

Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug-metabolite atlas

Jun Liu^{1,2}, Lies Lahousse^{1,3}, Michel G. Nivard^{4,5}, Mariska Bot⁶, Lianmin Chen^{7,8}, Jan Bert van Klinken⁹⁻¹¹, Carisha S. Thesing⁶, Marian Beekman¹², Erik Ben van den Akker¹²⁻¹⁴, Roderick C. Slieker^{15,16}, Eveline Waterham¹⁷, Carla J.H. van der Kallen^{18,19}, Irene de Boer²⁰, Ruifang Li-Gao²¹, Dina Vojinovic¹, Najaf Amin¹, Djawad Radjabzadeh²², Robert Kraaij²², Louise J.M. Alferink²³, Sarwa Darwish Murad²³, André G. Uitterlinden^{1,22}, Gonneke Willemsen^{4,5}, Rene Pool^{4,5}, Yuri Milaneschi⁶, Diana van Heemst²⁴, H. Eka D. Suchiman¹², Femke Rutters¹⁵, Petra J.M. Elders²⁵, Joline W.J. Beulens¹⁵, Amber A.W.A. van der Heijden²⁵, Marleen M.J. van Greevenbroek^{18,19}, Ilja C.W. Arts^{19,26,27}, Gerrit L.J. Onderwater²⁰, Arn M.J.M. van den Maagdenberg^{9,20}, Dennis O. Mook-Kanamori^{21,28}, Thomas Hankemeier^{29,30}, Gisela M. Terwindt²⁰, Coen D.A. Stehouwer^{18,19}, Johanna M. Geleijnse¹⁷, Leen M. 't Hart^{12,15,16}, P. Eline Slagboom¹², Ko Willems van Dijk^{9,10,31}, Alexandra Zhernakova⁷, Jingyuan Fu^{7,8}, Brenda W.J.H. Penninx⁶, Dorret I. Boomsma^{4,5}, Ayse Demirkan^{1,7,32}, Bruno H.C. Stricker^{1,22,33}, Cornelia M. van Duijn^{1,2,29}

Affiliations:

1. Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands.
2. Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom.
3. Department of Bioanalysis, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium.
4. Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands.
5. Amsterdam Public Health Research Institute (APH), Amsterdam, the Netherlands.
6. Department of Psychiatry, Amsterdam University Medical Center (UMC), Vrije Universiteit, Amsterdam Public Health research institute, Amsterdam, the Netherlands.
7. Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands.
8. Department of Pediatrics, University Medical Center Groningen, Groningen, the Netherlands.
9. Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands.
10. Eindhoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands.
11. Department of Clinical Chemistry, Laboratory Genetic Metabolic Disease, Amsterdam University Medical Center, Amsterdam, the Netherlands.
12. Department of Biomedical Data Sciences, section of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.
13. Department of Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, the Netherlands.
14. Leiden Computational Biology Center, Leiden University Medical Center, Leiden, the Netherlands.

15. Department of Epidemiology and Biostatistics, Amsterdam University Medical Center (UMC), Vrije Universiteit, location VUmc, Amsterdam Public Health research institute, Amsterdam, the Netherlands.
16. Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands.
17. Division of Human Nutrition and Health, Wageningen University, Wageningen, the Netherlands.
18. Department of Internal Medicine, Maastricht University, Maastricht, the Netherlands.
19. School for Cardiovascular Diseases (CARIM), Maastricht University, Maastricht, the Netherlands.
20. Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands.
21. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.
22. Department of Internal Medicine, Erasmus MC, University Medical Center, Rotterdam, the Netherlands.
23. Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands.
24. Department of Internal Medicine, section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands.
25. Department of General Practice and Elderly Care Medicine, Amsterdam University Medical Center (UMC), Vrije Universiteit, Amsterdam Public Health research institute, Amsterdam, the Netherlands.
26. Department of Epidemiology, Maastricht University, Maastricht, the Netherlands.
27. Maastricht Center for Systems Biology (MaCSBio), Maastricht University, Maastricht, the Netherlands.
28. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands.
29. Leiden Academic Center for Drug Research, Leiden University, Leiden, the Netherlands.
30. Netherlands Metabolomics Center, Leiden, the Netherlands.
31. Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands.
32. Section of Statistical Multi-omics, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK.
33. Inspectorate of Healthcare, The Hague, the Netherlands.

Address:

Big Data Institute, Old Road Campus, Oxford, United Kingdom. OX3 7FL.

Correspondence

Correspondence and requests for materials should be addressed to Jun Liu (email: jun.liu@ndph.ox.ac.uk) or to Cornelia van Duijn (email: cornelia.vanduijn@ndph.ox.ac.uk).

Abstract

Progress in high-throughput metabolic profiling provides unprecedented opportunities to obtain insights into the effects of drugs on human metabolism. The Biobanking BioMolecular Research Infrastructure of the Netherlands (BBMRI-NL) has constructed an atlas of drug-metabolite associations for 87 commonly prescribed drugs and 150 clinically relevant plasma-based metabolites assessed by proton nuclear magnetic resonance (^1H -NMR). The atlas involves a meta-analysis of ten cohorts (18,873 persons) and uncovers 1,071 drug-metabolite associations after evaluating confounders including co-treatment. We show the effect estimates of statins on metabolites from the cross-sectional study are comparable to those from intervention and genetic observational studies. Further data integration links proton pump inhibitors to circulating metabolites, liver function, hepatic steatosis and the gut microbiome. Our atlas provides a tool for targeted experimental pharmaceutical research and clinical trials to improve drug efficacy, safety and repurposing. We provide a web-based resource for visualization of the atlas (<http://bbmri.researchlumc.nl/atlas/>).

In the past decade, metabolomics technology has developed rapidly¹, facilitating large-scale studies which highlighted the importance of differential molecular dynamics captured in a wide range of common complex diseases, including diabetes, cardiovascular disease, asthma and dementia²⁻⁹. The human metabolome is in part driven by the human genome and new genetic drivers of these metabolites continue to be revealed.¹⁰⁻¹³. The past decade has also seen major successes in understanding the relation of the human metabolome to the exposome, e.g. lifestyle, nutrition, environment and microbiome¹⁴⁻¹⁶. Although the use of drugs is recognized to have a major effect on the metabolism, our knowledge on the drug-metabolite associations [is](#) incomplete and limited to the most commonly prescribed drugs, e.g. statins, metformin and antihypertensives¹⁷⁻²². In addition, even for the commonly prescribed drugs, the metabolic and physiologic effects, including on- or off-target effects, are virtually unexplored. Mapping these unexplored drug-metabolite associations is crucial for pharmaco-epidemiological research and practice as it may offer new avenues to improve drug efficacy, enable repurposing of drugs²³⁻²⁵ and improve our understanding of the off-target effects of drugs occurring in an individual patient^{26,27}. However, pointing out such associations is complicated since confounding may occur due to the metabolic changes that are the cause or the consequence of the pathology for which the drug is prescribed. Furthermore, many patients are treated with multiple drugs for multiple diseases, raising the important question of whether drug-metabolite associations are confounded by co-treatment²⁸. Last not but least, longitudinal observations are often lacking for relatively rare off-target effects, forcing clinical decision making to be based on cross-sectional data.

The aim of the present study was to develop a comprehensive atlas of the associations between a wide range of commonly prescribed drugs (**Supplementary Table 1**) and 150 plasma-based metabolites as measured by proton nuclear magnetic resonance (¹H-NMR) platform of Nightingale Health—(**Supplementary Table 2**). The platform allows rapid and cost-effective characterization of metabolites in human blood and it has been successfully used globally to discover and validate disease-metabolite associations²⁹, such as diabetes³⁰, dementia⁶, cardiovascular diseases^{31,32}, migraine³³, graves' disease³⁴ and mortality^{35,36}. Nightingale Health is now being validated for use in clinical care, which makes it timely to develop a pharmacological metabolomics atlas for this platform that can be used in research as well as clinical care. The term “metabolite” throughout the manuscript does not refer to the products of drug metabolism but to endogenous metabolites that are naturally produced by an organism and in

this context includes lipoprotein particles as well. In the present paper, we work through a series of examples of applications of the atlas, including disentangling the disease effect of the drug-metabolite associations and exploring in-depth the interaction of metabolites and two drugs, statins and proton pump inhibitors (PPIs).

Results

Overall drug-metabolite associations

We meta-analyzed 12 datasets of ten Dutch cohorts (**Supplementary Table 3**, 18,873 individuals) from Biobanking and BioMolecular resources Research Infrastructure of the Netherlands (BBMRI-NL). We discovered 2,087 significant associations out of 13,050 meta-analyzed tests involving 87 drugs and 150 metabolites in model 1 with adjustment for age and sex (Bonferroni P-value threshold = 1.9×10^{-5}). The number of drug users ranged from 3,023 (16.0%, for lipophilic statin) down to 20 (0.11%, for leukotriene receptor antagonists). Among the 13,050 tests, 543 (4%) showed heterogeneity across datasets and for these we used the random-effect model to pool data across datasets. **Supplementary Table 4** shows all drug-metabolite associations tested across different models, as well as disease-metabolite associations. Over the metabolites studied, effect estimates derived from different datasets agreed convincingly (P-values ranging from 1.67×10^{-11} to 1.0×10^{-318} of pairwise correlation tests) (**Supplementary Figure 1** and **Supplementary Table 5**). **Figure 1** shows the associations of model 1 for the top 15 drugs that were associated with the largest number of metabolites. The 15 drugs belong to five clinical pharmacological groups: (1) six antihypertensives, i.e. selective beta-blockers, angiotensin II antagonists, ACE inhibitors, high-ceiling diuretics, low-ceiling diuretics and potassium-sparing agents, (2) two glucose-lowering drugs, i.e. metformin and sulfonamides-urea derivatives, (3) two lipid-modifying drugs, i.e. lipophilic statin and hydrophilic statin, (4) three other cardiovascular-related drugs, i.e. vitamin K antagonists, antithrombotic agents-acetylsalicylic acid and digoxin, and (5) two others including PPIs and selective serotonin reuptake inhibitors (SSRIs). Thirteen of the top 15 drugs that were associated with the largest number of metabolites were cardio-metabolic related drugs, which may be for a large part explained by the fact that the numbers of users were large and the current metabolome spectrum contains mainly lipids and-which correlated with each other (**Extended Data Figure 1**).

Effects of BMI, smoking and co-treatment as major confounders

Next, we studied the potential confounding effect of BMI and smoking. In total, 1,640 of the 2,087 significant associations (78.6%) in model 1 were still significant after the adjustment for BMI and smoking in model 2 (**Extended Data Figure 2** and **Figure 1**). The drugs for which the evidence for association was most dramatically impacted by adjustment for BMI and smoking were SSRIs: 59 of the initial 65 SSRIs-metabolite associations (90.8%) were no longer significant after adjustment for BMI and smoking. A major impact of adjustment was also seen for two antihypertensives: 56 (60.9%) associations with high-ceiling-diuretics were no longer significant, and 53 (49.1%) associations with angiotensin-II-antagonists lost their significance. After we additionally excluded the confounding of other drugs by adjustment for co-treatments (**Extended Data Figure 3** and **Figure 2**), 1,071 significant associations remained to be investigated. For five out of six antihypertensives in the top 15 drugs (**Figure 2**), the associations with LDL and IDL particles were explained by co-treatments. Notably, statin use was correlated with antihypertensives and associated with LDL and IDL particles, which leads to a false discovery association of LDL and IDL particles and antihypertensives. Most antihypertensives associations disappeared after adjusting for co-treatment including statins, except for 15.4% (4/26) of the selective beta-blocker and all of the angiotensin II antagonists which remained significantly associated with LDL and IDL particles, suggesting that these associations are independent of co-treatments. In our epidemiological study, metformin was co-prescribed with hydrophilic statins and both drugs were associated to similar circulating metabolites, i.e. there were 85 metabolites associated with metformin in model 2, and 59 of these were also associated with hydrophilic statins. However, none of the metformin-metabolite associations were explained by hydrophilic statins, suggesting that metformin and hydrophilic statins are independently associated with the metabolites (**Figure 2**). These results above were confirmed by our sensitivity analysis from sub-samples of patients who use one drug only: all significant associations in this sensitivity analyses remained significant in the model with-adjusted for co-treatment adjusted-for (**Extended Data Figure 4**).

Examples of applications of the atlas

Effect of indicated disease: drug-metabolite associations explained by the indication

First, we tested whether indicated diseases causally related to the drug-related metabolites using genetic risk score of the disease as an instrumental variable in Mendelian randomization (MR) (**Supplementary Table 6, 7**). Second, we associated the drug-related metabolites with the indicated disease in those who were not receiving the treatment, i.e. the on-target-treatment-naïve population (**Supplementary Table 4**). For instance, in the current study, metformin use is associated with increasing alanine, but we also know that type 2 diabetes (causally by MR) increases alanine levels in the blood⁴. This finding raises the question of whether the disease (type 2 diabetes) or its endophenotype partially or fully explain the association of metformin and alanine. This hypothesis was supported by the finding that after excluding all metformin users, type 2 diabetes was still associated with increasing alanine levels (beta = 0.42, P-value = 8.3×10^{-19}). Integration of the findings on drug-metabolite and disease-metabolite associations suggests that alanine levels in blood are most likely raised by type 2 diabetes effect rather than by metformin effect.

Following the line of research outlined above, we noticed that hypertension or high blood pressure partially or fully explained the associations of very-low-density lipoprotein (VLDL) particles and various triglycerides with beta-blockers and low-ceiling diuretics. Depression partially or fully explained the association of estimated degree of unsaturation of fatty acids and SSRIs, but not for those high-density lipoprotein (HDL) particles. Notably, type 2 diabetes or its endophenotype, fasting glucose, partially or fully explained a substantial part of associations, including 98.8% of metabolite associations with metformin and 100% with sulfonamides-urea derivatives, based on a nominal significance level in the disease-metabolite associations (P-value < 0.05, **Figure 3**). With such a strict exclusion of effect of the indicated disease, we still found acetate was negatively associated with metformin effect, and there is no evidence that the relationship is resulted from the effect of type 2 diabetes or fasting glucose levels.

Effects of drugs in cross-sectional and longitudinal studies

We compared our results on statin-metabolite associations in the present cross-sectional study with that of the longitudinal study published earlier by Wurtz and co-workers¹⁷. In their paper, the changes of metabolite concentrations in blood (two time points per individual) were compared between 716 patients who started statin therapy during follow-up and 4,874 persistent non-users¹⁷. There are 48 metabolites that overlapped with our

study¹⁷, in which metabolite and statin use were assessed at the same time in 3,023 individuals who were using lipophilic statins and 15,850 non-users, providing a cross-sectional snapshot. Twenty-nine (60%) of the metabolites showed consistently significant results between the two studies (Figure 4A). We further checked the metabolite associations with genetic variant rs12916-T located in gene *HMGCR* (3-Hydroxy-3-Methylglutaryl-CoA Reductase). This genetic variant was used as an instrumental variable for the effect of statins as the protective T allele results in low functioning HMG-CoA reductase, which is one-of-the pharmacological targeted effects of statins^{17,37}. Figure 4A shows that 20 of the 29 associations (69.0%) were consistently and significantly associated with rs12916-T in both the cross-sectional and longitudinal analyses. The 20 statin-metabolite associations involved mainly fatty acids (30.0%) and non-HDL cholesterol and lipoprotein particles (50.0%). Meanwhile, 15 of the 19 metabolites (80%) that were inconsistently associated with statins between our study and the previous study¹⁷ were not associated with rs12916-T.

We additionally identified 35 of the tested 55 statin-related metabolites (63.6%) associated with rs12916-T in the same direction as with lipophilic statins (Figure 4B and Supplementary Table 8). Twenty-five of them are new and complement the findings of the above-mentioned study by Wurtz and co-workers¹⁷. The new metabolites emerging, by association with rs12916-T in our cross-sectional analyses, involved very small to medium VLDL particles, IDL particles, LDL particles and the total phosphatidylcholine and other choline.

Cross-omics analysis exploring the association of PPIs, circulating metabolites, liver function and gut microbiome

In our study, PPIs were found to be associated with 55 metabolites after adjustment of co-treatment (Figure 5A), involving small to extremely large VLDL, large HDL, triglycerides particles, mono-unsaturated fatty acids, isoleucine, creatinine and glycoprotein acetyls mainly α 1-acid glycoprotein (glycoprotein). These associations were validated by drug-dose-metabolite associations. Analysis in the population-based cohort, Rotterdam Study (n = 700), shows a high consistency of the association between PPI (yes/no) and metabolites and the used Defined Daily Doses in-of PPI (continuous)users and metabolites (Extended Data Figure 5).

PPIs are often used by patients with cirrhosis and in these patients PPIs are associated with infections and worsening prognosis³⁸. We next studied in Rotterdam Study (n = 3,436) whether the PPI-associated metabolites are

also associated to liver function, including biochemical variables of liver function test and hepatic steatosis. **Figure 5A** and **Figure 5B** show a high consistency of the patterns of association between PPIs and metabolites and between metabolites and liver function (**Supplementary Table 9**). The consistency of associations in terms of the number of significant associations overlapping is for hepatic steatosis 98.2% (54/55), gamma-glutamyl transferase (GGT) 80.0% (44/55) and alanine transaminase (ALT) 81.8% (45/55; positively associated), and 90.9% (50/55) for the ratio of aspartate transaminase and ALT (AST/ALT) and 69.1% (38/55) for total bilirubin (inversely associated). Of these liver function variables, total bilirubin and GGT were significantly associated with reported PPI use in Rotterdam Study (**Figure 5B** and **Supplementary Table 10**).

We then studied the PPI-associated metabolites in relation to microbial diversity and the abundance of microbiota that are pharmacologically driven by PPI use in population³⁹⁻⁴³. We found that 94.4% (51/54) of the metabolites associated with PPIs are also associated with gut microbial (alpha) diversity in a meta-analysis of 2,305 participants that did not use antibiotics (**Figure 5C** and **Supplementary Table 11**). Of the 92 gut microbiota of which the abundances were associated with PPI use³⁹, 45 were available to test the association with metabolites (**Supplementary Table 12**). We found that three common microbiota (phylum *Tenericutes*, class *Mollicutes* and family *Ruminococcaceae*) which showed reduced abundance in PPI users had a consistent metabolite association pattern with the PPI-metabolite association pattern but in the opposite direction (**Figure 5D** and **Supplementary Table 13**). The genera of *Scardovia* showed an increased abundance in the gut of patients using PPIs. Although the genera of *Scardovia* showed a similar metabolite association pattern to PPIs, ~~of note is that~~ only the association to glycoprotein reached statistical significance when adjusted for multiple testing (**Figure 5D**).

Discussion

To our knowledge, we performed the most comprehensive analysis of the interaction between 87 commonly prescribed drugs and as many as 150 circulating metabolites measured by ¹H-NMR in 18,873 individuals. We uncovered 1,071 drug-metabolite associations after adjustment for age, sex, BMI, smoking and co-treatment, covering a wide range of drug-metabolite associations which were not studied before. We also demonstrated three examples of applications of the atlas, disentangling disease (e.g. type 2 diabetes) and therapy (e.g. metformin)

effects, aligning longitudinal and genetic analysis with our large-scale cross-sectional findings, and ultimately, linking PPI-metabolite interactions to the gut microbiome abundance and liver function.

Although many of the metabolites cluster strongly in populations (**Extended Data Figure 1**), our analysis shows [that](#) the direction and significance of drug-metabolite associations are not always the same among different metabolites in the same cluster. This [is](#) especially true for VLDL and HDL particles. ~~This and~~ [is](#) consistent with previous studies of the role of lipid particle profiles and diseases^{4,6,31-34,44,45}. This is also true for amino acids. In the Rotterdam Study, histone is clustering strongly with leucine, valine and isoleucine (P-values of correlation tests 3.3×10^{-23}). But histone is negatively associated with selective beta-blocker use (**Figure 2**), and leucine, valine and isoleucine are positively associated with selective beta-blocker use. We showed that BMI is a major confounder of associations with SSRIs. The high proportion of elimination in the SSRIs-metabolite associations (90.8%) after adjustment for smoking and BMI may be explained by the fact that body weight is a strong determinant of circulating metabolites and significant weight loss when not dieting or weight gain is part of the diagnostic criteria for depression⁴⁶. After adjustment for co-treatment, the similar significant association patterns between different drugs (e.g. angiotensin II antagonists and metformin) may imply that drug-metabolite associations are independently associated with a similar shift in metabolism, but this is only true if the pathology for which the two drugs are prescribed does not explain the drug-metabolite association. For instance, if metabolic syndrome is associated with a shift in circulating metabolites, this may result in a false discovery association with drugs often prescribed to these patients (e.g. statins, antihypertensives and metformin). This type of confounding was further addressed by investigating whether drug-metabolite associations are related to the pathology (e.g. diabetes, hypertension, dyslipidemia) that indicated prescription. As a typical metabolic disorder, evidences [show](#) [s](#) that type 2 diabetes explains a substantial [part of](#) glucose-lowering-drug-metabolite associations. The validation of the effects awaits clinical trials or prospective studies, but our example illustrates how the drug-metabolite atlas can be used in combination with disease-metabolite studies to tease out drug and disease effects and generate testable hypotheses for future trials. We further showed that to some extent the statin-metabolite associations in a large-scale cross-sectional study can mimic that of longitudinal effect of statin administration, which are preferred from a

methodological perspective. It is strengthened as the two studies are benchmarked by MR. These findings suggest that the atlas does yield informative associations that may be tested in future trials and follow-up studies.

The third and by far the most exciting example integrates the atlas data into state-of-the-art research questions. The finding that PPIs are associated with lower gut microbial diversity and a shift of the composition of the gut microbiome has been long recognized^{39,41,47}. Interestingly, a recent study⁴⁸ reported that non-diabetic obese patients with hepatic steatosis have low microbial gene richness and increased genetic potential for processing of dietary lipids and dysregulation of branched-chain amino acid metabolism, which is very much consistent with our findings. Zooming in into oral bacteria, genus *Scardovia* is found to be increased in the gut microbiome of PPI users³⁹. This raises the hypothesis that due to the PPI related changes of the gastric acid secretion in stomach, these microbiotas are reaching the gastrointestinal tract, very similar to the mechanism described in mice⁴⁰ and in the study of human gut microbiome in patients with liver cirrhosis⁴⁹. Genus *Scardovia* was most strongly and significantly associated with glycoprotein, which is an intriguing metabolite from a clinical and epidemiological perspective, as this acute phase glycoprotein is synthesized in the liver⁵⁰ and associated to a wide spectrum of incident diseases⁵¹, such as cardiovascular disease⁵², type 2 diabetes⁵³, cognition⁶ and all-cause mortality³⁶. A key question to answer in future studies is to what extent glycoprotein plays a mediating role in the relation of gut microbiome and morbidity. Our analysis validated the previous findings that human gut microbiome is changed in patients with liver cirrhosis⁴⁹ and withdrawal of PPIs in the cirrhosis patients decrease oral-origin taxa³⁸ in a general population study which has very low prevalence of severe diseases such as advanced liver or kidney disease (less than 3%). Our study also showed associations of PPIs with liver function variables, gut microbiota and metabolites in the blood circulation. Again, a longitudinal or intervention study is still required to examine this hypothesis.

Another note of interest is that the experimental study of the effect of PPIs on the gut microbiome in patients with cirrhosis is based on omeprazole³⁸. If we compared the different drugs that are included in the PPI category, we found that omeprazole is indeed associated to the metabolites identified in the drug category analysis (**Extended Data Figure 6**). However, we also found that other drugs such as lansoprazole are even more strongly and significantly associated, while the association to rabeprazole and esomeprazole is less strong and non-significant. Also these are interesting findings to follow-up.

221 This first comprehensive drug-metabolite atlas provides a basis for future exploration of drug-metabolite
222 interactions, using the omics-based approach as we used, or other (un)targeted experimental and longitudinal
223 pharmaceutical research in the future. Our study includes examples of how to use the atlas which can be extended
224 to other settings. We have limited the atlas to the most common drugs, but the atlas can be extended in the future
225 for more rare drugs as such data for this platform are generated in larger cohorts such as UK Biobank. These “mega”
226 cohorts would also allow studying the interaction of multiple drugs intake with sufficient statistical power
227 systematically. On the other hand, the current atlas can be a starting point for the future researches which focus on
228 certain limited number of drugs with metabolomics to check drug interactions. Another future challenge is to extend
229 the atlas to a wider range of metabolites measured by other platforms (e.g. mass spectrometry) and tissues (e.g.
230 urine). The use of MR is a strength of the current study, as it enables us to disentangle the effect of drugs and
231 indicated diseases. However, we are not always able to capture strong instruments for the MR test, which may
232 reduce the power of our analyses aiming to exclude the disease effects. Since our knowledge of the genes mimicking
233 effects of drugs and diseases is rapidly growing, we are optimistic that more powerful genetic instrumental variables
234 will be identified in the near future, opening windows of opportunities to MR analyses in pharmacometabolomics
235 research and in clinical trials.

236 Our comprehensive *in vivo* reference atlas will empower future clinical and pharmacological research in a
237 number of areas. These not only advance knowledge on the mechanisms of on-target drug effects as well as off-
238 target drug effects but may also provide evidence for the discovery of novel therapeutic applications of known drugs.
239 By making the atlas freely available through a web-based browser with downloadable datasets
240 (<http://bbmri.researchlumc.nl/atlas/>), we hope to facilitate the use of the data by pharmacists, drug developers and
241 clinical researchers on their drug or disease of interest.

Acknowledgements

The authors gratefully acknowledge all the participants included in the cohorts. We also acknowledge the BBMRI Metabolomics Consortium (See **Supplementary Information**, <http://www.bbMRI.nl/omics-metabolomics/>) funded by Biobanking and BioMolecular resources Research Infrastructure of the Netherlands (BBMRI-NL), a research infrastructure financed by the Dutch government through Netherlands Organisation for Scientific Research (NWO) (Grant Nos. 184.021.007 and 184033111). This work is part of the CardioVasculair Onderzoek Nederland (CVON 2012-03), the Common mechanisms and pathways in Stroke and Alzheimer's disease (CoSTREAM) project (www.costream.eu, grant agreement No 667375), Memorabel program (project number 733050814), Netherlands X-omics Research Infrastructure and U01-AG061359 NIA. The full list of funding information of each cohort is found in the **Cohort Acknowledgements**. J.L., C.M.v.D. and A.D. have used exchange grants from the Personalized pREvention of Chronic Diseases consortium (PRECeDI) (H2020-MSCA-RISE-2014). A.D. is supported by the Dutch Science Organization (ZonMW-VENI Grant 2015). L.C. is supported by a joint PhD fellowship from China Scholarship Council (CSC 201708320268) and University of Groningen. M.G.N. is supported by Royal Netherlands Academy of Science Professor Award (PAH/6635) to D.I.B.. D.O.M.K. is supported by the Dutch Science Organization (ZonMW- VENI Grant 916.14.023). B.H.Ch.S. received a grant from TransQST, 116030-2; IMI2. D.R. is funded by an Erasmus MC mRACE grant "Profiling of the human gut microbiome".

Cohort Acknowledgements

Erasmus Rucphen Family study (ERF)

We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work, and P. Snijders M.D. for his help in data collection. ERF was supported by the Consortium for Systems Biology (NCSB), both within the frame-work of the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO). ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of the 5th Framework Programme (No. QLG2-CT-2002-01254) as well as FP7 project EUROHEADPAIN (nr 602633). High-throughput analysis of the ERF data was supported by a joint grant from the Netherlands Organisation for Scientific Research (NWO) and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). High throughput metabolomics measurements of the ERF study has been supported by BBMRI-NL (Biobanking and Biomolecular Resources Research Infrastructure Netherlands).

Rotterdam Study

The authors thank the study subjects, the staff from Rotterdam Study, and the participating pharmacists and general practitioners. Rotterdam Study is supported by the Erasmus MC and Erasmus University Rotterdam; the Netherlands Organisation for Scientific Research (NWO), the Netherlands Organisation for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Netherlands Genomics Initiative (NGI), the Ministry of Education, Culture and Science, the Ministry of HealthWelfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. L.L. reports expert consultation for Boehringer Ingelheim GmbH and Novartis and unrestricted grants from AstraZeneca and Chiesi. The authors are also deeply grateful to nurse ultrasonographer, Mrs. van Wijngaarden (PW) for performing the abdominal ultrasonography and liver stiffness measurements. The generation and management of stool microbiome data for Rotterdam Study (Rotterdam Study III-2) were executed

by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. We thank Nahid El Faquir and Jolande Verkroost-Van Heemst for their help in sample collection and registration, and Pelle van der Wal, Kamal Arabe, Hedayat Razawy and Karan Singh Asra for their help in DNA isolation and sequencing. Furthermore, we thank drs. Jeroen Raes and Jun Wang (KU Leuven, Belgium) for their guidance in 16S rRNA profiling and dataset generation.

Netherlands Twin Register (NTR)

We are grateful to all twins and their relatives for their continued participation in the Netherlands Twin Register. Funding was obtained from the Netherlands Organization for Scientific Research (NWO) and MagW/ZonMW grants 904-61-090, 985-10-002, 904-61-193,480-04-004, 400-05-717, Addiction- 31160008, Middelgroot-911-09-032, Spinozapremie 56-464-14192, Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL, 184.021.007); VUUniversity's Institute for Health and Care Research (EMGO1) and Neuroscience Campus Amsterdam (NCA); the European Community's Seventh Framework Program (FP7/2007-2013), ENGAGE (HEALTH-F4-2007-201413); and the European ScienceCouncil (ERCAAdvanced, 230374), "MGN is supported by a ZonMw grant: "Genetics as a research tool: A natural experiment to elucidate the causal effects of social mobility on health." (pnr: 531003014), ZonMw project: "Can sex- and gender-specific gene expression and epigenetics explain sex-differences in disease prevalence and etiology?" (pnr:849200011) and grant nr: R01AG054628 02S.

Netherlands Study of Depression and Anxiety (NESDA)

The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organization for Health Research and Development (ZonMw, grant nr. 10-000-1002) and through the financial contributions of participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum).

Leiden Longevity Study (LLS)

We thank all the participants of the Leiden Longevity Study. This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology, and the Netherlands Consortium for Healthy Ageing (grant nr. 050-060-810), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), by the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) Metabolomics Consortium funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO, grant nrs. 184.021.007 and 184033111).

LifeLines-DEEP cohort (LifeLines DEEP)

We thank participants and staff of LifeLines DEEP cohort for their collaboration. We thank Jackie Dekens, Mathieu Platteel, Astrid Maatman and Jody Arends for management and technical support. This project was funded by the Netherlands Heart Foundation (IN-CONTROL CVON grant nr. 2012-03 to A.Z., and J.F.; by the Netherlands Organization for Scientific Research (NWO) (NWO-VIDI 864.13.013 to J.F., NWO-VIDI 016.178.056 to A.Z.; by the European Research Council Starting Grant 715772 to A.Z.; A.Z. also holds a Rosalind Franklin Fellowship from the University of Groningen.

Hoorn Diabetes Care System Cohort (Hoorn DCS)

The authors thank participants of this study and research staff of the Diabetes Care System West-Friesland. High throughput metabolomics measurements of the DCS study has been supported by BBMRI-NL (Biobanking and Biomolecular Resources Research Infrastructure Netherlands) and the Parelsnoer Initiative, which is part of and is funded by the Dutch Federation of University Medical Centres and from 2007 to 2011 received initial funding from the Dutch Government. To perform additional research in (subsamples of) the DCS cohort, funding has been received from several instances including the VUMC, the Dutch Federation of University Medical Centres, health insurers, the Netherlands Organisation for Scientific Research (NWO), the Dutch Organization for Health Research and Development (ZonMw), the Dutch Diabetes Foundation, the European Foundation for the Study of Diabetes, International Diabetes Federation, the European Innovative Medicine Initiative and European Union.

Alpha Omega Cohort

The Alpha Omega Cohort is registered with ClinicalTrials.gov. Identifier: NCT03192410. The Alpha Omega Cohort was funded by the Netherlands Heart Foundation (grant nr. 200T401) and the National Institutes of Health (NIH, grant nr. R01HL076200). J.M.G. received funding from Unilever for analyses of dietary and circulating fatty acids in the Alpha Omega Cohort. High throughput metabolomics measurements of the Alpha Omega Cohort has been supported by BBMRI-NL (Biobanking and Biomolecular Resources Research Infrastructure Netherlands).

The Maastricht Study (TMS)

This study was supported by the European Regional Development Fund via OP-Zuid, the Province of Limburg, the Dutch Ministry of Economic Affairs (grant nr. 31O.041), Stichting De Weijerhorst (Maastricht, the Netherlands), the Pearl String Initiative Diabetes (Amsterdam, the Netherlands), CARIM School for Cardiovascular Diseases (Maastricht, the Netherlands), Stichting Annadal (Maastricht, the Netherlands), Health Foundation Limburg (Maastricht, the Netherlands) and by unrestricted grants from Janssen-Cilag B.V. (Tilburg, the Netherlands), Novo Nordisk Farma B.V. (Alphen aan den Rijn, the Netherlands) and Sanofi-Aventis Netherlands B.V. (Gouda, the Netherlands).

Leiden University Migraine Neuro Analysis (LUMINA)

The Leiden University Migraine Neuro-Analysis (LUMINA) study is funded by grants obtained from the Netherlands Organization for the Health Research and Development (ZonMw; grant nr. 90700217) and VIDI (ZonMw; grant nr. 91711319) (to G.M.T.); The Netherlands Consortium for Systems Biology (NCSB) and the Centre for Medical System Biology (CMSB), both within the framework of the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) (to A.M.J.M.v.d.M.); the FP7 EU project EUROHEADPAIN (grant nr. 602633) (to A.M.J.M.v.d.M. & G.M.T.).

Netherlands Epidemiology of Obesity Study (NEO)

The authors of the NEO study thank all individuals who participated in the study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO

366 study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management
367 of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génomique (Paris,
368 France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the
369 Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research
370 Profile Area Vascular and Regenerative Medicine. NEO study is supported by the Netherlands Cardiovascular
371 Research Initiative: an initiative with support of the Dutch Heart Foundation (CVON2014-02 ENERGISE).

372

373

Author Contributions

A.D., J.F., J.M.G., L.L., J.L., M.N., L.M.H, C.M.v.D. and A.Z. contributed to study design. M.Beekman, J.W.J.B., D.I.B., I.d.B., P.J.M.E., J.F., J.M.G., D.O.M.K., G.L.J.O., B.P., F.R., P.E.S., C.D.A.S., B.H.S., L.M.H., G.M.T., A.M.J.M.v.d.M., A.A.W.A.v.d.H., G.J.H.v.d.K., K.W.v.D., C.M.v.D., M.M.J.v.G., G.W., R.K., S.D.M., A.G.U. and A.Z. contributed to cohort design and management. I.C.W.A., M.Beekman, D.I.B., I.d.B., J.F., J.M.G., T.H., G.L.J.O., R.P., P.E.S., C.D.A.S., B.H.S., H.E.D.S., L.M.H, G.M.T., A.M.J.M.v.d.M., A.A.W.A.v.d.H., G.J.H.v.d.K., C.M.v.D., M.M.J.v.G., D.v.H., G.W., D.V., N.A., D.R., R.K., S.D.M., A.G.U., and A.Z. contributed to cohort data collection. M.Beekman, M.Bot, L.C., I.d.B., J.M.G., L.L., R.L.G., J.L., M.N., R.C.S., L.M.H, C.T., E.B.v.d.A., D.V., D.R., L.J.M.A. and E.W. contributed to data analysis. J.B.v.K. contributed to web development. A.D., J.M.G., L.L., J.L. and C.M.v.D. contributed to writing of manuscript. I.C.W.A., M.Beekman, J.W.J.B., D.I.B., M.Bot, L.C., I.d.B., A.D., P.J.M.E., J.F., J.M.G., L.L., R.L.G., J.L., Y.M., D.O.M.K., M.N., G.L.J.O., B.P., R.P., F.R., R.C.S., B.H.S., L.M.H, G.M.T., C.T., E.B.v.d.A., A.M.J.M.v.d.M., A.A.W.A.v.d.H., G.J.H.v.d.K., K.W.v.D., C.M.v.D., M.M.J.v.G., G.W., D.V., N.A., R.K., L.J.M.A., S.D.M. and A.Z. contributed to critical review of manuscript.

389 **Competing interests**

390 D.O.M.K. is a part-time clinical research consultant for Metabolon, Inc. All other authors have nothing to disclose.

391 The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the

392 manuscripts.

393

394

References (for main text only)

1. Patti, G.J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **13**, 263-269 (2012).
2. Park, J.E., Lim, H.R., Kim, J.W. & Shin, K.H. Metabolite changes in risk of type 2 diabetes mellitus in cohort studies: A systematic review and meta-analysis. *Diabetes Res Clin Pract* **140**, 216-227 (2018).
3. McGarrah, R.W., Crown, S.B., Zhang, G.F., Shah, S.H. & Newgard, C.B. Cardiovascular Metabolomics. *Circ Res* **122**, 1238-1258 (2018).
4. Liu, J., *et al.* A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. *Diabetes* **66**, 2915-2926 (2017).
5. Wang, T.J., *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med* **17**, 448-453 (2011).
6. van der Lee, S.J., *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimers Dement* **14**, 707-722 (2018).
7. Mapstone, M., *et al.* Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* **20**, 415-418 (2014).
8. Thorburn, A.N., *et al.* Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nat Commun* **6**, 7320 (2015).
9. Mabalirajan, U., *et al.* Linoleic acid metabolite drives severe asthma by causing airway epithelial injury. *Sci Rep* **3**, 1349 (2013).
10. Illig, T., *et al.* A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* **42**, 137-141 (2010).
11. Kettunen, J., *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* **7**, 11122 (2016).
12. Draisma, H.H., *et al.* Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat Commun* **6**, 7208 (2015).
13. Suhre, K., *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54-60 (2011).
14. Beger, R.D., *et al.* Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics* **12**, 149 (2016).
15. Rappaport, S.M., Barupal, D.K., Wishart, D., Vineis, P. & Scalbert, A. The blood exposome and its role in discovering causes of disease. *Environ Health Perspect* **122**, 769-774 (2014).
16. Schutte, B.A., *et al.* The effect of standardized food intake on the association between BMI and (1)H-NMR metabolites. *Sci Rep* **6**, 38980 (2016).
17. Wurtz, P., *et al.* Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J Am Coll Cardiol* **67**, 1200-1210 (2016).
18. Altmaier, E., *et al.* Metabolomics approach reveals effects of antihypertensives and lipid-lowering drugs on the human metabolism. *Eur J Epidemiol* **29**, 325-336 (2014).
19. Elbadawi-Sidhu, M., *et al.* Pharmacometabolomic signature links simvastatin therapy and insulin resistance. *Metabolomics* **13**(2017).
20. Kaddurah-Daouk, R., *et al.* Lipidomic analysis of variation in response to simvastatin in the Cholesterol and Pharmacogenetics Study. *Metabolomics* **6**, 191-201 (2010).
21. Xu, T., *et al.* Effects of metformin on metabolite profiles and LDL cholesterol in patients with type 2 diabetes. *Diabetes Care* **38**, 1858-1867 (2015).
22. t Hart, L.M., *et al.* Blood Metabolomic Measures Associate With Present and Future Glycemic Control in Type 2 Diabetes. *J Clin Endocrinol Metab* **103**, 4569-4579 (2018).

23. Moosavinasab, S., *et al.* 'RE: fine drugs': an interactive dashboard to access drug repurposing opportunities. *Database (Oxford)* **2016**(2016).
24. Voora, D. & Shah, S.H. Pharmacometabolomics meets genetics: a "natural" clinical trial of statin effects. (*Journal of the American College of Cardiology*, 2016).
25. Wishart, D.S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nat Rev Drug Discov* **15**, 473-484 (2016).
26. Van Norman, G.A. Drugs, devices, and the FDA: Part 1: an overview of approval processes for drugs. *JACC: Basic to Translational Science* **1**, 170-179 (2016).
27. (FDA), U.S.F.D.A. 22 Case studies where phase 2 and phase 3 trials had divergent results <https://www.fda.gov/> (2017).
28. Brahma, D.K., Wahlang, J.B., Marak, M.D. & Ch Sangma, M. Adverse drug reactions in the elderly. *J Pharmacol Pharmacother* **4**, 91-94 (2013).
29. Wurtz, P., *et al.* Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol* **186**, 1084-1096 (2017).
30. Ahola-Olli, A.V., *et al.* Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia* (2019).
31. Ference, B.A., *et al.* Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. *JAMA* **321**, 364-373 (2019).
32. Holmes, M.V., *et al.* Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol* **71**, 620-632 (2018).
33. Onderwater, G.L.J., *et al.* Large-scale plasma metabolome analysis reveals alterations in HDL metabolism in migraine. *Neurology* **92**, e1899-e1911 (2019).
34. Struja, T., *et al.* Metabolomics for Prediction of Relapse in Graves' Disease: Observational Pilot Study. *Front Endocrinol (Lausanne)* **9**, 623 (2018).
35. Deelen, J., *et al.* A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat Commun* **10**, 3346 (2019).
36. Fischer, K., *et al.* Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: an observational study of 17,345 persons. *PLoS Med* **11**, e1001606 (2014).
37. Teslovich, T.M., *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713 (2010).
38. Bajaj, J.S., *et al.* Proton Pump Inhibitor Initiation and Withdrawal affects Gut Microbiota and Readmission Risk in Cirrhosis. *Am J Gastroenterol* **113**, 1177-1186 (2018).
39. Imhann, F., *et al.* Proton pump inhibitors affect the gut microbiome. *Gut* **65**, 740-748 (2016).
40. Llorente, C., *et al.* Gastric acid suppression promotes alcoholic liver disease by inducing overgrowth of intestinal Enterococcus. *Nat Commun* **8**, 837 (2017).
41. Jackson, M.A., *et al.* Proton pump inhibitors alter the composition of the gut microbiota. *Gut* **65**, 749-756 (2016).
42. Liu, R., *et al.* Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat Med* **23**, 859-868 (2017).
43. Pedersen, H.K., *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376-381 (2016).
44. Kontush, A. HDL particle number and size as predictors of cardiovascular disease. *Front Pharmacol* **6**, 218 (2015).
45. Ahola-Olli, A.V., *et al.* Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *BioRxiv*, 513648 (2019).

46. Mitchell, A.J., Vaze, A. & Rao, S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* **374**, 609-619 (2009).
47. Bajaj, J.S., *et al.* Systems biology analysis of omeprazole therapy in cirrhosis demonstrates significant shifts in gut microbiota composition and function. *Am J Physiol Gastrointest Liver Physiol* **307**, G951-957 (2014).
48. Hoyles, L., *et al.* Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. *Nat Med* **24**, 1070-1080 (2018).
49. Qin, N., *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59-64 (2014).
50. Bates, C., Adams, W. & Handschumacher, R. Control of the formation of uridine diphospho-N-acetyl-hexosamine and glycoprotein synthesis in rat liver. *Journal of Biological Chemistry* **241**, 1705-1712 (1966).
51. Kettunen, J., *et al.* Biomarker Glycoprotein Acetyls Is Associated With the Risk of a Wide Spectrum of Incident Diseases and Stratifies Mortality Risk in Angiography Patients. *Circ Genom Precis Med* **11**, e002234 (2018).
52. Akinkuolie, A.O., Buring, J.E., Ridker, P.M. & Mora, S. A novel protein glycan biomarker and future cardiovascular disease events. *J Am Heart Assoc* **3**, e001221 (2014).
53. Akinkuolie, A.O., Pradhan, A.D., Buring, J.E., Ridker, P.M. & Mora, S. Novel protein glycan side-chain biomarker and risk of incident type 2 diabetes mellitus. *Arterioscler Thromb Vasc Biol* **35**, 1544-1550 (2015).

Figure legends

Figure 1 Drug-metabolite associations in baseline model versus model 2 with adjustment for BMI and smoking

The top 15 drugs that were associated with the largest number of metabolites in the baseline model of linear regression are ordered and shown in the figure. The first letter of the ATC code is shown in front of the drug names to identify different categories. Sample sizes of the drug users and non-users in the baseline model and BMI and smoking adjusted regression model are shown behind the drug names, respectively. Red: positive significant associations in the baseline model ($P\text{-value} < 1.9 \times 10^{-5}$). Light red: positive non-significant associations in the baseline model ($P\text{-value} \geq 1.9 \times 10^{-5}$). Blue: negatively significant associations in the baseline model ($P\text{-value} < 1.9 \times 10^{-5}$). Light blue: negatively non-significant associations in the baseline model ($P\text{-value} \geq 1.9 \times 10^{-5}$). Star in boxes (*): The direction and significance status did not change between baseline model and model 2 ($P\text{-value} < 1.9 \times 10^{-5}$). Two-tailed tests were used.

Figure 2 Drug-metabolite associations in model 2 versus model 3 with adjustment for co-treatments

The top 15 drugs that were associated with the largest number of metabolites are ordered and shown in the figure. The first letter of the ATC code is shown in front of the drug names to identify different categories. Sample sizes of the drug users and non-users in regression model 2 and model 3 are shown behind the drug names, respectively. Red: positive significant associations in model 2 ($P\text{-value} < 1.9 \times 10^{-5}$). Light red: positive non-significant associations in in model 2 ($P\text{-value} \geq 1.9 \times 10^{-5}$). Blue: negatively significant associations in in model 2 ($P\text{-value} < 1.9 \times 10^{-5}$). Light blue: negatively non-significant associations in in model 2 ($P\text{-value} \geq 1.9 \times 10^{-5}$). Star in boxes (*): The direction and significance status did not change between model 2 and model 3 ($P\text{-value}$ threshold is multiple testing corrected per drug; See **Supplementary Table 4**). Two-tailed tests were used.

Figure 3 Drug-metabolite associations in model 3 versus significance after disentangling the indicated disease/endophenotype effect

The drugs in the top 15 drugs that were associated with the largest number of metabolites are ordered and shown in the figure. The first letter of the ATC code is shown in front of the drug names to identify the different categories.

Sample size of the drug users and non-users in regression model 3 and sample size of the cases and controls in the disease-metabolite associations are shown behind the drug names, respectively. Red: positive significant associations in model 3. Light red: positive non-significant associations in model 3. Blue: negatively significant associations in model 3. Light blue: negatively non-significant associations in model 3. Star in boxes (*): The significant associations confirmed after disentangling the disease/endophenotype effect (P-value < 0.05 in the disease-metabolite associations). Common in boxes (,): The associations confirmed after disentangling the disease/endophenotype effect (P-value threshold after multiple testing corrected per disease \leq P-value < 0.05; See **Supplementary Table 4**). Two-tailed tests were used.

Figure 4 Comparison of statin-metabolite associations between cross-sectional study, longitudinal study and genetic study

Figure 4A shows the comparison of statin-metabolite associations between the current cross-sectional study, longitudinal study¹⁷ and genetic study. The results of statin-metabolite associations in the longitudinal study (n = 716/4,874) are shown as the previous study in standard deviation (SD)-scaled metabolite concentration units (top axis) and ratio to the lowering effect estimate on LDL cholesterol (bottom axis). The results of rs12916-T-metabolite associations (n = 27,914) in Figure 4A are shown in effect estimate per SD and ratio to the lowering effect estimate on LDL cholesterol (bottom axis). Figure 4B shows the comparison of significant statin-metabolite associations in the cross-sectional study (n = 3,023/15,850 for lipophilic statin, n = 849/17,631 for hydrophilic statin) and genetic study (n = 24,925). The results of statin-metabolite associations are shown in the effect estimate (standardized metabolite concentration unit; bottom axis) and the results of rs12916-T-metabolite associations are shown in five times of the effect estimate (standardized metabolite concentration unit; top axis). The measure of center is the calculated estimates as mentioned above. The error bar is 95% confidence intervals (CI) which were statistically corrected for multiple testing. It means that if the error bar crosses the line zero, the association is not significant at the multiple testing significance level. * Statistical data were extracted from the previous longitudinal study¹⁷. Two-tailed tests were used.

Figure 5 Integrating data of PPIs, metabolites, liver function measurements and gut microbiome

The figure shows the significant results after integrating the directions of the associations of between PPIs, metabolites, liver function measurements and gut microbiome by linear regression. Red: positive association. Blue: negative association. The depth of red and blue presents the value of effect estimate per standard error. Grey: associations were not performed. Star in boxes (*): significance of the associations. Two-tailed tests were used.

Extended Data Figure 1 Correlation between metabolites in Rotterdam Study

The correlation matrix of metabolites were performed by Pearson's correlation ($n = 5,191$). The hierarchical cluster analysis was used in the clustering. Color in the boxes: correlation coefficient.

Extended Data Figure 2 Drug-metabolite associations in baseline model versus model 2 with adjustment for BMI and smoking

The drugs with at least one significant metabolite association by linear regression are shown. The first letter of the ATC code is shown in front of the drug names to identify different categories. Sample sizes of the drug users and non-users in the baseline model and BMI and smoking adjusted model are shown behind the drug names, respectively. Red: positive significant associations in the baseline model ($P\text{-value} < 1.9 \times 10^{-5}$). Light red: positive non-significant associations in the baseline model ($P\text{-value} \geq 1.9 \times 10^{-5}$). Blue: negatively significant associations in the baseline model ($P\text{-value} < 1.9 \times 10^{-5}$). Light blue: negatively non-significant associations in the baseline model ($P\text{-value} \geq 1.9 \times 10^{-5}$). Star in boxes (*): The direction and significance status did not change between baseline model and model 2 ($P\text{-value} < 1.9 \times 10^{-5}$). Two-tailed tests were used.

Extended Data Figure 3 Drug-metabolite associations in model 2 versus model 3 with adjustment for co-treatments

The drugs with at least one significant metabolite association by linear regression are shown. The first letter of the ATC code is shown in front of the drug names to identify different categories. Sample sizes of the drug users and non-users in model 2 and model 3 are shown behind the drug names, respectively. Red: positive significant

associations in model 2 ($P\text{-value} < 1.9 \times 10^{-5}$). Light red: positive non-significant associations in in model 2 ($P\text{-value} \geq 1.9 \times 10^{-5}$). Blue: negatively significant associations in in model 2 ($P\text{-value} < 1.9 \times 10^{-5}$). Light blue: negatively non-significant associations in in model 2 ($P\text{-value} \geq 1.9 \times 10^{-5}$). Star in boxes (*): The direction and significance status did not change between model 2 and model 3 ($P\text{-value}$ threshold is multiple testing corrected per drug; See **Supplementary Table 4**). Two-tailed tests were used.

Extended Data Figure 4 Drug-metabolite Associations in model 3 versus single drug test

The first letter of the ATC code is shown in front of the drug names to identify different categories. Single drug test: Association analysis (linear regression) in the sub-samples of patients who use one drug only (one-drug-users) and all-treatment-naïve controls. Sample size of the drug users and non-users in model 3 and the single drug test are shown behind the drug names, respectively. Red: positive significant associations in model 3 which are available for the single drug test. Light red: positive non-significant associations in model 3 or not available for the single drug test. Blue: negatively significant associations in model 3 which are available for the single drug test. Light blue: negatively non-significant associations in model 3 or not available for the single drug test. Star in boxes (*) The significant associations confirmed in the single drug test (P threshold is multiple testing corrected per drug; see **Supplementary Table 4**). Two-tailed tests were used.

Extended Data Figure 5 Association of PPI/dosage and the PPI-related metabolites

The association of dosage of PPI and metabolites were tested by linear regression ($n = 700$). The PPI-related metabolites were selected in model 3. DDD: Defined Daily Dose. (/): sample size of user/non-user. Red: positive association. Blue: negative association. The depth of the color refers to the association estimates. Star in boxes (*) significance after correcting for multiple test ($P\text{-value} < 0.004$). Two-tailed tests were used.

Extended Data Figure 6 Association of specific PPI drugs and the PPI-related metabolites

The association of PPI drugs and metabolites were tested by linear regression ($n = 700$). The PPI-related metabolites were selected in model 3. (/): sample size of user/non-user. Red: positive association. Blue: negative association. The

depth of the color refers to the association estimates. Star in boxes (*): significance after correcting for multiple test (P-value < 0.004). Two-tailed tests were used.

Extended Data Figure 7 The effect of population structure on metabolite clustering across datasets

Principal component (PC) analysis was performed using joint metabolite data from the cohorts (AlphaOmega, n = 877; ERF: n = 778; RS1: RS Dataset 1, n = 2,975; RS2: RS Dataset 2, n = 729; RS3: RS Dataset 3, n = 1,487; TMS: n = 854). Two-tailed tests were used.

Extended Data Figure 8 Correlation between drugs

The correlation matrix of metabolites were performed by Spearman's correlation (n = 6,631). The first letter of the ATC code is shown in front of the drug names to identify different categories. Sample size of the drug users and non-users is shown behind the drug names. The depth of the color refers to the correlation coefficients. Star in boxes (*): the positively significant correlations (P-value < 5.9×10^{-4}). Two-tailed tests were used.

Online Methods

Study population

The research was performed within the BBMRI-NL. The study included 18,873 individuals from 12 datasets of ten Dutch cohorts who had metabolites measured by Nightingale Health, drug information based on the Anatomical Therapeutic Chemical (ATC) Classification and clinical phenotypes which allow us to control for confounders. These cohorts included Rotterdam Study with three datasets (RS Dataset 1: n = 2,975, RS Dataset 2: n = 729, RS Dataset 3: n = 1,487)⁵⁴, Netherlands Twin Register (NTR, n = 3,563)⁵⁵, Netherlands Study of Depression and Anxiety (NESDA, n = 2,914)⁵⁶, Leiden Longevity Study (LLS, n = 1,873)⁵⁷, LifeLines DEEP cohort (n = 1,435)⁵⁸, Hoorn Diabetes Care System Cohort (Hoorn DCS, n = 995)⁵⁹, Alpha Omega Cohort (n = 877)⁶⁰, The Maastricht Study (TMS, n = 854)⁶¹, Erasmus Rucphen Family study (ERF, n = 778)⁶² and Leiden University Migraine Neuro-Analysis (LUMINA, n = 393)⁶³.

In the examples of application atlas, we additionally involved Netherlands Epidemiology of Obesity Study (NEO; n = 6,603)⁶⁴, which is an obese cohort but have adjusted for BMI in type 2 diabetes-metabolite associations by inverse probability weighting on BMI to make the results comparable with the Dutch general population. Cohort descriptions, specific data processing and ethical compliance can be found in **Supplementary Table 3**. All studies have been approved by their respective Institutional Review Boards Local research ethics committees, and all participants have provided written informed consent to the original study.

Metabolite measurements

The present study included 150 absolute-value-based metabolites measured by high-throughput ¹H-NMR metabolomics (Brainshake Ltd./Nightingale Health, Helsinki, Finland). The explanation of the metabolites was shown in **Supplementary Table 2**. The metabolites include the quantitative molecular data on 14 lipoprotein subclasses, apolipoprotein A-I and B, multiple cholesterol and triglyceride measures, albumin, various fatty acids as well as on numerous low-molecular-weight metabolites, including amino acids, glycolysis-related measures and ketone bodies. The 14 lipoprotein subclasses included IDL, six VLDL subclasses, three LDL subclasses and four HDL subclasses based on the particle diameters. The components of these lipoprotein subclasses were quantified on total lipids (L), total

cholesterol (C), particle concentration (P), phospholipids (PL), triglycerides (TG), free cholesterol (FC) and cholesterol esters (CE). The values of the representative coefficients of variations (CVs) for the metabolites ranged between 0.3% and 19.5% (mean 4.5%) and most values are comparable to the clinical chemistry assays^{11,65}.

The blood samples of different cohorts have been centralized in Leiden University Medical Center (LUMC) and were shipped to and analyzed by Nightingale Health as part of a national initiative. A standardized protocol of metabolite measurement was applied for all the cohorts following the comprehensive quantitative platform generated by Nightingale Health and described originally by Soininen et al^{11,65,66}. The protocol includes sample quality control and sample preparation, data storage and automated spectral analyses. The metabolite values which were suggested to be uncreditable in the quality control provided by Nightingale Health during the measurement procedure were treated as missing. Within the consortium, we checked and reported the distribution of zero values in our previous study by van den Akker, et al⁶⁷. The quality control was unified and included an in-depth evaluation of the consistency of findings across datasets, a metabolite correlation matrix and the principal component analysis on cohorts with different population structure. Pearson's correlation test was used to check the pairwise correlation of the overall estimate values of drug-metabolite associations in model 1 between datasets. We also checked the correlation matrix of metabolites in a population-based cohort, Rotterdam study (n = 5,191), by Pearson's correlation and hierarchical cluster analysis, reporting that the distinct clustering groups were in accordance with the biochemical pathways (**Extended Data Figure 1** and **Supplementary Table 14**). The effect of population structure on metabolite clustering was checked by principal component analysis using joint data from four cohorts that differ extremely in population: 1) one population-based study, Rotterdam Study⁵⁴, 2) one family-based study, ERF⁶², 3) one disease-based study, TMS⁶¹, which includes only patients with type 2 diabetes in the current dataset, and 4) a case-control study, Alpha Omega Cohort⁶⁰, including patients with cardiovascular disease and non-disease controls (details in **Supplementary Table 3**). The obvious difference between Alpha Omega Cohort and TMS underscore that meta-analysis should be performed instead of a joint analysis with pooled data (**Extended Data Figure 7**): the fixed-effect meta-analysis assumes a similar effect and structure over cohorts, while the random-effect meta-analysis allows for high heterogeneity across cohorts.

As some distributions of metabolites were skewed, we transformed the metabolite values in each cohort to normal distribution. We first added the value of one to all the metabolites before doing the natural logarithm transformation, to include samples labeled zero that had metabolite levels below the detectable value. Then we scaled these transformed values to standard deviation units.

Drug categories

The drug information was classified by ATC codes in each cohort. In brief, the drug information per cohort was obtained either from the pharmacy records or from the questionnaires during the interview. Details on drug information of each cohort can be found in **Supplementary Table 3**. We used the drug category instead of the individual compound in all the analyses. We merged drugs with similar chemical, pharmacodynamics, pharmacokinetics and/or therapeutic characteristics into one category. For the ATC codes used for combinations of active ingredients, we categorized them into separate categories if possible. We excluded categories with five or fewer users in each cohort or less than 20 users in total from all the cohorts. Thus, we ended up with 87 drug categories (**Supplementary Table 1**). The drug categorization was confirmed by two experienced pharmacologists: Lies Lahousse and Bruno H.C. Stricker. Throughout the text, the term *drug category* is further referred to *drug*. The individuals with metabolite and drug information available were included in the analysis.

Statistical analysis

All statistical analyses were performed using *R* statistical software and the two-tailed test was considered.

Association between drug and metabolite

To check for drug-metabolite associations, linear regression was performed in each cohort with drug use as an independent variable and metabolite as a dependent variable. Linear regression was used in the individual cohorts. The specific family relationship was considered in the three family-based cohorts (see details in **Supplementary Table 3**). In the baseline analysis, we used age and sex as the covariates (model 1). we additionally adjusted for smoking (current smoking: yes/no) which is a major common risk factor of pathology⁶⁸ and body mass index (BMI, kg/m²) which is a major determinant of circulating metabolites that captures the effects of diet and

physical activity⁶⁹ (model 2). Meta-analysis was performed with either the inverse-variance weighted fixed-effect model (no heterogeneity between cohorts) or a maximum likelihood random-effect model (significant heterogeneity between cohorts). The degree of heterogeneity was based on Cochran's Q test. The P-value threshold of both the Cochran's Q test and the meta-analysis was Bonferroni corrected with 30 independent equivalents of the 150 metabolites and 87 drugs tested ($P\text{-value} < 1.9 \times 10^{-5}$). Matrix Spectral Decomposition was used to calculate the number of independent equivalents⁷⁰ in the largest population-based dataset: RS Dataset 1. R-package "*metafor*" was used for the meta-analysis⁷¹.

Effects of co-treatment: drugs prescribed together

We next checked the potential confounding of drugs which were prescribed together (model 3) in each significant drug-metabolite pair. A co-treatment matrix with Spearman's correlation was made in the two population-based cohorts (Rotterdam Study and LifeLines DEEP, $n = 6,631$) separately and meta-analyzed. Potential confounding co-treatment for each drug-metabolite pair was defined if: (1) a drug was positively correlated with the target drug (explained as prescribed together, **Extended Data Figure 8** and **Supplementary Table 15**), and (2) this drug and the target drug were associated with the target metabolite in the same direction. We used the Bonferroni P-value correction with 85 drugs available in the co-treatment matrix ($P\text{-value} < 5.9 \times 10^{-4}$). We then performed the same regression analysis as above in each dataset (12 datasets) and meta-analyzed with age, sex, BMI, smoking and all the available confounding co-treatments as covariates in each significant drug-metabolite pair (model 3). A sensitivity analysis was performed in the sub-samples of patients who use one drug only (one-drug-users) and all-treatment-naïve controls adjusting for age, sex, BMI and smoking. We used the Bonferroni P-value threshold by correcting the independent equivalents of the number of tested significant metabolites for each drug.

Mendelian randomization to check the effect of indicated disease on metabolites

We further focused on the drugs in the top 15 drug lists that had the largest number of related metabolites and the metabolite associations after adjustment for co-treatments. We explored the confounding effect of the disease indicating the prescription of the drug by MR. MR is a statistical method which uses the effect of genetic variants determining an exposure and test its association with the outcome under study, based on the assumption that the genetic variant is inherited independent of the confounding variables⁷². Thus, we tested whether the genetic

determinants driving indicated diseases are also related to metabolites, using the genetic risk score of the disease as an instrumental variable of exposure. Genetic risk scores comprising > 5 genetic single nucleotide polymorphisms (SNPs) and explaining > 1% of variance in exposure were taken forward. For type 2 diabetes, we looked up the results from our previous well-organized MR research⁴, and 16 metabolites were found to be associated with either metformin or sulfonamides-urea derivatives. In brief, this MR research was a two-sample bi-direction MR study checking the causation of metabolites and type 2 diabetes and fasting glucose, following by biological knowledge-based sensitivity analysis to control for the pleiotropic effect of the SNPs in the instrumental variables⁴. We currently used the results of the backward MR that was checking the association of the genetic score of type 2 diabetes and metabolites.

For hypertension and depression, we performed two-sample MR based on the previous GWAS results on blood pressure⁷³ (n = 317,754), major depression⁷⁴ (n = 135,458 cases and n = 344,901 controls) and NMR metabolite GWAS¹¹ (n = 24,925). Among the 123 metabolites associated with either antihypertensives, 96 metabolites were available to perform MR with systolic and diastolic blood pressure. We also performed MR of major depression with six metabolites associated with SSRIs. We did not perform MR of dyslipidemia over the statin-associated metabolites because most of the metabolites were lipoproteins which are part of the dyslipidemia definition.

The R-package “*TwoSampleMR*” was used for the two-sample MR tests⁷⁵. Genetic loci of major depression were extracted from previous paper as its original GWAS was not available⁷⁴. The default pipeline in the *TwoSampleMR* package⁷⁵ was used. In brief, the genetic score was based on the top genetic determinant SNPs (P-value < 5×10^{-8}) with linkage disequilibrium (LD) $R^2 < 0.001$ within 10,000bps clumping distance. Proxy SNPs were searched for if SNPs not available in the metabolite GWAS ($R^2 > 0.8$). The palindromic SNPs with minor allele frequency less than 0.3 were excluded. It resulted in 161 independent SNPs for systolic blood pressure ($R^2 = 2.6\%$), 174 SNPs for diastolic blood pressure ($R^2 = 2.8\%$) and 40 SNPs for major depression ($R^2 = 1.1\%$). Inverse variance weighted MR, Maximum likelihood MR, MR Egger analysis and median-based estimator were also performed to check the significant results⁷⁵. We used the Bonferroni P-value threshold by correcting the independent equivalents of the number of tests per disease: P-value < 2.3×10^{-3} for blood pressure, and P-value < 0.025 for depression.

Indicated disease-metabolite associations: effect of indicated disease

We associated the drug-related metabolites with the indicated disease in those who were not receiving the drug under study, i.e. the on-target-treatment-naïve population. This was focused on type 2 diabetes, dyslipidemia, hypertension and depression. The type 2 diabetes analyses were performed based on Rotterdam Study and NEO. Type 2 diabetes was defined as fasting glucose ≥ 7.0 mmol/L, and the cases who used glucose-lowering drugs were excluded in the analysis (815 cases and 10,619 non-diabetics controls in meta-analysis). We performed a regression model with type 2 diabetes status as an independent variable, glucose-lowering-drug-related metabolite as the dependent variable. Covariates included age, sex, BMI, smoking and lipid-modifying drugs.

Dyslipidemia and hypertension were tested in ERF and Rotterdam Study. We tested the association of lipid-modifying-drug-related metabolites and dyslipidemia. Dyslipidemia was defined according to the National Cholesterol Education Program-Adult Treatment Panel III as either total cholesterol ≥ 240 mg/dL, LDL-C ≥ 160 mg/dL, HDL-C < 40 mg/dL, or triglyceride ≥ 200 mg/dL⁷⁶ (2,451 cases and 2,956 controls in meta-analysis). We excluded the subjects with lipid-modifying drugs and adjusted for age, sex, BMI and smoking in the model. The associations of antihypertensives-related metabolites and hypertension were performed. Hypertension was defined as either systolic blood pressure ≥ 140 mmHg or diastolic blood pressure ≥ 90 mmHg (2,506 cases and n = 2,263 controls in meta-analysis). We excluded the subjects with antihypertensives and adjusted for age, sex, BMI, smoking and lipid-modifying drugs in the model.

For depression, we tested the associations between the six SSRIs-related metabolites and depressed mood in the participants without any antidepressant drug (ATC code as N06A)⁷⁷. Depressed mood was measured by either diagnostic interviews or validated depression questionnaires (3,966 cases and 8,887 controls in the meta-analysis). The detailed definition of cases and control in cohorts was described in our previous publication⁷⁷. We adjusted for age, sex, fasting status, lipid-modifying drug and current smoking status.

In addition, we checked the association of fasting glucose and glucose-lowering-drug-related metabolites in the non-diabetes population (n = 5,871) and the association of systolic and diastolic blood pressure and antihypertensives-related metabolites in the non-hypertension population (n = 2,263) in ERF and Rotterdam Study. The non-diabetes population were those fasting glucose ≤ 6.9 mg/dl and without any anti-diabetics treatment; the non-hypertension population were those systolic blood pressure less than 140 mmHg, diastolic blood pressure less

than 90 mmHg and without any antihypertensives. Linear regression was performed with adjustment for age, sex, BMI, smoking and lipid-modifying drugs in the model. The P-value threshold for significance of associations was corrected for the number of independently tested metabolite equivalents per disease or endophenotype. Nominal significance between disease/endophenotype and metabolite was also considered (P-value < 0.05).

A comparison of cross-sectional and longitudinal studies and benchmarking findings by genetics: using statin as an example

Forty-eight metabolites in the current cross-sectional study were also studied in the previous longitudinal study by Wurtz and co-workers which also quantified the ¹H-NMR metabolic profiles in blood samples but focused on the change of metabolite concentrations of two time points: baseline and follow-up¹⁷. As the longitudinal study only adjusted for age and sex, we used the same model in the present cross-sectional study to allow a fair comparison. Since the effects of lipophilic statin and hydrophilic statin are similar in the current study, we used the results of lipophilic statin which had the largest sample size to do the comparison. The results of MR analysis, association of rs12916-T and metabolites, from Wurtz and co-workers were also used in the comparison¹⁷.

We then compared the significant statin-metabolite associations in the current cross-sectional study with the associations of rs12916-T and metabolites. We used the GWAS results of the NMR metabolites from our previous paper which included 24,925 individuals without lipid-modifying drug usage¹¹. It resulted in 55 metabolites in the comparison.

PPIs, circulating metabolite and liver function

We studied biochemical variables in liver function test, i.e. ALT, AST, GGT, AST/ALT, total bilirubin and alkaline phosphatase, and hepatic steatosis. The liver function test used automatic enzymatic procedures (Roche Diagnostics GmbH, Mannheim, Germany)⁷⁸. Abdominal ultrasonography was performed by a certified and experienced technician (Pavel Taimr) on Hitachi HI VISION 900 (Highland Heights, OH). Images were stored digitally and re-evaluated by a single hepatologist with more than ten years of experience in ultrasonography. The diagnosis of steatosis was determined by the ultrasound technician according to the protocol by Hamaguchi et al⁷⁹.

Linear regression was performed in Rotterdam Study (n = 3,436) with liver function measurements as an independent variable and metabolite levels as a dependent variable. The covariates included age, sex, BMI, smoking,

lipid-modifying drugs, PPIs and alcohol intake per day calculated from questionnaires. The P-value threshold was Bonferroni corrected with 10 independent equivalents of 55 PPI-related metabolites and six independent equivalents of the seven liver function measurements ($P\text{-value} < 8.3 \times 10^{-4}$). We further checked the association of PPI use and liver function measurements by linear regression with adjustment for age, sex, BMI, smoking and alcohol intake per day ($P\text{-value} < 8.3 \times 10^{-3}$).

PPIs, circulating metabolites and gut microbiome

We extracted the associations of PPIs with gut microbiota and (alpha) diversity from our previous paper by Imhann and co-workers³⁹. Age, sex, BMI, antibiotics use and sequence read depth were corrected in the association analysis³⁹. In total, 92 bacterial taxa abundance assessed by tag sequencing of the 16S rRNA gene⁵⁸ and Shannon's diversity index (alpha diversity) were reported to be significantly different between PPI users and non-users (211 PPI users and 1,594 non-users, $FDR < 0.05$). Forty-five of the 92 bacterial taxa abundance and alpha diversity were also tested association with metabolites measured by Nightingale Health in our previous study⁸⁰. In brief, it included 2,309 individuals who were not using antibiotics from Rotterdam Study ($n = 1,390$) and LifeLines DEEP ($n = 915$)^{47,58}. Age, sex, BMI, technical covariates (time in mail and storage time) and medication use (lipid-modifying drugs, metformin and PPIs) were adjusted in the association analysis. The P-value threshold for gut microbiota was Bonferroni corrected with 10 independent equivalents of 55 PPI-related metabolites and 15 independent equivalents of the 45 gut microbiota ($P\text{-value} < 3.3 \times 10^{-4}$). P-value threshold for alpha diversity was 5.0×10^{-3} .

Reporting Summary

Further information on life sciences study design is available in the Life Science Reporting Summary linked to this article.

Data Availability

All the summary statistics of the meta-analysis are made available through the supplementary tables. Also, the data to make the figures are available in the supplementary tables. As to the availability of the raw data, the analyses are based on a meta-analysis of multiple Dutch studies. The raw metabolomics data of the studies are pooled in a single database. The quantified metabolic biomarker datasets that participated in this study are available through the Biobanking and Biomolecular Resources Research Infrastructure of the Netherlands (BBMRI-NL) website <http://www.bbmri.nl/omics-metabolomics/>, where details of how to access the data through centralized computational facilities are described. To request data, researchers will have to fill out and sign the data access request and code of conduct forms. Applications compliant with ethical and legal legislations will be reviewed by the BBMRI-NL board for overlap with other ongoing projects before access is granted. Data on medication used in the current study are available through the individual studies upon reasonable request. To obtain these, the principal investigator of the cohorts can be contacted through <http://www.bbmri.nl/omics-metabolomics/>. No custom code or mathematical algorithm is used in the current study.

Methods-only References

54. Ikram, M.A., *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* **32**, 807-850 (2017).
55. Boomsma, D.I., *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* **9**, 849-857 (2006).
56. Penninx, B.W., *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int J Methods Psychiatr Res* **17**, 121-140 (2008).
57. Schoenmaker, M., *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* **14**, 79-84 (2006).
58. Tigchelaar, E.F., *et al.* Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5**, e006772 (2015).
59. van der Heijden, A.A., *et al.* The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with type 2 diabetes treated in primary care in the Netherlands. *BMJ Open* **7**, e015599 (2017).
60. Geleijnse, J.M., *et al.* Effect of low doses of n-3 fatty acids on cardiovascular diseases in 4,837 post-myocardial infarction patients: design and baseline characteristics of the Alpha Omega Trial. *Am Heart J* **159**, 539-546 e532 (2010).
61. Schram, M.T., *et al.* The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. *Eur J Epidemiol* **29**, 439-451 (2014).
62. Sayed-Tabatabaei, F.A., *et al.* Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke* **36**, 2351-2356 (2005).
63. van Oosterhout, W.P., *et al.* Validation of the web-based LUMINA questionnaire for recruiting large cohorts of migraineurs. *Cephalalgia* **31**, 1359-1367 (2011).
64. de Mutsert, R., *et al.* The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. *Eur J Epidemiol* **28**, 513-523 (2013).
65. Soininen, P., Kangas, A.J., Wurtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* **8**, 192-206 (2015).
66. Inouye, M., *et al.* Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol* **6**, 441 (2010).
67. van den Akker, E., *et al.* Predicting biological age based on the BBMRI-NL 1H-NMR metabolomics repository. *bioRxiv*, 632919 (2019).
68. Sturm, R. The effects of obesity, smoking, and drinking on medical problems and costs. *Health Aff (Millwood)* **21**, 245-253 (2002).
69. Van Gaal, L.F., Mertens, I.L. & De Block, C.E. Mechanisms linking obesity with cardiovascular disease. *Nature* **444**, 875-880 (2006).
70. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221-227 (2005).
71. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* **36**, 1-48 (2010).
72. Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* **27**, 1133-1163 (2008).

73. lab, N. Details and considerations of the UK Biobank GWAS, available 10 Dec, 2018.
<http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas> (2017).
74. Wray, N.R., *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* **50**, 668-681 (2018).
75. Hemani, G., *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**(2018).
76. National Cholesterol Education Program Expert Panel on Detection, E. & Treatment of High Blood Cholesterol in, A. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* **106**, 3143-3421 (2002).
77. Bot, M., *et al.* Metabolomics profile in depression: a pooled analysis of 230 metabolic markers in 5,283 cases with depression and 10,145 controls. *Biological Psychiatry* (2019).
78. Koehler, E.M., *et al.* Presence of diabetes mellitus and steatosis is associated with liver stiffness in a general population: The Rotterdam study. *Hepatology* **63**, 138-147 (2016).
79. Hamaguchi, M., *et al.* The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. *Am J Gastroenterol* **102**, 2708-2715 (2007).
80. Dina Vojinovic, *et al.* Relationship between gut microbiota and circulating metabolites in population-based cohorts. *Nature communications* **In press**(2019).