth and Invasion. *Neoplasia* **20**, 1161–1174 (2018).

264. Seo, B. R. *et al.* Obesity-dependent changes in interstitial ECM mechanics promote breast tumorigenesis. *Sci. Transl. Med.* **7**, 301ra130–301ra130 (2015).

265. Huang, C.-K. *et al.* Adipocytes promote malignant growth of breast tumours with monocarboxylate transporter 2 expression via β-hydroxybutyrate. *Nat. Commun.* **8**, 14706 (2017).

266. Dirat, B. *et al.* Cancer-associated adipocytes exhibit an activated phenotype and contribute to breast cancer invasion. *Cancer Res.* **71**, 2455–2465 (2011).

267. Nieman, K. M. *et al.* Adipocytes promote ovarian cancer metastasis and provide energy for rapid tumor growth. *Nat. Med.* **17**, 1498–1503 (2011).

268. Thomou, T. *et al.* Adipose-derived circulating miRNAs regulate gene expression in other tissues. *Nature* **542**, 450–455 (2017).

269. Lazar, I. *et al.* Adipocyte Exosomes Promote Melanoma Aggressiveness through Fatty Acid Oxidation: A Novel Mechanism Linking Obesity and Cancer. *Cancer Res.* **76**, 4051–4057 (2016).

270. Lin, R., Wang, S. & Zhao, R. C. Exosomes from human adipose-derived mesenchymal stem cells promote migration through Wnt signaling pathway in a breast cancer cell

model. *Molecular and Cellular Biochemistry* **383**, 13–20 (2013).

271. Lope, L. R. de *et al.* Tumour–adipose tissue crosstalk: fuelling tumour metastasis by extracellular vesicles. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20160485 (2018).

272. Endo, M. *et al.* Tumor Cell-Derived Angiopoietin-like Protein ANGPTL2 Is a Critical Driver of Metastasis. *Cancer Research* **72**, 1784–1794 (2012).

273. Horbinski, C., Mojesky, C. & Kyprianou, N. Live free or die: tales of homeless (cells) in cancer. *Am. J. Pathol.* **177**, 1044–1052 (2010).

274. Luey, B. C. & May, F. E. B. Insulin-like growth factors are essential to prevent anoikis in oestrogen-responsive breast cancer cells: importance of the type I IGF receptor and PI3-kinase/Akt pathway. *Mol. Cancer* **15**, 8 (2016).

275. Gauthier, M.-S. *et al.* Decreased AMP-activated protein kinase activity is associated with increased inflammation in visceral adipose tissue and with whole-body insulin resistance in morbidly obese humans. *Biochem. Biophys. Res. Commun.* **404**, 382–387 (2011).

276. Wang, N. *et al.* The IGF-Trap: Novel Inhibitor of Carcinoma Growth and Metastasis. *Mol. Cancer Ther.* **14**, 982–993 (2015).

277. Masuda, T. *et al.* ANGPTL2 increases bone metastasis of breast cancer cells through enhancing CXCR4 signaling. *Sci. Rep.* **5**, 9170 (2015).

278. Wang, Y. *et al.* TNF-α-induced LRG1 promotes angiogenesis and mesenchymal stem cell migration in the subchondral bone during osteoarthritis. *Cell Death Dis.* **8**, e2715 (2017).

279. Zhang, J., Zhu, L., Fang, J., Ge, Z. & Li, X. LRG1 modulates epithelial-mesenchymal transition and angiogenesis in colorectal cancer via HIF-1α activation. *J. Exp. Clin. Cancer Res.* **35**, 29 (2016).

280. Hos, D., Regenfuss, B., Bock, F., Onderka, J. & Cursiefen, C. Blockade of insulin receptor substrate-1 inhibits corneal lymphangiogenesis. *Invest. Ophthalmol. Vis. Sci.* **52**, 5778–5785 (2011).

281. Kang, J. S. *et al.* Interleukin-18 increases metastasis and immune escape of stomach cancer via the downregulation of CD70 and maintenance of CD44. *Carcinogenesis* **30**, 1987–1996 (2009).

282. Marhaba, R. & Zöller, M. CD44 in Cancer Progression: Adhesion, Migration and Growth Regulation. *Journal of Molecular Histology* **35**, 211–231 (2003).

283. Naor, D., Sionov, R. V. & Ish-Shalom, D. CD44: structure, function, and association with the malignant process. *Adv. Cancer Res.* **71**, 241–319 (1997).

284. Olee, T., Hashimoto, S., Quach, J. & Lotz, M. IL-18 is produced by articular chondrocytes and induces proinflammatory and catabolic responses. *J. Immunol.* **162**, 1096–1100 (1999).

285. Krumm, B., Meng, X., Xiang, Y. & Deng, J. Identification of small molecule inhibitors of Interleukin-18. *Sci. Rep.* **7**, 483 (2017).

286. Park, C. C. *et al.* Evidence of IL-18 as a novel angiogenic mediator. *J. Immunol.* **167**, (2001).

287. Druso, J. E. & Fischbach, C. Biophysical Properties of Extracellular Matrix: Linking Obesity and Cancer. *Trends Cancer Res.* **4**, (2018).

288. Bousquenaud, M., Fico, F., Solinas, G., Rüegg, C. & Santamaria-Martínez, A. Obesity promotes the expansion of metastasis-initiating cells in breast cancer. *Breast Cancer Res.* **20**, 104 (2018).

289. Choi, B.-J., Park, S.-A., Lee, S.-Y., Cha, Y. N. & Surh, Y.-J. Hypoxia induces epithelial-mesenchymal transition in colorectal cancer cells through ubiquitin-specific protease 47-mediated stabilization of Snail: A potential role of Sox9. *Sci. Rep.* **7**, 15918 (2017).

290. Imai, T. *et al.* Hypoxia attenuates the expression of E-cadherin via up-regulation of SNAIL in ovarian carcinoma cells. *Am. J. Pathol.* **163**, 1437–1447 (2003).

291. Renaud, S., Guenot, D., Falcoz, P.-E., Massard, G. & Beau-Faller, M. Role of hypoxia in epithelial-to-mesenchymal transition (EMT) in non-small cell lung cancer (NSCLC). *Eur. Respir. J.* **44**, (2014).

292. Wei, L. *et al.* Twist may be associated with invasion and metastasis of hypoxic NSCLC cells. *Tumour Biol.* **37**, 9979–9987 (2016).

293. Zhang, J., Cheng, Q., Zhou, Y., Wang, Y. & Chen, X. Slug is a key mediator of hypoxia induced cadherin switch in HNSCC: correlations with poor prognosis. *Oral Oncol.* **49**, 1043–1050 (2013).

294. D'Amato, N. C. *et al.* A TDO2-AhR signaling axis facilitates anoikis resistance and metastasis in triple-negative breast cancer. *Cancer Res.* **75**, 4651–4664 (2015).

295. Lee, E. *et al.* Breast cancer cells condition lymphatic endothelial cells within pre-metastatic niches to promote metastasis. *Nat. Commun.* **5**, 4715 (2014).

296. Quail, D. F. *et al.* Obesity alters the lung myeloid cell landscape to enhance breast cancer metastasis through IL5 and GM-CSF. *Nat. Cell Biol.* **19**, 974–987 (2017).

297. Snoderly, H. T., Boone, B. A. & Bennewitz, M. F. Neutrophil extracellular traps in breast cancer and beyond: current perspectives on NET stimuli, thrombosis and metastasis, and clinical utility for diagnosis and treatment. *Breast Cancer Res.* **21**, 145 (2019).

298. Yang, L. *et al.* DNA of neutrophil extracellular traps promotes cancer metastasis via CCDC25. *Nature* **583**, 133–138 (2020).

299. Setayesh, T. *et al.* Impact of obesity and overweight on DNA stability: Few facts and many hypotheses. *Mutat. Res.* **777**, 64–91 (2018).

300. Quagliaro, L. *et al.* Intermittent high glucose enhances apoptosis related to oxidative stress in human umbilical vein endothelial cells: the role of protein kinase C and NAD(P)H-oxidase activation. *Diabetes* **52**, 2795–2804 (2003).

301. Jang, Y. *et al.* Visceral fat accumulation determines postprandial lipemic response, lipid peroxidation, DNA damage, and endothelial dysfunction in nonobese Korean men. *J. Lipid Res.* **44**, 2356–2364 (2003).

302. Hansel, D. E. *et al.* Telomere length variation in biliary tract metaplasia, dysplasia, and carcinoma. *Mod. Pathol.* **19**, 772–779 (2006).

303. Roberts, M. J., Wondrak, G. T., Laurean, D. C., Jacobson, M. K. & Jacobson, E. L. DNA

damage by carbonyl stress in human skin cells. *Mutat. Res.* **522**, 45–56 (2003).

304. Morisset, A.-S., Blouin, K. & Tchernof, A. Impact of diet and adiposity on circulating levels of sex hormone-binding globulin and androgens. *Nutr. Rev.* **66**, 506–516 (2008).

305. Dwivedy, I., Devanesan, P., Cremonesi, P., Rogan, E. & Cavalieri, E. Synthesis and characterization of estrogen 2,3- and 3,4-quinones. Comparison of DNA adducts formed by the quinones versus horseradish peroxidase-activated catechol estrogens. *Chem. Res. Toxicol.* **5**, 828–833 (1992).

306. Zhao, Z. *et al.* Mutagenic Activity of 4-Hydroxyestradiol, but Not 2-Hydroxyestradiol, in BB Rat2 Embryonic Cells, and the Mutational Spectrum of 4-Hydroxyestradiol. *Chemical Research in Toxicology* **19**, 475–479 (2006).

307. Tsuchiya, Y., Nakajima, M. & Yokoi, T. Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Lett.* **227**, 115–124 (2005).

308. Azzarà, A., Pirillo, C., Giovannini, C., Federico, G. & Scarpato, R. Different repair kinetic of DSBs induced by mitomycin C in peripheral lymphocytes of obese and normal weight adolescents. *Mutat. Res.* **789**, 9–14 (2016).

309. Wang, X. *et al.* Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med.* **8**, 87 (2010).

310. Milagro, F. I. *et al.* A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss. *FASEB J.* **25**, 1378–1389 (2011).

311. Mijnes, J. *et al.* Promoter methylation of DNA damage repair (DDR) genes in human tumor entities: RBBP8/ CtIP is almost exclusively methylated in bladder cancer. *Clin. Epigenetics* **10**, (2018).

312. Costanzo, V. *et al.* An ATR- and Cdc7-dependent DNA damage checkpoint that inhibits initiation of DNA replication. *Mol. Cell* **11**, 203–213 (2003).

313. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).

314. Cimprich, K. A. & Cortez, D. ATR: an essential regulator of genome integrity. *Nat. Rev. Mol. Cell Biol.* **9**, 616–627 (2008).

315. Kass, E. M., Lim, P. X., Helgadottir, H. R., Moynahan, M. E. & Jasin, M. Robust homology-directed repair within mouse mammary tissue is not specifically affected by Brca2 mutation. *Nat. Commun.* **7**, 13241 (2016).

316. Chen, C.-C., Feng, W., Lim, P. X., Kass, E. M. & Jasin, M. Homology-Directed Repair and the Role of BRCA1, BRCA2, and Related Proteins in Genome Integrity and Cancer. *Annu Rev Cancer Biol* **2**, 313–336 (2018).

317. Heikkinen, K., Mansikka, V., Karppinen, S.-M., Rapakko, K. & Winqvist, R. Mutation analysis of the ATR gene in breast and ovarian cancer families. *Breast Cancer Res.* **7**, R495 (2005).

318. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

319. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).

320. Blackburn, E. H. Structure and function of telomeres. *Nature* vol. 350 569–573 (1991).

321. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883–899 (2010).

322. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).

323. Nagareddy, P. R. *et al.* Adipose tissue macrophages promote myelopoiesis and monocytosis in obesity. *Cell Metab.* **19**, 821–835 (2014).

324. Kratz, M. *et al.* Metabolic dysfunction drives a mechanistically distinct proinflammatory phenotype in adipose tissue macrophages. *Cell Metab.* **20**, 614–625 (2014).

325. Matarese, G., Moschos, S. & Mantzoros, C. S. Leptin in immunology. *J. Immunol.* **174**, (2005).

326. Exley, M. A., Hand, L., O'Shea, D. & Lynch, L. Interplay between the immune system

and adipose tissue in obesity. *Journal of Endocrinology* **223**, 41–48 (2014).

327. Floris, G. *et al.* Body Mass Index and Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancer. *J. Natl. Cancer Inst.* **113**, 146–153 (2021).
328. Tiwari, P. *et al.* Metabolically activated adipose tissue macrophages link obesity to triple-negative breast cancer. *J. Exp. Med.* **216**, 1345–1358 (2019).
329. Kolb, R. *et al.* Obesity-associated NLRC4 inflammasome activation drives breast cancer progression. *Nat. Commun.* **7**, 1–12 (2016).
330. Hotamisligil, G. S. & Erbay, E. Nutrient sensing and inflammation in metabolic diseases. *Nat. Rev. Immunol.* **8**, (2008).
331. Cani, P. D. & Jordan, B. F. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 671–682 (2018).
332. Cipolletta, D., Cohen, P., Spiegelman, B. M., Benoist, C. & Mathis, D. Appearance and disappearance of the mRNA signature characteristic of Treg cells in visceral adipose tissue: age, diet, and PPARγ effects. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 482–487 (2015).
333. Murano, I. *et al.* Dead adipocytes, detected as crown-like structures, are prevalent in visceral fat depots of genetically obese mice. *J. Lipid Res.* **49**, 1562–1568 (2008).
334. Vandanmagsar, B. *et al.* The NLRP3 inflammasome instigates obesity-induced inflammation and insulin resistance. *Nat. Med.* **17**, 179–188 (2011).
335. Bu, D. *et al.* Human endotrophin as a driver of malignant tumor growth. *JCI Insight* **5**, (2019).
336. Castaner, O. *et al.* The Gut Microbiome Profile in Obesity: A Systematic Review. *Int. J. Endocrinol.* **2018**, (2018).
337. Cani, P. D. & Jordan, B. F. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. *Nat. Rev. Gastroenterol. Hepatol.* **15**, 671–682 (2018).
338. Lassenius, M. I. *et al.* Bacterial endotoxin activity in human serum is associated with dyslipidemia, insulin resistance, obesity, and chronic inflammation. *Diabetes Care* **34**,

1809–1815 (2011).

339. Yoshimoto, S. *et al.* Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* **499**, 97–101 (2013).

340. Lo, C.-H. *et al.* Periodontal disease, tooth loss, and risk of oesophageal and gastric adenocarcinoma: a prospective study. *Gut* **70**, 620–621 (2021).

341. Fan, X. *et al.* Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. *Gut* **67**, 120 (2018).

342. Kanneganti, T.-D. & Dixit, V. D. Immunological complications of obesity. *Nature Immunology* **13**, 707–712 (2012).

343. Sheridan, P. A. *et al.* Obesity is associated with impaired immune response to influenza vaccination in humans. *Int. J. Obes.* **36**, 1072–1077 (2012).

344. Vinay, D. S. *et al.* Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Seminars in Cancer Biology* **35**, S185–S198 (2015).

345. Zamarron, B. F. & Chen, W. Dual roles of immune cells and their factors in cancer development and progression. *Int. J. Biol. Sci.* **7**, 651–658 (2011).

346. Andersen, C. J., Murphy, K. E. & Fernandez, M. L. Impact of Obesity and Metabolic Syndrome on Immunity. *Advances in Nutrition* vol. 7 66–75 (2016).

347. Mathis, D. Immunological goings-on in visceral adipose tissue. *Cell Metab.* **17**, 851–859 (2013).

348. Cha, Y. J., Kim, E.-S. & Koo, J. S. Tumor-associated macrophages and crown-like structures in adipose tissue in breast cancer. *Breast Cancer Res. Treat.* **170**, 15–25 (2018).

349. Yang, H. *et al.* Obesity accelerates thymic aging. *Blood* **114**, 3803–3812 (2009).

350. CD36-mediated ferroptosis dampens intratumoral CD8+ T cell effector function and impairs their antitumor ability. *Cell Metab.* **33**, 1001–1012.e5 (2021).

351. de Heredia, F. P., Gómez-Martínez, S. & Marcos, A. Obesity, inflammation and the immune system. *Proc. Nutr. Soc.* **71**, 332-8 (2012).

352. Nojima, I. *et al.* Dysfunction of CD8 + PD-1 + T cells in type 2 diabetes caused by the impairment of metabolism-immune axis. *Sci. Rep.* **10**, 1–12 (2020).

353. Murphy, W. J. & Longo, D. L. The Surprisingly Positive Association Between Obesity and Cancer Immunotherapy Efficacy. *JAMA* **321**, 1247–1248 (2019).

354. Indini, A. *et al.* Impact of BMI on Survival Outcomes of Immunotherapy in Solid Tumors: A Systematic Review. *Int. J. Mol. Sci.* **22**, (2021).

355. Francisco, V. *et al.* Obesity, Fat Mass and Immune System: Role for Leptin. *Front. Physiol.* **9**, 640 (2018).

356. Lo, C. K. C. *et al.* Leptin signaling protects NK cells from apoptosis during development in mouse bone marrow. *Cell. Mol. Immunol.* **6**, 353–360 (2009).

357. Lam, Q. L. K., Liu, S., Cao, X. & Lu, L. Involvement of leptin signaling in the survival and maturation of bone marrow-derived dendritic cells. *Eur. J. Immunol.* **36**, 3118–3130 (2006).

358. Lam, Q. L. K. & Lu, L. Role of leptin in immunity. *Cell. Mol. Immunol.* **4**, 1–13 (2007).

359. Wong, C. K., -Y. Cheung, P. F. & Lam, C. W. K. Leptin-mediated cytokine release and migration of eosinophils: Implications for immunopathophysiology of allergic inflammation. *European Journal of Immunology* **37**, 2337–2348 (2007).

360. Sánchez-Jiménez, F., Pérez-Pérez, A., de la Cruz-Merino, L. & Sánchez-Margalet, V. Obesity and Breast Cancer: Role of Leptin. *Frontiers in Oncology* **9**, 596 (2019).

361. Canter, R. J., Le, C. T., Beerthuijzen, J. M. T. & Murphy, W. J. Obesity as an immune-modifying factor in cancer immunotherapy. *J. Leukoc. Biol.* **104**, 487–497 (2018).

362. Seung, L. P., Rowley, D. A., Dubey, P. & Schreiber, H. Synergy between T-cell immunity and inhibition of paracrine stimulation causes tumor rejection. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6254–6258 (1995).

363. Clements, V. K. *et al.* Frontline Science: High fat diet and leptin promote tumor progression by inducing myeloid-derived suppressor cells. *J. Leukoc. Biol.* **103**, 395–407 (2018).

364. Michelet, X. *et al.* Metabolic reprogramming of natural killer cells in obesity limits antitumor responses. *Nat. Immunol.* **19**, 1330–1340 (2018).

365. Stephens, G. L. *et al.* Kynurenine 3-monooxygenase mediates inhibition of Th17 differentiation via catabolism of endogenous aryl hydrocarbon receptor ligands. *Eur. J. Immunol.* **43**, 1727–1734 (2013).

366. Umansky, V. & Sevko, A. Overcoming immunosuppression in the melanoma microenvironment induced by chronic inflammation. *Cancer Immunol. Immunother.* **61**, 275–282 (2012).

367. Becker, J. C., Andersen, M. H., Schrama, D. & Straten, P. T. Immune-suppressive properties of the tumor microenvironment. *Cancer Immunology, Immunotherapy* **62**,1137–1148 (2013).

368. Ringel, A. E. *et al.* Obesity Shapes Metabolism in the Tumor Microenvironment to Suppress Anti-Tumor Immunity. *Cell* **183**, 1848–1866.e26 (2020).

369. Kaelin, W. G., Jr & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol. Cell* **30**, 393–402 (2008).

370. Zhang, C. *et al.* STAT3 Activation-Induced Fatty Acid Oxidation in CD8 + T Effector Cells Is Critical for Obesity-Promoted Breast Tumor Growth. *Cell Metab.* **31**, 148-161 (2020).

371. Chang, C.-H. *et al.* Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression. *Cell* **162**, 1229–1241 (2015).

372. Baek, A. E. *et al.* The cholesterol metabolite 27 hydroxycholesterol facilitates breast cancer metastasis through its actions on immune cells. *Nat. Commun.* **8**, 1–11 (2017).

373. Mullarky, E. *et al.* Identification of a small molecule inhibitor of 3-phosphoglycerate dehydrogenase to target serine biosynthesis in cancers. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1778–1783 (2016).

374. Seyfried, T. N. Cancer as a Metabolic Disease: Implications for Novel Therapeutics. *AACR Education book* 31–36 (2013).

375. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
376. Bartrons, R. *et al.* Fructose 2,6-Bisphosphate in Cancer Cell Metabolism. *Front. Oncol.* **8**, 331 (2018).
377. Yin, L., Kosugi, M. & Kufe, D. Inhibition of the MUC1-C oncoprotein induces multiple myeloma cell death by down-regulating TIGAR expression and depleting NADPH. *Blood* **119**, 810–816 (2012).
378. Sivridis, E., Giatromanolaki, A. & Koukourakis, M. I. Proliferating fibroblasts at the invading tumour edge of colorectal adenocarcinomas are associated with endogenous markers of hypoxia, acidity, and oxidative stress. *J. Clin. Pathol.* **58**, 1033–1038 (2005).
379. Martinez-Outschoorn, U. E., Lisanti, M. P. & Sotgia, F. Catabolic cancer-associated fibroblasts transfer energy and biomass to anabolic cancer cells, fueling tumor growth. *Seminars in Cancer Biology* **25**, 47–60 (2014).
380. Eales, K. L., Hollinshead, K. E. R. & Tennant, D. A. Hypoxia and metabolic adaptation of cancer cells. *Oncogenesis* **5**, e190–e190 (2016).
381. Dengler, V. L., Galbraith, M. D. & Espinosa, J. M. Transcriptional regulation by hypoxia inducible factors. *Critical Reviews in Biochemistry and Molecular Biology* **49**, 1–15 (2014).
382. Wise, D. R. *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18782–18787 (2008).
383. Luo, W. *et al.* Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1. *Cell* **145**, 732–744 (2011).
384. Yehuda-Shnaidman, E. *et al.* Secreted human adipose leptin decreases mitochondrial respiration in HCT116 colon cancer cells. *PLoS One* **8**, e74843 (2013).
385. Lehuédé, C., Dupuy, F., Rabinovitch, R., Jones, R. G. & Siegel, P. M. Metabolic Plasticity as a Determinant of Tumor Growth and Metastasis. *Cancer Res.* **76**, 5201–

5208 (2016).

386. Rabin-Court, A., Rodrigues, M. R., Zhang, X.-M. & Perry, R. J. Obesity-associated, but not obesity-independent, tumors respond to insulin by increasing mitochondrial glucose oxidation. *PLOS ONE* **14**, e0218126 (2019).

387. Koundouros, N. & Poulogiannis, G. Reprogramming of fatty acid metabolism in cancer. *British Journal of Cancer* **122**, 4–22 (2020).

388. Nomura, D. K. *et al.* Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell* **140**, 49–61 (2010).

389. Motawi, T. K., Shaker, O. G., Ismail, M. F. & Sayed, N. H. Peroxisome Proliferator-Activated Receptor Gamma in Obesity and Colorectal Cancer: the Role of Epigenetics. *Sci. Rep.* **7**, 10714 (2017).

390. Howe, L. R. *et al.* HER2/neu-induced mammary tumorigenesis and angiogenesis are reduced in cyclooxygenase-2 knockout mice. *Cancer Res.* **65**, 10113–10119 (2005).

391. Bowers, L. W. & deGraffenried, L. A. Targeting the COX-2 Pathway to Improve Therapeutic Response in the Obese Breast Cancer Patient Population. *Curr Pharmacol Rep* **1**, 336–345 (2015).

392. van Weverwijk, A. *et al.* Metabolic adaptability in metastatic breast cancer by AKR1B10-dependent balancing of glycolysis and fatty acid oxidation. *Nat. Commun.* **10**, 2698 (2019).

393. Nguyen, P. L. *et al.* Fatty acid synthase polymorphisms, tumor expression, body mass index, prostate cancer risk, and survival. *J. Clin. Oncol.* **28**, 3958–3964 (2010).

394. Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* **7**, 763–777 (2007).

395. Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **169**, 361–371 (2017).

396. Laplante, M. & Sabatini, D. M. mTOR Signaling in Growth Control and Disease. *Cell* **149,** 274–293 (2012).

397. Wu, L. *et al.* Rapamycin upregulates autophagy by inhibiting the mTOR-ULK1 pathway, resulting in reduced podocyte injury. *PLoS One* **8**, e63799 (2013).
398. Shimano, H. & Sato, R. SREBP-regulated lipid metabolism: convergent physiology - divergent pathophysiology. *Nat. Rev. Endocrinol.* **13**, 710–730 (2017).
399. Hudson, C. C. *et al.* Regulation of Hypoxia-Inducible Factor 1α Expression and Function by the Mammalian Target of Rapamycin. *Molecular and Cellular Biology* **22**, 7004–7014 (2002).
400. Düvel, K. *et al.* Activation of a metabolic gene regulatory network downstream of mTOR complex 1. *Mol. Cell* **39**, 171–183 (2010).
401. Cunningham, J. T. *et al.* mTOR controls mitochondrial oxidative function through a YY1-PGC-1alpha transcriptional complex. *Nature* **450**, 736–740 (2007).
402. Linke, M., Fritsch, S. D., Sukhbaatar, N., Hengstschläger, M. & Weichhart, T. mTORC1 and mTORC2 as regulators of cell metabolism in immunity. *FEBS Lett.* **591**, 3089–3103 (2017).
403. Catalán, V. *et al.* Expression of S6K1 in human visceral adipose tissue is upregulated in obesity and related to insulin resistance and inflammation. *Acta Diabetologica* **52**, 257–266 (2015).
404. Hardie, D. G. Keeping the home fires burning: AMP-activated protein kinase. *J. R. Soc. Interface* **15**, (2018).
405. Bijland, S., Mancini, S. J. & Salt, I. P. Role of AMP-activated protein kinase in adipose tissue metabolism and inflammation. *Clin. Sci.* **124**, (2013).
406. Erik A. Richter, N. B. R. AMPK and the biochemistry of exercise: Implications for human health and disease. *Biochem. J.* **418**, 261 (2009).
407. Gibala, M. J. *et al.* Brief intense interval exercise activates AMPK and p38 MAPK signaling and increases the expression of PGC-1α in human skeletal muscle. *Journal of Applied Physiology* **106**, 929–934 (2009).
408. The role of physical activity and exercise in obesity and weight management: Time for

critical appraisal. *Journal of Sport and Health Science* **5**, 151–154 (2016).

409. Soussi, T., Dehouche, K. & Béroud, C. p53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis. *Hum. Mutat.* **15**, 105–113 (2000).

410. Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).

411. Kharitonenkov, A. *et al.* FGF-21 as a novel metabolic regulator. *Journal of Clinical Investigation* **115**, 1627–1635 (2005).

412. Esau, C. *et al.* miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab.* **3**, 87–98 (2006).

413. Gil-Zamorano, J. *et al.* Docosahexaenoic acid modulates the enterocyte Caco-2 cell expression of microRNAs involved in lipid metabolism. *J. Nutr.* **144**, 575–585 (2014).

414. Murata, Y. *et al.* Up-Regulated MicroRNA-27b Promotes Adipocyte Differentiation via Induction of Acyl-CoA Thioesterase 2 Expression. *Biomed Res. Int.* **2019**, 2916243 (2019).

415. Wu, Q. *et al.* Exosomes from the tumour-adipocyte interplay stimulate beige/brown differentiation and reprogram metabolism in stromal adipocytes to promote tumour progression. *J. Exp. Clin. Cancer Res.* **38**, 223 (2019).

416. Wu, Q. *et al.* Tumour-originated exosomal miR-155 triggers cancer-associated cachexia to promote tumour progression. *Mol. Cancer* **17**, 155 (2018).

417. Chafey, P. *et al.* Proteomic analysis of beta-catenin activation in mouse liver by DIGE analysis identifies glucose metabolism as a new target of the Wnt pathway. *Proteomics* **9**, 3889–3900 (2009).

418. Ham, S. A. *et al.* Ligand-activated interaction of PPARδ with c-Myc governs the tumorigenicity of breast cancer. *International journal of cancer* **143**, (2018).

419. Maguire, O. A. *et al.* Creatine-mediated crosstalk between adipocytes and cancer cells regulates obesity-driven breast cancer. *Cell Metab.* **33**, 499–512.e6 (2021).

420. Brändstedt, J. *et al.* Gender, anthropometric factors and risk of colorectal cancer with particular reference to tumour location and TNM stage: a cohort study. *Biol. Sex Differ.* **3**, 23 (2012).

421. Haakinson, D. J. *et al.* The impact of obesity on breast cancer: a retrospective review. *Ann. Surg. Oncol.* **19**, 3012–3018 (2012).

422. Porter, G. A., Inglis, K. M., Wood, L. A. & Veugelers, P. J. Effect of obesity on presentation of breast cancer. *Ann. Surg. Oncol.* **13**, 327–332 (2006).

423. Njor, S.H. *et al.* Body weight and sensitivity of screening mammography. *Eur. J. Cancer* **60**, 93–100 (2016).

424. Maruthur, N. M., Bolen, S., Brancati, F. L. & Clark, J. M. Obesity and Mammography: A Systematic Review and Meta-Analysis. *J. Gen. Intern. Med.* **24**, 665 (2009).

425. Friedman, A. M., Hemler, J. R., Rossetti, E., Clemow, L. P. & Ferrante, J. M. Obese women's barriers to mammography and pap smear: the possible role of personality. *Obesity* **20**, 1611–1617 (2012).

426. Aldenkortt, M., Lysakowski, C., Elia, N., Brochard, L. & Tramèr, M. R. Ventilation strategies in obese patients undergoing surgery: a quantitative systematic review and meta-analysis†. *Br. J. Anaesth.* **109**, 493–502 (2012).

427. Postlethwait, R. W. & Johnson, W. D. Complications following surgery for duodenal ulcer in obese patients. *Arch. Surg.* **105**, 438–440 (1972).

428. Abdel-Moneim, R. I. The hazards of surgery in the obese. *Int. Surg.* **70**, (1985).

429. Dandapani, S. V., Zhang, Y., Jennelle, R. & Lin, Y. G. Radiation-Associated Toxicities in Obese Women with Endometrial Cancer: More Than Just BMI? *Scientific World Journal* **2015**, 483208 (2015).

430. Twardella, D. *et al.* Personal characteristics, therapy modalities and individual DNA repair capacity as predictive factors of acute skin toxicity in an unselected cohort of breast cancer patients receiving radiotherapy. *Radiotherapy and Oncology* **69**, 145–153 (2003).

431. Thigpen, J. T. Appropriate Chemotherapy Dosing for Obese Adult Patients With Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *Yearbook of Oncology* **2012,** 9–10 (2012).

432. Ligibel, J. A. & Winer, E. P. Aromatase inhibition in obese women: how much is enough? *Journal of Clinical Oncology: official journal of the American Society of Clinical Oncology* **30**, 2940–2942 (2012).

433. Pfeiler, G. *et al.* Impact of body mass index on the efficacy of endocrine therapy in premenopausal patients with breast cancer: an analysis of the prospective ABCSG-12 trial. *J. Clin. Oncol.* **29**, 2653–2659 (2011).

434. Yom, C. K. *et al.* Leptin as a Potential Target for Estrogen Receptor-Positive Breast Cancer. *J. Breast Cancer* **16**, 138 (2013).

435. Creighton, C. J. *et al.* A gene transcription signature of obesity in breast cancer. *Breast Cancer Res. Treat.* **132**, (2012).

436. Mamidi, T. K. K., Wu, J., Tchounwou, P. B., Miele, L. & Hicks, C. Whole Genome Transcriptome Analysis of the Association between Obesity and Triple-Negative Breast Cancer in Caucasian Women. *Int. J. Environ. Res. Public Health* **15**, (2018).

437. Sanchez, A. *et al.* Transcriptomic signatures related to the obesity paradox in patients with clear cell renal cell carcinoma: a cohort study. *Lancet Oncol.* **21**, 283–293 (2020).

438. Sharad, S. *et al.* Prostate cancer gene expression signature of patients with high body mass index. *Prostate Cancer Prostatic Dis.* **14**, (2011).

439. Hines, R. B. et al. Effect of comorbidity and body mass index on the survival of African-American and Caucasian patients with colon cancer. Cancer **115**, 5798–5806 (2009).

440. Navarro, W. H. *et al.* Effect of body mass index on mortality of patients with lymphoma undergoing autologous hematopoietic cell transplantation. *Biol. Blood Marrow Transplant.* **12**, 541–551 (2006).

441. Parker, A. S. *et al.* Greater body mass index is associated with better pathologic features and improved outcome among patients treated surgically for clear cell renal cell

carcinoma. *Urology* **68**, 741–746 (2006).

442. Schlesinger, S. *et al.* Postdiagnosis body mass index and risk of mortality in colorectal cancer survivors: a prospective study and meta-analysis. *Cancer Causes & Control* **25**, 1407–1418 (2014).

443. Amptoulach, S., Gross, G. & Kalaitzakis, E. Differential impact of obesity and diabetes mellitus on survival after liver resection for colorectal cancer metastases. *J. Surg. Res.* **199**, 378–385 (2015).

444. Brunner, A. M. *et al.* Association between baseline body mass index and overall survival among patients over age 60 with acute myeloid leukemia. *Am. J. Hematol.* **88**, 642–646 (2013).

445. Tsang, N. M. *et al.* Overweight and obesity predict better overall survival rates in cancer patients with distant metastases. *Cancer Med.* **5**, 665–675 (2016).

446. McQuade, J. L. *et al.* The association of BMI and outcomes in metastatic melanoma: A retrospective, multicohort analysis of patients treated with targeted therapy, immunotherapy, or chemotherapy. *Lancet Oncol.* **19**, 310 (2018).

447. Welch, H. G., Schwartz, L. M. & Woloshin, S. Are increasing 5-year survival rates evidence of success against cancer? *JAMA* **283**, 2975–2978 (2000).

448. Lennon, H., Sperrin, M., Badrick, E. & Renehan, A. G. The Obesity Paradox in Cancer: a Review. *Current Oncology Reports* **18** (2016).

449. Xia, S. *et al.* Prevention of dietary fat-fueled ketogenesis attenuates BRAF V600E tumor growth. *Cell Metab.* **25**, 358 (2017).

450. Sergentanis, T. N. *et al.* Obesity and risk of malignant melanoma: a meta-analysis of cohort and case-control studies. *Eur. J. Cancer* **49**, 642–657 (2013).

451. Kichenadasse, G. *et al.* Association Between Body Mass Index and Overall Survival With Immune Checkpoint Inhibitor Therapy for Advanced Non–Small Cell Lung Cancer. *JAMA Oncol* **6**, 512–518 (2020).

452. De Giorgi, U. *et al.* Association of Systemic Inflammation Index and Body Mass Index

with Survival in Patients with Renal Cell Cancer Treated with Nivolumab. *Clin. Cancer Res.* **25**, 3839–3846 (2019).

453. The prognostic role of obesity is independent of sex in cancer patients treated with immune checkpoint inhibitors: A pooled analysis of 4090 cancer patients. *Int. Immunopharmacol.* **74**, 105745 (2019).
454. Naik, G. S. *et al.* Complex inter-relationship of body mass index, gender and serum creatinine on survival: exploring the obesity paradox in melanoma patients treated with checkpoint inhibition. *Journal for ImmunoTherapy of Cancer* **7**, 1–17 (2019).
455. Yang, R. & Barouch, L. A. Leptin Signaling and Obesity. *Circulation Research* **101**, 545–559 (2007).
456. Matarese, G. *et al.* Leptin increase in multiple sclerosis associates with reduced number of CD4(+)CD25+ regulatory T cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5150–5155 (2005).
457. Rosa, V. D. & De Rosa, V. Leptin neutralization interferes with pathogenic T cell autoreactivity in autoimmune encephalomyelitis. *Journal of Clinical Investigation* **116**, 447–455 (2006).
458. Rajapurohitam, V., Javadov, S., Purdham, D. M., Kirshenbaum, L. A. & Karmazyn, M. An autocrine role for leptin in mediating the cardiomyocyte hypertrophic effects of angiotensin II and endothelin-1. *J. Mol. Cell. Cardiol.* **41**, 265–274 (2006).
459. Singh, U. P. *et al.* The emerging role of leptin antagonist as potential therapeutic option for inflammatory bowel disease. *Int. Rev. Immunol.* **33**, 23–33 (2014).
460. Gertler, A. & Solomon, G. Leptin-activity blockers: development and potential use in experimental biology and medicine. *Can. J. Physiol. Pharmacol.* **91**, 873–882 (2013).
461. Riera-Guardia, N. & Rothenbacher, D. The effect of thiazolidinediones on adiponectin serum level: a meta-analysis. *Diabetes Obes. Metab.* **10**, (2008).
462. Gómez-Arbeláez, D. *et al.* Aged Garlic Extract Improves Adiponectin Levels in Subjects with Metabolic Syndrome: A Double-Blind, Placebo-Controlled, Randomized, Crossover

Study. *Mediators Inflamm.* **2013**, (2013).

463. Achari, A. E. & Jain, S. K. L-Cysteine supplementation increases adiponectin synthesis and secretion, and GLUT4 and glucose utilization by upregulating disulfide bond A-like protein expression mediated by MCP-1 inhibition in 3T3-L1 adipocytes exposed to high glucose. *Mol. Cell. Biochem.* **414**, 105–113 (2016).

464. Burlet, E. & Jain, S. K. Manganese supplementation increases adiponectin and lowers ICAM-1 and creatinine blood levels in Zucker type 2 diabetic rats, and downregulates ICAM-1 by upregulating adiponectin multimerization protein (DsbA-L) in endothelial cells. *Mol. Cell. Biochem.* **429**, 1–10 (2017).

465. Jürimäe, J., Purge, P. & Jürimäe, T. Adiponectin is altered after maximal exercise in highly trained male rowers. *Eur. J. Appl. Physiol.* **93**, 502–505 (2005).

466. Kriketos, A. D. *et al.* Exercise Increases Adiponectin Levels and Insulin Sensitivity in Humans. *Diabetes Care* **27**, 629–630 (2004).

467. Frodermann, V. *et al.* Exercise reduces inflammatory cell production and cardiovascular inflammation via instruction of hematopoietic progenitor cells. *Nat. Med.* **25**, (2019).

468. Flossmann, E., Rothwell, P. M. & British Doctors Aspirin Trial and the UK-TIA Aspirin Trial. Effect of aspirin on long-term risk of colorectal cancer: consistent evidence from randomised and observational studies. *Lancet* **369**, 1603–1613 (2007).

469. Movahedi, M. *et al.* Obesity, Aspirin, and Risk of Colorectal Cancer in Carriers of Hereditary Colorectal Cancer: A Prospective Investigation in the CAPP2 Study. *J. Clin. Oncol.* **33**, (2015).

470. Incio, J. *et al.* Obesity promotes resistance to anti-VEGF therapy in breast cancer by up-regulating IL-6 and potentially FGF-2. *Sci. Transl. Med.* **10**, (2018).

471. Luo, J. *et al.* Intentional Weight Loss and Obesity-Related Cancer Risk. *JNCI Cancer Spectr* **3**, (2019).

472. Rapp, K. *et al.* Weight change and cancer risk in a cohort of more than 65 000 adults in Austria. *Annals of Oncology* **19**, 641–648 (2008).

473. Wiggins, T., Antonowicz, S. S. & Markar, S. R. Cancer Risk Following Bariatric Surgery—Systematic Review and Meta-analysis of National Population-Based Cohort Studies. *Obes. Surg.* **29**, 1031–1039 (2018).
474. Feigelson, H. S. *et al.* Bariatric Surgery is Associated With Reduced Risk of Breast Cancer in Both Premenopausal and Postmenopausal Women. *Annals of Surgery* **272**, 1053–1059 (2020).
475. Emerging Risk Factors Collaboration *et al.* Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. *Lancet* **377**, 1085–1095 (2011).
476. Singh, G. M. *et al.* The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: a pooled analysis. *PLoS One* **8**, e65174 (2013).
477. Quetelet, L. A. J. Sur l'Homme et le Développement de ses Facultés, ou Essai de Physique Sociale. (1835).
478. WHO | Obesity: preventing and managing the global epidemic. (2015).
479. Nuttall, F. Q. Body Mass Index: Obesity, BMI, and Health: A Critical Review. *Nutr. Today* **50**, (2015).
480. Adab, P., Pallan, M. & Whincup, P. H. Is BMI the best measure of obesity? *BMJ* **360** (2018).
481. Rothman, K. J. BMI-related errors in the measurement of obesity. *Int. J. Obes.* **32 Suppl 3**, S56–9 (2008).
482. Heymsfield, S. B. & Cefalu, W. T. Does Body Mass Index Adequately Convey a Patient's Mortality Risk? *JAMA* **309** 87 (2013).
483. Gonzalez, M. C., Correia, M. I. T. D. & Heymsfield, S. B. A requiem for BMI in the clinical setting. *Curr. Opin. Clin. Nutr. Metab. Care* **20**, 314–321 (2017).
484. Deurenberg, P., Weststrate, J. A. & Seidell, J. C. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. *Br. J. Nutr.* **65**, 105–114 (1991).
485. Roche, A. F., Sievogel, R. M., Chumlea, W. C. & Webb, P. Grading body fatness from

limited anthropometric data. *Am. J. Clin. Nutr.* **34**, 2831–2838 (1981).

486. Fenger-Grøn, M., Overvad, K., Tjønneland, A. & Frost, L. Lean Body Mass Is the Predominant Anthropometric Risk Factor for Atrial Fibrillation. *J. Am. Coll. Cardiol.* **69**, (2017).

487. Magnani, J. W., Hylek, E. M. & Apovian, C. M. Obesity begets atrial fibrillation: a contemporary summary. *Circulation* **128**, 401–405 (2013).

488. Murphy, N. F. *et al.* Long-term cardiovascular consequences of obesity: 20-year follow-up of more than 15 000 middle-aged men and women (the Renfrew-Paisley study). *Eur. Heart J.* **27**, 96–106 (2006).

489. Tedrow, U. B. *et al.* The Long- and Short-Term Impact of Elevated Body Mass Index on the Risk of New Atrial Fibrillation. *Journal of the American College of Cardiology* **55** 2319–2327 (2010).

490. Wang, T. J., Parise, H. & Levy, D. Obesity and the risk of new-onset atrial fibrillation. *ACC Current Journal Review* **14** 45 (2005).

491. Worm, M. S. *et al.* Atrial fibrillation is associated with lean body mass in postmenopausal women. *Sci. Rep.* **10**, 573 (2020).

492. Staunstrup, L. M. *et al.* Cancer risk in relation to body fat distribution, evaluated by DXA-scans, in postmenopausal women – the Prospective Epidemiological Risk Factor (PERF) study. *Sci. Rep.* **9**, 1–9 (2019).

493. Lahmann, P. H., Lissner, L., Gullberg, B., Olsson, H. & Berglund, G. A prospective study of adiposity and postmenopausal breast cancer risk: The Malmö diet and cancer study. *International Journal of Cancer* **103**, 246–252 (2003).

494. Krebs, E. E. *et al.* Measures of adiposity and risk of breast cancer in older postmenopausal women. *J. Am. Geriatr. Soc.* **54**, 63–69 (2006).

495. MacInnis, R. J., English, D. R., Gertig, D. M., Hopper, J. L. & Giles, G. G. Body Size and Composition and Risk of Postmenopausal Breast Cancer. *Cancer Epidemiol. Biomarkers Prev.* **13**, 2117–2125 (2004).

496. Hoffmann, C. & Weigert, C. Skeletal Muscle as an Endocrine Organ: The Role of Myokines in Exercise Adaptations. *Cold Spring Harb. Perspect. Med.* **7**, (2017).
497. Pratesi, A. Skeletal muscle: an endocrine organ. *Clinical Cases In Mineral And Bone Metabolism* (2013) doi:10.11138/ccmbm/2013.10.1.011.
498. Hocking, S., Samocha-Bonet, D., Milner, K. L., Greenfield, J. R. & Chisholm, D. J. Adiposity and insulin resistance in humans: the role of the different tissue and cellular lipid depots. *Endocr. Rev.* **34**, (2013).
499. Srikanthan, P. & Karlamangla, A. S. Relative Muscle Mass Is Inversely Associated with Insulin Resistance and Prediabetes. Findings from The Third National Health and Nutrition Examination Survey. *J. Clin. Endocrinol. Metab.* **96**, 2898–2903 (2011).
500. Caan, B. J., Cespedes Feliciano, E. M. & Kroenke, C. H. The Importance of Body Composition in Explaining the Overweight Paradox in Cancer-Counterpoint. *Cancer Res.* **78**, 1906–1912 (2018).
501. Chen, M.-J. *et al.* Body mass index and breast cancer: analysis of a nation-wide population-based prospective cohort study on 1 393 985 Taiwanese women. *Int. J. Obes.* **40**, 524–530 (2016).
502. Munsell, M. F., Sprague, B. L., Berry, D. A., Chisholm, G. & Trentham-Dietz, A. Body Mass Index and Breast Cancer Risk According to Postmenopausal Estrogen-Progestin Use and Hormone Receptor Status. *Epidemiol. Rev.* **36**, 114 (2014).
503. Picon-Ruiz, M., Morata-Tarifa, C., Valle-Goffin, J. J., Friedman, E. R. & Slingerland, J. M. Obesity and adverse breast cancer risk and outcome: Mechanistic insights and strategies for intervention. *CA Cancer J. Clin.* **67**, 378–397 (2017).
504. Pierobon, M. & Frankenfeld, C. L. Obesity as a risk factor for triple-negative breast cancers: a systematic review and meta-analysis. *Breast Cancer Res. Treat.* **137**, 307–314 (2013).
505. White, A. J., Nichols, H. B., Bradshaw, P. T. & Sandler, D. P. Overall and central adiposity and breast cancer risk in the Sister Study. *Cancer* **121**, 3700–3708 (2015).

506. Nagrani, R. *et al.* Central obesity increases risk of breast cancer irrespective of menopausal and hormonal receptor status in women of South Asian Ethnicity. *Eur. J. Cancer* **66**, 153–161 (2016).

507. Park, B. *et al.* Attribution to Heterogeneous Risk Factors for Breast Cancer Subtypes Based on Hormone Receptor and Human Epidermal Growth Factor 2 Receptor Expression in Korea. *Medicine* **95**, e3063 (2016).

508. Ma, H., Bernstein, L., Ross, R. K. & Ursin, G. Hormone-related risk factors for breast cancer in women under age 50 years by estrogen and progesterone receptor status: results from a case–control and a case–case comparison. *Breast Cancer Research* **8**, R39 (2006).

509. Enger, S. M., Ross, R. K., Paganini-Hill, A., Carpenter, C. L. & Bernstein, L. Body size, physical activity, and breast cancer hormone receptor status: results from two case-control studies. *Cancer Epidemiol. Biomarkers Prev.* **9**, 681–687 (2000).

510. Michels, K. B., Terry, K. L. & Willett, W. C. Longitudinal study on the role of body size in premenopausal breast cancer. *Arch. Intern. Med.* **166**, 2395–2402 (2006).

511. Berstad, P. *et al.* A Case-Control Study of Body Mass Index and Breast Cancer Risk in White and African-American Women. *Cancer Epidemiology Biomarkers & Prevention* **19**, 1532–1544 (2010).

512. Harris, H. R., Willett, W. C., Terry, K. L. & Michels, K. B. Body fat distribution and risk of premenopausal breast cancer in the Nurses' Health Study II. *J. Natl. Cancer Inst.* **103**, 273–278 (2011).

513. Kunutsor, S. K. & Laukkanen, J. A. Gamma-glutamyltransferase and risk of prostate cancer: Findings from the KIHD prospective cohort study. *International Journal of Cancer* **140**, 818–824 (2017).

514. Choi, J. B. *et al.* The Impact of Diabetes on the Risk of Prostate Cancer Development according to Body Mass Index: A 10-year Nationwide Cohort Study. *J. Cancer* **7**, 2061–2066 (2016).

515. Rao, G. A. *et al.* Angiotensin receptor blockers and risk of prostate cancer among United States veterans. *J. Clin. Pharmacol.* **53**, 773–778 (2013).
516. Perez-Cornago, 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.* **15**, 115 (2017).
517. Möller, E. *et al.* Lifetime body size and prostate cancer risk in a population-based case–control study in Sweden. *Cancer Causes & Control* **24**, 2143–2155 (2013).
518. Häggström, C. *et al.* Prospective study on metabolic factors and risk of prostate cancer. *Cancer* **118**, 6199–6206 (2012).
519. Shafique, K. *et al.* Cholesterol and the risk of grade-specific prostate cancer incidence: evidence from two large prospective cohort studies with up to 37 years' follow up. *BMC Cancer* **12**, 25 (2012).
520. Harrison, S. *et al.* Systematic review and meta-analysis of the associations between body mass index, prostate cancer, advanced prostate cancer, and prostate-specific antigen. *Cancer Causes & Control* **31**, 431–449 (2020).
521. Harrison, S. *et al.* Investigating the prostate specific antigen, body mass index and age relationship: is an age–BMI-adjusted PSA model clinically useful? *Cancer Causes Control* **27**, 1465 (2016).
522. Dai, Z. Obesity and colorectal cancer risk: A meta-analysis of cohort studies. *World Journal of Gastroenterology* **13**, 4199 (2007).
523. Moghaddam, A. A., Woodward, M. & Huxley, R. Obesity and risk of colorectal cancer: a meta-analysis of 31 studies with 70,000 events. *Cancer Epidemiol. Biomarkers Prev.* **16**, (2007).
524. Guh, D. P. *et al.* The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health* **9**, 88 (2009).
525. Rapp, K. *et al.* Obesity and incidence of cancer: a large cohort study of over 145,000 adults in Austria. *Br. J. Cancer* **93**, (2005).

526. Engeland, A., Tretli, S., Austad, G. & Bjørge, T. Height and body mass index in relation to colorectal and gallbladder cancer in two million Norwegian men and women. *Cancer Causes Control* **16**, 987–996 (2005).

527. Dong, Y. *et al.* Abdominal obesity and colorectal cancer risk: systematic review and meta-analysis of prospective studies. *Biosci. Rep.* **37**, (2017).

528. Guo, W., Key, T. J. & Reeves, G. K. Adiposity and breast cancer risk in postmenopausal women: Results from the UK Biobank prospective cohort. *Int. J. Cancer* **143**, 1037–1046 (2018).

529. Perez-Cornago, A. *et al.* Prospective investigation of risk factors for prostate cancer in the UK Biobank cohort study. *Br. J. Cancer* **117**, 1562–1571 (2017).

530. Pietrobelli, A., Rubiano, F., St-Onge, M. P. & Heymsfield, S. B. New bioimpedance analysis system: improved phenotyping with whole-body analysis. *Eur. J. Clin. Nutr.* **58**, (2004).

531. Kelly, J. S. & Metcalfe, J. Validity and reliability of body composition analysis using the Tanita BC418-MA. *J. Exerc. Physiol. Online* **15**, 74–84 (2012).

532. UK Biobank. Anthropometry, 2014. Online: https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/Anthropometry.pdf

533. Brown, J. C., Cespedes Feliciano, E. M. & Caan, B. J. The evolution of body composition in oncology—epidemiology, clinical trials, and the future of patient care: facts and numbers. *J. Cachexia Sarcopenia Muscle* **9**, 1200 (2018).

534. Lei, Y.-Y. *et al.* Dietary changes in the first 3 years after breast cancer diagnosis: a prospective Chinese breast cancer cohort study. *Cancer Manag. Res.* **10**, 4073 (2018).

535. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202 (2016).

536. Yano, K. *et al.* GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21262–21267 (2019).

537. Marie-Pierre St-Onge, D. G. Body composition changes with aging: The cause or the result of alterations in metabolic rate and macronutrient oxidation? *Nutrition* **26**, 152 (2010).

538. Kim, H. & Giovannucci, E. L. Sex differences in the association of obesity and colorectal cancer risk. *Cancer Causes Control* **28**, 1–4 (2017).

539. Poehlman, E. T., Toth, M. J. & Gardner, A. W. Changes in energy balance and body composition at menopause: a controlled longitudinal study. *Ann. Intern. Med.* **123**, 673–675 (1995).

540. Poehlman, E. T. & Tchernof, A. Traversing the menopause: changes in energy expenditure and body composition. *Coron. Artery Dis.* **9**, 799–803 (1998).

541. Mellemkjaer, L. *et al.* Body composition and breast cancer in postmenopausal women: a Danish prospective cohort study. *Obesity* **14**, 1854–1862 (2006).

542. Heymsfield, S. B., Heo, M., Thomas, D. & Pietrobelli, A. Scaling of body composition to height: relevance to height-normalized indexes. *Am. J. Clin. Nutr.* **93**, 736–740 (2011).

543. James, F. R. *et al.* Obesity in breast cancer – What is the risk factor? *European Journal of Cancer* **51**, 705–720 (2015).

544. Nelson, L. R. & Bulun, S. E. Estrogen production and action. *J. Am. Acad. Dermatol.* **45**, (2001).

545. Larionov, A. A. *et al.* Aromatase in skeletal muscle. *J. Steroid Biochem. Mol. Biol.* **84**, (2003).

546. Grinspoon, S. *et al.* Loss of lean body and muscle mass correlates with androgen levels in hypogonadal men with acquired immunodeficiency syndrome and wasting. *J. Clin. Endocrinol. Metab.* **81**, 4051–4058 (1996).

547. Juul, A. *et al.* Serum insulin-like growth factor-I in 1030 healthy children, adolescents, and adults: relation to age, sex, stage of puberty, testicular size, and body mass index. *J. Clin. Endocrinol. Metab.* **78**, (1994).

548. Renehan, A. G. *et al.* Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer

risk: systematic review and meta-regression analysis. *Lancet* **363**, 1346–1353 (2004).

549. Ma, J. *et al.* A prospective study of plasma levels of insulin-like growth factor I (IGF-I) and IGF-binding protein-3, and colorectal cancer risk among men. *Growth Hormone & IGF Research* **10**, S28–S29 (2000).

550. Conejo, R., Valverde, A. M., Benito, M. & Lorenzo, M. Insulin produces myogenesis in C2C12 myoblasts by induction of NF-kappaB and downregulation of AP-1 activities. *J. Cell. Physiol.* **186**, (2001).

551. Gallagher, E. J. & LeRoith, D. Hyperinsulinaemia in cancer. *Nature Reviews Cancer* **20**, 629–644 (2020).

552. Boyd, N. F. *et al.* Mammographic density and the risk and detection of breast cancer. *N. Engl. J. Med.* **356**, (2007).

553. Duffy, S. W. *et al.* Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. *Eur. J. Cancer* **88**, 48–56 (2018).

554. Juge-Aubry, C. E. *et al.* Adipose tissue is a major source of interleukin-1 receptor antagonist: upregulation in obesity and inflammation. *Diabetes* **52**, 1104–1110 (2003).

555. Bonn, S. E. *et al.* Body mass index in relation to serum prostate-specific antigen levels and prostate cancer risk. *Int. J. Cancer* **139**, 50–57 (2016).

556. Littlejohns, T. J., Travis, R. C., Key, T. J. & Allen, N. E. Lifestyle factors and prostate-specific antigen (PSA) testing in UK Biobank: Implications for epidemiological research. *Cancer Epidemiol.* **45**, (2016).

557. Pollard, M., Luckert, P. H. & Schmidt, M. A. Induction of prostate adenocarcinomas in Lobund Wistar rats by testosterone. *The Prostate* **3**, 563–568 (1982).

558. Glass, A. R., Swerdloff, R. S., Bray, G. A., Dahms, W. T. & Atkinson, R. L. Low serum testosterone and sex-hormone-binding-globulin in massively obese men. *J. Clin. Endocrinol. Metab.* **45**, 1211–1219 (1977).

559. Michaud, J. E., Billups, K. L. & Partin, A. W. Testosterone and prostate cancer: an

evidence-based review of pathogenesis and oncologic risk. *Ther. Adv. Urol.* **7**, 378 (2015).

560. Dupree, J. M., Langille, G. M., Khera, M. & Lipshultz, L. I. The safety of testosterone supplementation therapy in prostate cancer. *Nat. Rev. Urol.* **11**, 526–530 (2014).
561. Pastuszak, A. W. *et al.* Testosterone Therapy after Radiation Therapy for Low, Intermediate and High Risk Prostate Cancer. *J. Urol.* **194**, (2015).
562. Caton, J. R., Molé, P. A., Adams, W. C. & Heustis, D. S. Body composition analysis by bioelectrical impedance: effect of skin temperature. *Med. Sci. Sports Exerc.* **20**, 489–491 (1988).
563. Liang, M. T., Su, H. F. & Lee, N. Y. Skin temperature and skin blood flow affect bioelectric impedance study of female fat-free mass. *Med. Sci. Sports Exerc.* **32**, 221–227 (2000).
564. Mahshid Dehghan, A. T. M. Is bioelectrical impedance accurate for use in large epidemiological studies? *Nutr. J.* **7**, 26 (2008).
565. Völgyi, E. *et al.* Assessing body composition with DXA and bioimpedance: effects of obesity, physical activity, and age. *Obesity* **16**, 700–705 (2008).
566. Neovius, M., Hemmingsson, E., Freyschuss, B. & Uddén, J. Bioelectrical impedance underestimates total and truncal fatness in abdominally obese women. *Obesity* **14**, (2006).
567. Kutáč, P. The effect of intake of water on the final values of body composition parameters in active athletes using two different bioimpedance analyzers. *Acta Gymnica* **44** 107–116 (2014).
568. Randhawa, A. K., Jamnik, V., Fung, M. D. T., Fogel, A. S. & Kuk, J. L. No differences in the body fat after violating core bioelectrical impedance measurement assumptions. *BMC Public Health* **21**, 495 (2021).
569. Abe, T., Bemben, M. G., Kondo, M., Kawakami, Y. & Fukunaga, T. Comparison of skeletal muscle mass to fat-free mass ratios among different ethnic groups. *J. Nutr.*

Health Aging **16**, 534–538 (2012).

570. Evans, J. M. M., Donnelly, L. A., Emslie-Smith, A. M., Alessi, D. R. & Morris, A. D. Metformin and reduced risk of cancer in diabetic patients. *BMJ* **330** 1304–1305 (2005).

571. Lindsay, C. *et al.* Antitumour effects of metformin and curcumin in human papillomavirus positive and negative head and neck cancer cells. *Mol. Carcinog.* **58**, 1946-1959 (2019).

572. Zakikhani, M., Dowling, R., Fantus, I. G., Sonenberg, N. & Pollak, M. Metformin is an AMP kinase-dependent growth inhibitor for breast cancer cells. *Cancer Res.* **66**, 10269-73 (2006).

573. Deng, X. S. *et al.* Metformin targets Stat3 to inhibit cell growth and induce apoptosis in triple-negative breast cancers. *Cell Cycle* **11**, 367-376 (2012).

574. Lee, M.-S. *et al.* Type 2 diabetes increases and metformin reduces total, colorectal, liver and pancreatic cancer incidences in Taiwanese: a representative population prospective cohort study of 800,000 individuals. *BMC Cancer* **11**, 20 (2011).

575. Sehdev, A. *et al.* Metformin for primary colorectal cancer prevention in patients with diabetes: A case-control study in a US population. *Cancer*. **121**, 1071–1078 (2015).

576. Bodmer, M., Becker, C., Meier, C., Jick, S. S. & Meier, C. R. Use of Metformin Is Not Associated with a Decreased Risk of Colorectal Cancer: A Case–Control Analysis. *Cancer Epidemiol. Biomarkers Prev.* **21**, 280–286 (2012).

577. Kowall, B., Stang, A., Rathmann, W. & Kostev, K. No reduced risk of overall, colorectal, lung, breast, and prostate cancer with metformin therapy in diabetic patients: database analyses from Germany and the UK. *Pharmacoepidemiol. Drug Saf.* **24**, (2015).

578. Knapen, L. M. *et al.* Use of biguanides and the risk of colorectal cancer: a register-based cohort study. *Curr. Drug Saf.* **8**, 349–356 (2013).

579. Suissa, S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol. Drug Saf.* **16**, (2007).

580. Suissa, S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. *Am. J. Respir. Crit. Care Med.* **168**,

49-53 (2003).

581. Suissa, S., Dell'aniello, S., Vahey, S. & Renoux, C. Time-window bias in case-control studies: statins and lung cancer. *Epidemiology* **22**, 228-231 (2011).

582. Birsoy, K. *et al.* Metabolic determinants of cancer cell sensitivity to glucose limitation and biguanides. *Nature* **508**, 108–112 (2014).

583. Goldmann, W. H. *et al.* Saw palmetto berry extract inhibits cell growth and Cox-2 expression in prostatic cancer cells. *Cell Biol. Int.* **25**, 1117–1124 (2001).

584. Brasky, T. M. *et al.* Specialty supplements and prostate cancer risk in the VITamins and Lifestyle (VITAL) cohort. *Nutr. Cancer* **63**, 573–582 (2011).

585. Bonnar-Pizzorno, R. M., Littman, A. J., Kestin, M. & White, E. Saw palmetto supplement use and prostate cancer risk. *Nutr. Cancer* **55**, (2006).

586. Klein, E. A. *et al.* Vitamin E and the Risk of Prostate Cancer: Updated Results of The Selenium and Vitamin E Cancer Prevention Trial (SELECT). *JAMA* **306**, 1549 (2011).

587. Vivarelli, F. *et al.* Co-carcinogenic effects of vitamin E in prostate. *Sci. Rep.* **9**, 1–10 (2019).

588. Ruan, D. & So, S.-P. Prostaglandin E2 produced by inducible COX-2 and mPGES-1 promoting cancer cell proliferation in vitro and in vivo. *Life Sci.* **116**, 43–50 (2014).

589. Qin, G. *et al.* Palbociclib inhibits epithelial-mesenchymal transition and metastasis in breast cancer via c-Jun/COX-2 signaling pathway. *Oncotarget* **6**, 41794–41808 (2015).

590. Santos, R. de O. *et al.* Principal Component Analysis and Factor Analysis: differences and similarities in Nutritional Epidemiology application. *Rev. Bras. Epidemiol.* **22**, e190041 (2019).

591. Albanes, D., Jones, D. Y., Schatzkin, A., Micozzi, M. S. & Taylor, P. R. Adult stature and risk of cancer. *Cancer Res.* **48**, 1658-62 (1988).

592. Renehan, A. G. Height and cancer: consistent links, but mechanisms unclear. *The Lancet Oncology* **12**, 716–717 (2011).

593. Bhaskaran, K. *et al.* Body-mass index and risk of 22 specific cancers: a population-

based cohort study of 5·24 million UK adults. *The Lancet* **384**, 755–765 (2014).

594. Park, J., Morley, T. S., Kim, M., Clegg, D. J. & Scherer, P. E. Obesity and cancer--mechanisms underlying tumour progression and recurrence. *Nat. Rev. Endocrinol.* **10**, 455–465 (2014).

595. Benyi, E. *et al.* Adult height is associated with risk of cancer and mortality in 5.5 million Swedish women and men. *J. Epidemiol. Community Health* **73**, 730–736 (2019).

596. Clayton, P. E., Banerjee, I., Murray, P. G. & Renehan, A. G. Growth hormone, the insulin-like growth factor axis, insulin and cancer risk. *Nature Reviews Endocrinology* **7**, 11–24 (2011).

597. Magruder, J. T., Elahi, D. & Andersen, D. K. Diabetes and pancreatic cancer: chicken or egg? *Pancreas* **40**, 339-351 (2011).

598. Nunney, L. Size matters: height, cell number and a person's risk of cancer. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20181743 (2018).

599. Peto, R. Epidemiology, multistage models, and short-term mutagenicity tests. *Int. J. Epidemiol.* **45**, 621–637 (2016).

600. Arnold, M. *et al.* Obesity and cancer: An update of the global impact. *Cancer Epidemiol.* **41**, 8–15 (2016).

601. Dalamaga, M., Diakopoulos, K. N. & Mantzoros, C. S. The role of adiponectin in cancer: a review of current evidence. *Endocr. Rev.* **33**, (2012).

602. Li, F. P. Soft-Tissue Sarcomas, Breast Cancer, and Other Neoplasms. *Annals of Internal Medicine* **71**, 747 (1969).

603. Liu, S., Kim, T.-H., Franklin, D. A. & Zhang, Y. Protection against High-Fat-Diet-Induced Obesity in MDM2 Mice Due to Reduced p53 Activity and Enhanced Energy Expenditure. *Cell Rep.* **18**, 1005–1018 (2017).

604. Bignon, Y. J. *et al.* Expression of a retinoblastoma transgene results in dwarf mice. *Genes Dev.* **7**, 1654–1662 (1993).

605. Matsuura, H. *et al.* Prolyl Hydroxylase Domain Protein 2 Plays a Critical Role in Diet-

Induced Obesity and Glucose Intolerance. *Circulation* **127**, 2078–2087 (2013).

606. Zhang, X. *et al.* Adipose tissue-specific inhibition of hypoxia-inducible factor 1{alpha} induces obesity and glucose intolerance by impeding energy expenditure in mice. *J. Biol. Chem.* **285**, 32869–32877 (2010).

607. Al Alwan, I. *et al.* Turner Syndrome Genotype and phenotype and their effect on presenting features and timing of Diagnosis. *Int. J. Health Sci.* **8**, 195–202 (2014).

608. Rao, E. *et al.* Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature Genetics* **16**, 54–63 (1997).

609. Stuppia, L. *et al.* SHOX mutations detected by FISH and direct sequencing in patients with short stature. *J. Med. Genet.* **40**, E11 (2003).

610. Song, H. H., Shi, W., Xiang, Y.-Y. & Filmus, J. The loss of glypican-3 induces alterations in Wnt signaling. *J. Biol. Chem.* **280**, 2116–2125 (2005).

611. Filmus, J. & Capurro, M. The role of glypican-3 in the regulation of body size and cancer. *Cell Cycle* **7**, 2787–2790 (2008).

612. Neri, G., Gurrieri, F., Zanni, G. & Lin, A. Clinical and molecular aspects of the Simpson-Golabi-Behmel syndrome. *Am. J. Med. Genet.* **79**, 279–283 (1998).

613. UK Biobank. Genotyping of 500,000 UK Biobank participants: Description of sample processing workflow and preparation of DNA for genotyping. (2015) Online at: https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/genotyping_sample_workflow.pdf.

614. Voight, B. F. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

615. UK10K Consortium *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).

616. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).

617. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

618. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–6 (2014).

619. Carithers, L. J. & Moore, H. M. The Genotype-Tissue Expression (GTEx) Project. *Biopreservation and Biobanking* **13**, 307–308 (2015).

620. Gong, J. *et al.* PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.* **46**, D971–D976 (2018).

621. Jansen, R. *et al.* Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* **26**, 1444–1451 (2017).

622. Kundaje, A. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

623. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).

624. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Regulatory changes underlying expression differences within and between Drosophila species. *Nat. Genet.* **40**, 346–350 (2008).

625. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120362 (2013).

626. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics* **11**, 499–511 (2010).

627. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).

628. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).

629. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological

architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).

630. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **49**, 1126–1132 (2017).

631. Duncan, R. *et al.* A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes Dev.* **8**, 465–480 (1994).

632. Rothman, N. *et al.* A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat. Genet.* **42**, 978–984 (2010).

633. Figueroa, J. D. *et al.* Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum. Mol. Genet.* **23**, 1387–1398 (2014).

634. Ohtsubo, M., Theodoras, A. M., Schumacher, J., Roberts, J. M. & Pagano, M. Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. *Mol. Cell. Biol.* **15**, 2612–2624 (1995).

635. Chadee, D. N., Yuasa, T. & Kyriakis, J. M. Direct activation of mitogen-activated protein kinase kinase kinase MEKK1 by the Ste20p homologue GCK and the adapter protein TRAF2. *Mol. Cell. Biol.* **22**, 737–749 (2002).

636. Rodríguez-Berriguete, G. *et al.* MAP Kinases and Prostate Cancer. *Journal of Signal Transduction* **2012**, 1–9 (2012).

637. Pham, T. T., Angus, S. P. & Johnson, G. L. MAP3K1: Genomic Alterations in Cancer and Function in Promoting Cell Survival or Apoptosis. *Genes Cancer* **4**, 419–426 (2013).

638. Abreu-Martin, M. T., Chari, A., Palladino, A. A., Craft, N. A. & Sawyers, C. L. Mitogen-activated protein kinase kinase kinase 1 activates androgen receptor-dependent transcription and apoptosis in prostate cancer. *Mol. Cell. Biol.* **19**, 5143–5154 (1999).

639. Cosmic. COSMIC. (2020).

640. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).

641. Turnbull, C. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42**, 504–507 (2010).
642. Ahsan, H. *et al.* A genome-wide association study of early-onset breast cancer identifies PFKM as a novel breast cancer gene and supports a common genetic spectrum for breast cancer at any age. *Cancer Epidemiol. Biomarkers Prev.* **23**, 658–669 (2014).
643. Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.* **47**, 373–380 (2015).
644. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–61, 361e1–2 (2013).
645. Geh, E. *et al.* Mitogen-activated protein kinase kinase kinase 1 (MAP3K1) integrates developmental signals for eyelid closure. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17349–17354 (2011).
646. Pearlman, A. *et al.* Mutations in MAP3K1 cause 46,XY disorders of sex development and implicate a common signal transduction pathway in human testis determination. *Am. J. Hum. Genet.* **87**, 898–904 (2010).
647. Paranjpe, M. *et al.* Transcriptomic Analysis of MAP3K1 and MAP3K4 in the Developing Marsupial Gonad. *Sex Dev.* **13**, (2019).
648. Kelley, M. R. *DNA Repair in Cancer Therapy: Molecular Targets and Clinical Applications*. (Academic Press, 2011).
649. Benitez, A. *et al.* FANCA Promotes DNA Double-Strand Break Repair by Catalyzing Single-Strand Annealing and Strand Exchange. *Mol. Cell* **71**, 621–628.e4 (2018).
650. Nan, H. *et al.* Genome-wide association study identifies novel alleles associated with risk of cutaneous basal cell carcinoma and squamous cell carcinoma. *Hum. Mol. Genet.* **20**, 3718–3724 (2011).
651. Ransohoff, K. J. *et al.* Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. *Oncotarget* **8**, 17586–17592 (2017).

652. Chahal, H. S. *et al.* Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nat. Commun.* **7**, 12510 (2016).

653. Zhang, M. *et al.* Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum. Mol. Genet.* **22**, 2948–2959 (2013).

654. Chahal, H. S. *et al.* Genome-wide association study identifies novel susceptibility loci for cutaneous squamous cell carcinoma. *Nat. Commun.* **7**, 12048 (2016).

655. Wong, J. C. Y. Targeted disruption of exons 1 to 6 of the Fanconi Anemia group A gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia. *Human Molecular Genetics* **12**, 2063–2076 (2003).

656. Parmar, K., D'Andrea, A. & Niedernhofer, L. J. Mouse models of Fanconi anaemia. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **668**, 133–140 (2009).

657. Levine, A. J. & Oren, M. The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer* **9**, 749–758 (2009).

658. Nigro, J. M. *et al.* Mutations in the p53 gene occur in diverse human tumour types. *Nature* **342**, 705–708 (1989).

659. Ozaki, T. & Nakagawara, A. Role of p53 in Cell Death and Human Cancers. *Cancers*. **3**, 994–1013 (2011).

660. Hock, A. K. & Vousden, K. H. The role of ubiquitin modification in the regulation of p53. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 137–149 (2014).

661. Cheng, Q., Chen, L., Li, Z., Lane, W. S. & Chen, J. ATM activates p53 by regulating MDM2 oligomerization and E3 processivity. *EMBO J.* **28**, 3857–3867 (2009).

662. He, G. *et al.* Induction of p21 by p53 following DNA damage inhibits both Cdk4 and Cdk2 activities. *Oncogene* **24**, 2929–2943 (2005).

663. Lee, M.-H. & Lozano, G. Regulation of the p53-MDM2 pathway by 14-3-3 σ and other proteins. *Seminars in Cancer Biology* **16**, 225–234 (2006).

664. Michalak, E. M., Villunger, A., Adams, J. M. & Strasser, A. In several cell types tumour suppressor p53 induces apoptosis largely via Puma but Noxa can contribute. *Cell Death Differ.* **15**, 1019–1029 (2008).

665. Li, M. *et al.* Mono- versus polyubiquitination: differential control of p53 fate by Mdm2. *Science* **302**, 1972–1975 (2003).

666. Lohrum, M. A., Woods, D. B., Ludwig, R. L., Bálint, E. & Vousden, K. H. C-terminal ubiquitination of p53 contributes to nuclear export. *Mol. Cell. Biol.* **21**, 8521–8532 (2001).

667. Kung, C.-P. & Murphy, M. E. The role of the p53 tumor suppressor in metabolism and diabetes. *Journal of Endocrinology* **231**, R61–R75 (2016).

668. Labuschagne, C. F., Zani, F. & Vousden, K. H. Control of metabolism by p53 – Cancer and beyond. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1870**, 32–42 (2018).

669. Liang, Y., Liu, J. & Feng, Z. The regulation of cellular metabolism by tumor suppressor p53. *Cell & Bioscience* **3**,9 (2013).

670. Itahana, Y. & Itahana, K. Emerging Roles of p53 Family Members in Glucose Metabolism. *Int. J. Mol. Sci.* **19**, (2018).

671. Matoba, S. p53 Regulates Mitochondrial Respiration. *Science*. **312**, 1650–1653 (2006).

672. Lee, P., Vousden, K. H. & Cheung, E. C. TIGAR, TIGAR, burning bright. *Cancer Metab* **2**, 1 (2014).

673. Boidot, R. *et al.* Regulation of monocarboxylate transporter MCT1 expression by p53 mediates inward and outward lactate fluxes in tumors. *Cancer Res.* **72**, 939–948 (2012).

674. Hu, W. *et al.* Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7455–7460 (2010).

675. Jones, R. G. *et al.* AMP-activated protein kinase induces a p53-dependent metabolic checkpoint. *Mol. Cell* **18**, 283–293 (2005).

676. Akeno, N., Miller, A. L., Ma, X. & Wikenheiser-Brokamp, K. A. p53 suppresses

carcinoma progression by inhibiting mTOR pathway activation. *Oncogene* **34**, 589–599 (2015).

677. Hasty, P., Sharp, Z. D., Curiel, T. J. & Campisi, J. mTORC1 and p53: Clash of the gods? *Cell Cycle* **12**, 20 (2013).

678. Stacey, S. N. *et al.* Germline sequence variants in TGM3 and RGS22 confer risk of basal cell carcinoma. *Hum. Mol. Genet.* **23**, 3045–3053 (2014).

679. Stacey, S. N. *et al.* New basal cell carcinoma susceptibility loci. *Nat. Commun.* **6**, 6825 (2015).

680. Melin, B. S. *et al.* Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet.* **49**, 789–794 (2017).

681. Stacey, S. N. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.* **43**, 1098–1103 (2011).

682. Guha, T. & Malkin, D. Inherited TP53 Mutations and the Li–Fraumeni Syndrome. *Cold Spring Harbor Perspectives in Medicine* **7**, a026187 (2017).

683. Swaminathan, M. *et al.* Hematologic malignancies and Li-Fraumeni syndrome. *Cold Spring Harb Mol Case Stud* **5**, (2019).

684. Bond, G. L. *et al.* A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell* **119**, 591–602 (2004).

685. Wang, P.-Y. *et al.* Increased oxidative metabolism in the Li-Fraumeni syndrome. *N. Engl. J. Med.* **368**, 1027–1032 (2013).

686. Wang, P.-Y. *et al.* Inhibiting mitochondrial respiration prevents cancer in a mouse model of Li-Fraumeni syndrome. *Journal of Clinical Investigation* **127**, 132–136 (2016).

687. Wheaton, W. W. *et al.* Metformin inhibits mitochondrial complex I of cancer cells to reduce tumorigenesis. *Elife* **3**, e02242 (2014).

688. Valdez, J. M., Nichols, K. E. & Kesserwan, C. Li-Fraumeni syndrome: a paradigm for the

understanding of hereditary cancer predisposition. *Br. J. Haematol.* **176**, 539–552 (2017).

689. Wang, Z. *et al.* Further Confirmation of Germline Glioma Risk Variant rs78378222 in TP53 and Its Implication in Tumor Tissues via Integrative Analysis of TCGA Data. *Hum. Mutat.* **36**, 684–688 (2015).

690. Zhao, Y., Li, X. & Zhu, S. rs78378222 polymorphism in the 3'-untranslated region of TP53 contributes to development of age-associated cataracts by modifying microRNA-125b-induced apoptosis of lens epithelial cells. *Mol. Med. Rep.* **14**, 2305–2310 (2016).

691. Haworth, S. *et al.* Low-frequency variation in TP53 has large effects on head circumference and intracranial volume. *Nat. Commun.* **10**, 357 (2019).

692. Deng, Q. *et al.* Tissue-specific microRNA expression alters cancer susceptibility conferred by a TP53 noncoding variant. *Nat. Commun.* **10**, 5061 (2019).

693. Williams, A. B. & Schumacher, B. p53 in the DNA-Damage-Repair Process. *Cold Spring Harbor Perspectives in Medicine*. **6**, a026070 (2016).

694. Yan, B. *et al.* mTORC1 regulates PTHrP to coordinate chondrocyte growth, proliferation and differentiation. *Nat. Commun.* **7**, 11151 (2016).

695. Karner, C. M., Lee, S.-Y. & Long, F. Bmp Induces Osteoblast Differentiation through both Smad4 and mTORC1 Signaling. *Mol. Cell. Biol.* **37**, (2017).

696. Chen, J. & Long, F. mTORC1 Signaling Promotes Osteoblast Differentiation from Preosteoblasts. *PLoS One* **10**, e0130627 (2015).

697. Zhang, X. *et al.* Adipose mTORC1 Suppresses Prostaglandin Signaling and Beige Adipogenesis via the CRTC2-COX-2 Pathway. *Cell Rep.* **24**, 3180 (2018).

698. Glass, D. J. Molecular mechanisms modulating muscle mass. *Trends in Molecular Medicine*. **9**, 344–350 (2003).

699. Risson, V. *et al.* Muscle inactivation of mTOR causes metabolic and dystrophin defects leading to severe myopathy. *J. Cell Biol.* **187**, 859–874 (2009).

700. Yoon, M.-S. mTOR as a Key Regulator in Maintaining Skeletal Muscle Mass. *Front.*

Physiol. **8**, 788 (2017).

701. Li, X. *et al.* Height, height-related SNPs, and risk of non-melanoma skin cancer. *Br. J. Cancer* **116**, 134–140 (2017).

702. Schork, N. J., Murray, S. S., Frazer, K. A. & Topol, E. J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **19**, 212–219 (2009).

703. Allen, H. L. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).

704. Chan, Y. *et al.* Genome-wide Analysis of Body Proportion Classifies Height-Associated Variants by Mechanism of Action and Implicates Genes Important for Skeletal Development. *The American Journal of Human Genetics* **96**, 695–708 (2015).

705. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).

706. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

707. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).

708. Loos, R. J. & Yeo, G. S. The bigger picture of FTO: the first GWAS-identified obesity gene. *Nat. Rev. Endocrinol.* **10**, 51-61 (2014).

709. Liang, B., Ding, H., Huang, L., Luo, H. & Zhu, X. GWAS in cancer: progress and challenges. *Mol. Genet. Genomics* **295**, 537–561 (2020).

710. Bossé, Y. & Amos, C. I. A Decade of GWAS Results in Lung Cancer. *Cancer Epidemiol. Biomarkers Prev.* **27**, 363–379 (2018).

711. Fachal, L. & Dunning, A. M. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Current Opinion in Genetics & Development* **30**, 32–41 (2015).

712. Thorgeirsson, T. E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–642 (2008).

713. Sud, A., Kinnersley, B. & Houlston, R. S. Genome-wide association studies of cancer: current insights and future perspectives. *Nat. Rev. Cancer* **17**, 692–704 (2017).

714. Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39**, 870–874 (2007).

715. Garcia-Closas, M. *et al.* Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet.* **4**, e1000054 (2008).

716. Udler, M. S. *et al.* FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum. Mol. Genet.* **18**, 1692–1703 (2009).

717. Salehi, Z., Afzali, S., Shabanipour, S. & Rahimi, A. Evaluation of FGFR2 gene polymorphism in women with breast cancer. *Cell. Mol. Biol.* **61**, 94–97 (2015).

718. Campbell, T. M. *et al.* FGFR2 risk SNPs confer breast cancer risk by augmenting oestrogen responsiveness. *Carcinogenesis* **37**, 741–750 (2016).

719. Wang, H. *et al.* Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci Adv* **5**, eaaw3538 (2019).

720. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49**, 1458–1467 (2017).

721. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).

722. Takeuchi, F. *et al.* Interethnic analyses of blood pressure loci in populations of East Asian and European descent. *Nat. Commun.* **9**, 5052 (2018).

723. Wan, E. S. *et al.* Genome-wide association analysis of body mass in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* **45**, 304–310 (2011).

724. Chiang, K.-M. *et al.* Genome-wide association study of morbid obesity in Han Chinese. *BMC Genetics* **20**, 97 (2019).

725. Kilpeläinen, T. O. *et al.* Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.* **43**, (2011).
726. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
727. Remy, I. & Michnick, S. W. Regulation of apoptosis by the Ft1 protein, a new modulator of protein kinase B/Akt. *Mol. Cell. Biol.* **24**, (2004).
728. van der Hoeven, F. *et al.* Programmed cell death is affected in the novel mouse mutant Fused toes (Ft). *Development* **120**, 2601–2607 (1994).
729. Burla, R. *et al.* AKTIP/Ft1, a New Shelterin-Interacting Factor Required for Telomere Maintenance. *PLoS Genet.* **11**, e1005167 (2015).
730. Gerken, T. *et al.* The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469–1472 (2007).
731. Sanchez-Pulido, L. & Andrade-Navarro, M. A. The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. *BMC Biochem.* **8**, 23 (2007).
732. Haupt, A. *et al.* Variation in the FTO Gene Influences Food Intake but not Energy Expenditure. *Experimental and Clinical Endocrinology & Diabetes* **117**, 194–197 (2008).
733. Speakman, J. R., Rance, K. A. & Johnstone, A. M. Polymorphisms of the FTO Gene Are Associated With Variation in Energy Intake, but not Energy Expenditure. *Obesity* **16**, 1961–1965 (2008).
734. Yamakawa, H. *et al.* Neuroligins 3 and 4X interact with syntrophin-gamma2, and the interactions are affected by autism-related mutations. *Biochem. Biophys. Res. Commun.* **355**, 41-46 (2007).
735. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. **28**, 166-174 (2019).
736. Winkler, T. W. *et al.* The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet.* **11**,

e1005378 (2015).

737. Hoffmann, T. J. *et al.* A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics*. **210**, 499–515 (2018).

738. Lu, Y. *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **7**, 10495 (2016).

739. Hübel, C. *et al.* Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **180**, 428–438 (2019).

740. Brina, D., Miluzio, A., Ricciardi, S. & Biffo, S. eIF6 anti-association activity is required for ribosome biogenesis, translational control and tumor progression. *Biochim. Biophys. Acta* **1849**, 830–835 (2015).

741. Clarke, K. *et al.* The Role of Eif6 in Skeletal Muscle Homeostasis Revealed by Endurance Training Co-expression Networks. *Cell Rep.* **21**, 1507–1520 (2017).

742. Brina, D. *et al.* eIF6 coordinates insulin sensitivity and lipid metabolism by coupling translation to transcription. *Nat. Commun.* **6**, 1–15 (2015).

743. Katayama, M. *et al.* Circulating Exosomal miR-20b-5p Is Elevated in Type 2 Diabetes and Could Impair Insulin Action in Human Skeletal Muscle. *Diabetes* **68**, 515–526 (2019).

744. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2018).

745. Hernandez Cordero, A. I. *et al.* Genome-wide Associations Reveal Human-Mouse Genetic Convergence and Modifiers of Myogenesis, CPNE1 and STC2. *Am. J. Hum. Genet.* **106**, 138 (2020).

746. Koyama-Nasu, R. *et al.* The critical role of cyclin D2 in cell cycle progression and tumorigenicity of glioblastoma stem cells. *Oncogene* **32**, 3840–3845 (2012).

747. Karagianni, P., Amazit, L., Qin, J. & Wong, J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol. Cell. Biol.* **28**,

705–717 (2008).

748. Wu, J. *et al.* The Rare Variant rs35356162 in UHRF1BP1 Increases Bladder Cancer Risk in Han Chinese Population. *Frontiers in Oncology*. **10**, 134 (2020).

749. Bi, X. *et al.* ILRUN , a Human Plasma Lipid GWAS Locus, Regulates Lipoprotein Metabolism in Mice. *Circulation Research*. **127**, 1347–1361 (2020).

750. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

751. Tomlinson, I. *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39**, 984–988 (2007).

752. Cui, R. *et al.* Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut*. **60**, 799–805 (2011).

753. Tanskanen, T. *et al.* Genome-wide association study and meta-analysis in Northern European populations replicate multiple colorectal cancer risk loci. *Int. J. Cancer* **142**, 540–546 (2018).

754. Tanikawa, C. *et al.* GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. *Carcinogenesis* **39**, 652–660 (2018).

755. Lu, Y. *et al.* Large-Scale Genome-Wide Association Study of East Asians Identifies Loci Associated With Risk for Colorectal Cancer. *Gastroenterology* **156**, 1455–1466 (2019).

756. Schmit, S. L. *et al.* Novel Common Genetic Susceptibility Loci for Colorectal Cancer. *J. Natl. Cancer Inst.* **111**, 146–157 (2019).

757. Lu, Y. *et al.* Identification of Novel Loci and New Risk Variant in Known Loci for Colorectal Cancer Risk in East Asians. *Cancer Epidemiol. Biomarkers Prev.* **29**, 477–486 (2020).

758. Schumacher, F. R. *et al.* Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat. Commun.* **6**, 7138 (2015).

759. Ling, H. *et al.* CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res.* **23**, (2013).

760. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).

761. Takata, R. *et al.* 12 new susceptibility loci for prostate cancer identified by genome-wide association study in Japanese population. *Nat. Commun.* **10**, 4422 (2019).

762. Brooks, J. Multiple loci identified in a genome-wide association study of prostate cancer. *Urologic Oncology: Seminars and Original Investigations*. **26**, 571 (2008).

763. Hayashi, H. *et al.* The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene* **34**, 199–208 (2015).

764. Meng, B., Xing, Y., Li, H., Gao, F. & Liu, Y. C. Knockdown of Long Noncoding RNA POU5F1B Promotes Radiosensitivity in Esophageal Carcinoma. *Med. Sci. Monit.* **25**, 1214–1219 (2019).

765. Pan, Y. *et al.* POU5F1B promotes hepatocellular carcinoma proliferation by activating AKT. *Biomed. Pharmacother.* **100**, 374–380 (2018).

766. Law, P. J. *et al.* Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* **10**, 2154 (2019).

767. Whiffin, N. *et al.* Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Human Molecular Genetics* **23**, 4729–4737 (2014).

768. Huyghe, J. R. *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nat. Genet.* **51**, 76–87 (2019).

769. Prognostic significance of stromal GREM1 expression in colorectal cancer. *Hum. Pathol.* **62**, 56–65 (2017).

770. Li, J. *et al.* A functional variant in GREM1 confers risk for colorectal cancer by disrupting a hsa-miR-185-3p binding site. *Oncotarget* **8**, 61318 (2017).

771. Hoffmann, T. J. *et al.* A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. *Cancer Discov.* **5**, 878–891 (2015).

772. Yang, Z., Zhao, S., Zhou, X., Zhao, H. & Jiang, X. PCAT-1: A pivotal oncogenic long

non-coding RNA in human cancers. *Biomed. Pharmacother.* **110**, 493–499 (2019).

773. Abrahamsson, P. A., Lilja, H., Falkmer, S. & Wadström, L. B. Immunohistochemical distribution of the three predominant secretory proteins in the parenchyma of hyperplastic and neoplastic prostate glands. *Prostate* **12**, 39–46 (1988).

774. Takeda, K. *et al.* Corepressive function of nuclear receptor coactivator 2 in androgen receptor of prostate cancer cells treated with antiandrogen. *BMC Cancer* **16**, 332 (2016).

775. Whitaker, H. C., Warren, A. Y., Eeles, R., Kote-Jarai, Z. & Neal, D. E. The potential value of microseminoprotein-β as a prostate cancer biomarker and therapeutic target. *The Prostate* **70**, 333–340 (2010).

776. Eeles, R. A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).

777. Takata, R. *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat. Genet.* **42**, 751–754 (2010).

778. Long, J. *et al.* Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet.* **6**, e1001002 (2010).

779. Lacson, J. C. A. *et al.* Genome-Wide Testing of Exonic Variants and Breast Cancer Risk in the California Teachers Study. *Cancer Epidemiol. Biomarkers Prev.* **26**, 1462–1465 (2017).

780. Fejerman, L. *et al.* Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. *Nat. Commun.* **5**, 5260 (2014).

781. Low, S.-K. *et al.* Identification of two novel breast cancer loci through large-scale genome-wide association study in the Japanese population. *Sci. Rep.* **9**, 17332 (2019).

782. Couch, F. J. *et al.* Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. *Nat. Commun.* **7**, 11375 (2016).

783. Chin, S. F. *et al.* High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* **8**, R215 (2007).

784. Jones, J. O. *et al.* TOX3 mutations in breast cancer. *PLoS One* **8**, e74102 (2013).

785. Zhang, L. & Long, X. Association of three SNPs in TOX3 and breast cancer risk: Evidence from 97275 cases and 128686 controls. *Sci. Rep.* **5**, 12773 (2015).

786. Taylor, E. M., Copsey, A. C., Hudson, J. J. R., Vidot, S. & Lehmann, A. R. Identification of the proteins, including MAGEG1, that make up the human SMC5-6 protein complex. *Mol. Cell. Biol.* **28**, 1197–1206 (2008).

787. Lange, E. M. *et al.* Genome-wide association scan for variants associated with early-onset prostate cancer. *PLoS One* **9**, e93436 (2014).

788. Eeles, R. A. *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat. Genet.* **41**, 1116–1121 (2009).

789. Tao, S. *et al.* A genome-wide search for loci interacting with known prostate cancer risk-associated genetic variants. *Carcinogenesis* **33**, 598–603 (2012).

790. Schumacher, F. R. *et al.* Author Correction: Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **51**, 363 (2019).

791. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).

792. Zhu, Z. *et al.* Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.* **145**, 537–549 (2020).

793. Lotta, L. A. *et al.* Association of Genetic Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease, and Cardiovascular Risk Factors. *JAMA* **320**, 2553–2563 (2018).

794. Ng, M. C. Y. *et al.* Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet.* **13**, e1006719 (2017).

795. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet.* **41**, 1122–1126 (2009).

796. Price, S. M., Stanhope, R., Garrett, C., Preece, M. A. & Trembath, R. C. The spectrum

of Silver-Russell syndrome: a clinical and molecular genetic study and new diagnostic criteria. *J. Med. Genet.* **36**, 837–842 (1999).

797. Demetriou, C. *et al.* Paternally Expressed, Imprinted Insulin-Like Growth Factor-2 in Chorionic Villi Correlates Significantly with Birth Weight. *PLoS ONE* **9** e85454 (2014).

798. Zhu, X. *et al.* C11orf21, a novel gene within the Beckwith-Wiedemann syndrome region in human chromosome 11p15.5. *Gene* **256**, 311–317 (2000).

799. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).

800. He, W. W. *et al.* A novel human prostate-specific, androgen-regulated homeobox gene (NKX3.1) that maps to 8p21, a region frequently deleted in prostate cancer. *Genomics* **43**, 69–77 (1997).

801. Bhatia-Gaur, R. *et al.* Roles for Nkx3.1 in prostate development and cancer. *Genes Dev.* **13**, (1999).

802. Bowen, C. *et al.* Loss of NKX3.1 expression in human prostate cancers correlates with tumor progression. *Cancer Res.* **60**, 6111–6115 (2000).

803. Zhang, P. *et al.* Gene expression profiles in the PC-3 human prostate cancer cells induced by NKX3.1. *Mol. Biol. Rep.* **37**, 1505–1512 (2010).

804. Lei, Q. *et al.* NKX3.1 stabilizes p53, inhibits AKT activation, and blocks prostate cancer initiation caused by PTEN loss. *Cancer Cell* **9**, 367–378 (2006).

805. Gurel, B. *et al.* NKX3.1 as a marker of prostatic origin in metastatic tumors. *Am. J. Surg. Pathol.* **34**, 1097–1105 (2010).

806. Chen, W. *et al.* Abcb10 physically interacts with mitoferrin-1 (Slc25a37) to enhance its stability and function in the erythroid mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 16263–16268 (2009).

807. Hsiao, C.-P., Wang, D., Kaushal, A. & Saligan, L. Mitochondria-related gene expression changes are associated with fatigue in patients with nonmetastatic prostate cancer receiving external beam radiation therapy. *Cancer Nurs.* **36**, 189–197 (2013).

808. Wang, M. *et al.* Large-scale association analysis in Asians identifies new susceptibility loci for prostate cancer. *Nat. Commun.* **6**, 8469 (2015).

809. Phornphutkul, C. *et al.* The effect of rapamycin on bone growth in rabbits. *Journal of Orthopaedic Research* **27** 1157–1161 (2009).

810. Liu, K. *et al.* ARHGEF38 as a novel biomarker to predict aggressive prostate cancer. *Genes & Diseases* **7**, 217–224 (2020).

811. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes & Development* **30**, 733–750 (2016).

812. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).

813. Riveros-McKay, F. *et al.* Genetic architecture of human thinness compared to severe obesity. *PLoS Genet.* **15**, e1007603 (2019).

814. Burla, R. *et al.* The telomeric protein AKTIP interacts with A- and B-type lamins and is involved in regulation of cellular senescence. *Open Biol.* **6**, (2016).

815. Donohoe, C. L., Lysaght, J., O'Sullivan, J. & Reynolds, J. V. Emerging Concepts Linking Obesity with the Hallmarks of Cancer. *Trends in Endocrinology & Metabolism* **28**, 46–62 (2017).

816. Cnattingius, S., Lundberg, F., Sandin, S., Gronberg, H. & Iliadou, A. Birth Characteristics and Risk of Prostate Cancer: the Contribution of Genetic Factors. *Cancer Epidemiology Biomarkers & Prevention* **18**, 2422–2426 (2009).

817. Bartonicek, N. *et al.* Intergenic disease-associated regions are abundant in novel transcripts. *Genome Biol.* **18**, 241 (2017).

818. Deveson, I. W. *et al.* Universal Alternative Splicing of Noncoding Exons. *Cell Syst* **6**, 245–255.e5 (2018).

819. Betts, J. A. *et al.* Long Noncoding RNAs CUPID1 and CUPID2 Mediate Breast Cancer Risk at 11q13 by Modulating the Response to DNA Damage. *Am. J. Hum. Genet.* **101**, 255–266 (2017).

820. Lieber, M. R. The Mechanism of Human Nonhomologous DNA End Joining. *Journal of Biological Chemistry* **283**, 1–5 (2008).

821. Lin, S.-L. *et al.* A Novel Long Non-Coding RNA-01488 Suppressed Metastasis and Tumorigenesis by Inducing miRNAs That Reduce Vimentin Expression and Ubiquitination of Cyclin E. *Cells* **9**, 1504 (2020).

822. Allott, E. H., Masko, E. M. & Freedland, S. J. Obesity and prostate cancer: weighing the evidence. *Eur. Urol.* **63**, 800–809 (2013).

823. Fu, L. *et al.* Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. *J. Infect.* **80**, (2020).

824. Cummings, M. J. *et al.* Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *medRxiv* (2020) doi:10.1101/2020.04.15.20067157.

825. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* **25**, 2000180 (2020).

826. Oran, D. P. & Topol, E. J. Prevalence of Asymptomatic SARS-CoV-2 Infection. *Annals of Internal Medicine* **173**, 362–367 (2020).

827. Tang, Y.-W., Schmitz, J. E., Persing, D. H. & Stratton, C. W. Laboratory Diagnosis of COVID-19: Current Issues and Challenges. *J. Clin. Microbiol.* **58**, e00512-20 (2020).

828. Winter, A. K. & Hegde, S. T. The important role of serology for COVID-19 control. *The Lancet Infectious Diseases* **20**, 758–759 (2020).

829. Lou, B. *et al.* Serology characteristics of SARS-CoV-2 infection after exposure and post-symptom onset. *Eur. Respir. J.* **56**, 2000763 (2020).

830. Chen, L., Xiong, J., Bao, L. & Shi, Y. Convalescent plasma as a potential therapy for COVID-19. *The Lancet Infectious Diseases* **20**, 398–400 (2020).

831. WHO | Pneumonia of unknown cause – China (2020) Online at: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/.

832. Perego, E. *et al.* Why the Patient-Made Term 'Long Covid' is needed. *Wellcome Open Research* **5**, 224 (2020).

833. National Insitute for Health and Care Excellence, SIGN, Royal College of General Practitioners. COVID-19 guideline scope: management of the long-term effects of COVID-19. (2020). Online at: https://www.rcgp.org.uk/about-us/news/2020/october/management-of-the-long-term-effects-of-covid-19.aspx.

834. Sudre, C. H. *et al.* Attributes and predictors of Long-COVID: analysis of COVID cases and their symptoms collected by the Covid Symptoms Study App. *Nat Med.* **27** 626-631 (2021).

835. Tedder, R. S. *et al.* Detection, characterization, and enrollment of donors of Ebola convalescent plasma in Sierra Leone. *Transfusion* **58**, 1289–1298 (2018).

836. Azkur, A. K. *et al.* Immune response to SARS-CoV-2 and mechanisms of immunopathological changes in COVID-19. *Allergy* **75**, 1564–1581 (2020).

837. Rosadas, C., Randell, P., Khan, M., McClure, M. O. & Tedder, R. S. Testing for responses to the wrong SARS-CoV-2 antigen? *Lancet* **396**, e23 (2020).

838. Qu, J. *et al.* Profile of Immunoglobulin G and IgM Antibodies Against Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Clin. Infect. Dis.* **71**, 2255–2258 (2020).

839. Huang, Y. *et al.* Obesity in patients with COVID-19: a systematic review and meta-analysis. *Metabolism* **113**, 154378 (2020).

840. Severe Covid-19 GWAS Group *et al.* Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).

841. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in Covid-19. *Nature* **591**, 92–98 (2021).

842. Greater London Authority (GLA). Coronavirus (COVID-19) Cases. London Datastore (2020). Online at: https://data.london.gov.uk/publisher/gla

843. Long, Q.-X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).

844. Lumley, S. F. *et al.* Antibodies to SARS-CoV-2 are associated with protection against reinfection. *medRxiv* 2020.11.18.20234369 (2020).

845. Rasa, S. *et al.* Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *J. Transl. Med.* **16**, (2018).

846. Comuzzie, A. G. *et al.* Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* **7**, e51954 (2012).

847. Malaponte, G. *et al.* IL-6-174 G>C and MMP-9-1562 C>T polymorphisms are associated with increased risk of deep vein thrombosis in cancer patients. *Cytokine* **62**, 64–69 (2013).

848. Sharma, A. *et al.* Impact of interleukin 6 promoter polymorphisms (−174 G > C, −572 G > C and −597 G > A) on plasma IL-6 levels and their influence on the development of DVT: a study from India. *Hematology* **23**, 833–838 (2018).

849. Ilya O. Blokhin, S. R. L. Mechanisms of thrombosis in obesity. *Curr. Opin. Hematol.* **20**, 437 (2013).

850. Katz, D. H., Wilson, J. G., Gerszten, R. E. & Jackson, Framingham, and Malmö Heart Study Group. Mining a GWAS of Severe Covid-19. *The New England Journal Of Medicine* **383**, 2589 (2020).

851. Jeffers, S. A. *et al.* CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15748–15753 (2004).

852. Lord, S. R. *et al.* Integrated Pharmacodynamic Analysis Identifies Two Metabolic Adaption Pathways to Metformin in Breast Cancer. *Cell Metab.* **28**, 679-688 (2018).

853. Ju, Y. S. *et al.* Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *Elife* **3**, e02935 (2014).

854. Skwarski, M. *et al.* Mitochondrial Inhibitor Atovaquone Increases Tumor Oxygenation and Inhibits Hypoxic Gene Expression in Patients with Non-Small Cell Lung Cancer.

Clin. Cancer Res. **27**, 2459–2469 (2021).

855. Zwart, N. R. *et al.* Quantification of Cerebral Blood Flow using arterial spin labeling in glioblastoma multiforme; challenges of calibration in the presence of oedema. SMRM 27th Annual Meeting Exhibition in Montreal, Canada. (2019) Online at: https://archive.ismrm.org/2019/2315.html.
856. Croal P. *et al.* Imaging pH in Glioblastoma Multiforme with CEST MRI (IMAGO Trial): Preliminary Findings. ISMRM British & Irish Chapter (2019) Online at: https://www.ismrm.org/british/2018FinalProgramme.pdf.
857. Harris B.H.L *et al.* Unpicking the links between obesity and cancer through obesity gene expression signatures OMH Symposium (2019). Online at: https://www.ndm.ox.ac.uk/events/university-events/omh-symposium-2019-1.
858. Flower, B. *et al.* Clinical and laboratory evaluation of SARS-CoV-2 lateral flow assays for use in a national COVID-19 seroprevalence survey. *Thorax* **75**, 1082–1088 (2020).
859. Harris, B. H. L. *et al.* Asymptomatic COVID-19 in a rehabilitation facility: evolution of the presence of nasopharyngeal SARS-CoV-2 and serological antibody responses. *J. Infect. Dis.* (2020) doi:10.1093/infdis/jiaa610.
860. Pereira, C. *et al.* The Association Between Antibody Response to Severe Acute Respiratory Syndrome Coronavirus 2 Infection and Post-COVID-19 Syndrome in Healthcare Workers. *J. Infect. Dis.* **223**, 1671-1676 (2021).
861. Tedder, R. *et al.* Detection and Quantification of Antibody to SARS-CoV-2 Receptor Binding Domain Provides Enhanced Sensitivity, Specificity and Utility. (2021).
862. Thwaites, R. S. *et al.* Elevated antiviral, myeloid and endothelial inflammatory markers in severe COVID-19. *medRxiv* 2020.10.08.20209411 (2020).