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

Liu, J.; Lahousse, L.; Nivard, M.G.; Bot, M.; Chen, L.M.; Klinken, J.B. van; ... ; Duijn, C.M. van

https://doi.org/10.1038/s41591-019-0722-x

nature medicine

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 Bot[4,5], Lianmin Chen[6,7], Jan Bert van Klinken[8,9,10], Carisha S. Thesing[4,5], Marian Beekman[11], Erik Ben van den Akker[11,12,13], Roderick C. Slieker[5,14,15], Eveline Waterham[16], Carla J. H. van der Kallen[17,18], Irene de Boer[19], Ruifang Li-Gao[20], Dina Vojinovic[1], Najaf Amin[1], Djawad Radjabzadeh[21], Robert Kraaij[21], Louise J. M. Alferink[22], Sarwa Darwish Murad[22], André G. Uitterlinden[1,21], Gonneke Willemsen[4,5], Rene Pool[4,5], Yuri Milaneschi[4,5], Diana van Heemst[23], H. Eka D. Suchiman[11], Femke Rutters[5,14], Petra J. M. Elders[5,24], Joline W. J. Beulens[5,14], Amber A. W. A. van der Heijden[5,24], Marleen M. J. van Greevenbroek[17,18], Ilja C. W. Arts[18,25,26], Gerrit L. J. Onderwater[19], Arn M. J. M. van den Maagdenberg[8,19], Dennis O. Mook-Kanamori[20,27], Thomas Hankemeier[28,29], Gisela M. Terwindt[19], Coen D. A. Stehouwer[17,18], Johanna M. Geleijnse[16], Leen M. 't Hart[5,11,14,15], P. Eline Slagboom[11], Ko Willems van Dijk[8,9,30], Alexandra Zhernakova[6], Jingyuan Fu[6,7], Brenda W. J. H. Penninx[4,5], Dorret I. Boomsma[4,5], Ayşe Demirkan[1,6,31], Bruno H. C. Stricker[1,21,32] and Cornelia M. van Duijn[1,2,28*]

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 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. The atlas includes a meta-analysis of ten cohorts (18,873 persons) and uncovers 1,071 drug–metabolite associations after evaluation of confounders including co-treatment. We show that 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 rapidly[1], facilitating large-scale studies that have highlighted the importance of differential molecular dynamics captured in a wide range of common complex diseases, including diabetes, cardiovascular disease, asthma and dementia[2–9]. 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—for example, lifestyle, nutrition, environment and microbiome[14–16]. Although the use of drugs is recognized as having a major effect on metabolism, our knowledge of drug–metabolite associations is incomplete and is limited to the most commonly prescribed drugs—for example, statins, metformin and antihypertensives[17–22]. In addition, even for commonly prescribed drugs, their 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 open new avenues to improve drug efficacy, enable repurposing of drugs[23–25] and improve our understanding of the off-target effects of drugs occurring in the individual patient[26,27]. However, pointing out such associations is complicated since confounding may occur due to the metabolic changes that are either 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-treatment[28]. Last, not but least, longitudinal observations are often lacking for relatively rare off-target effects, forcing the basing of clinical decision making 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 the proton nuclear magnetic resonance (^{1}H-NMR) platform of Nightingale Health (Supplementary Table 2). This 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 associations[29] including diabetes[30], dementia[6], cardiovascular diseases[31,32], migraine[33], Graves' disease[34] and mortality[35,36]. Nightingale Health is

A full list of affiliations appears at the end of the paper.

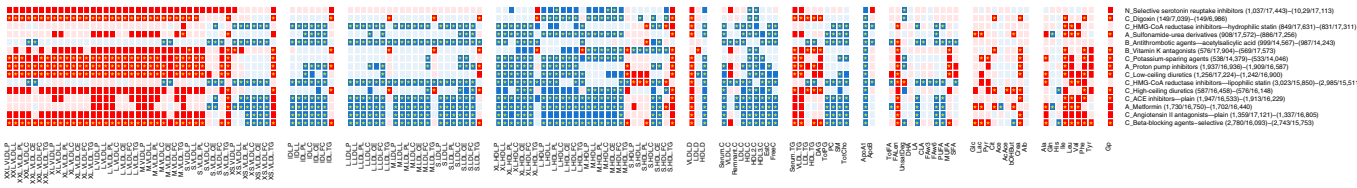


Fig. 1 | Drug–metabolite associations in model 1 versus model 2. The top 15 drugs that were associated with the largest number of metabolites in model 1 (age and sex adjusted) of linear regression are ordered and shown in the figure. The first letter of the ATC code precedes the drug name, to identify different categories. Sample sizes of drug users and non-users in model 1 and in model 2 (age, sex, BMI and smoking adjusted) are shown following drug names, respectively. Dark red, positive significant associations in model 1 ($P < 1.9 \times 10^{-5}$); light red, positive nonsignificant associations in model 1 ($P \geq 1.9 \times 10^{-5}$); dark blue, negatively significant associations in model 1 ($P < 1.9 \times 10^{-5}$); light blue, negatively nonsignificant associations in model 1 ($P \geq 1.9 \times 10^{-5}$). Asterisks in boxes denote that neither direction nor significance status were different between model 1 and model 2 ($P < 1.9 \times 10^{-5}$). Two-tailed tests were used.

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 both research and clinical care. The term 'metabolite' used 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 drug–metabolite associations and exploring in depth the interaction of metabolites with two drugs, statins and proton pump inhibitors (PPIs).

Results

Overall drug–metabolite associations. We meta-analyzed 12 datasets of ten Dutch cohorts ($n = 18{,}873$ individuals; Supplementary Table 3) 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 $= 1.9 \times 10^{-5}$). The number of drug users ranged from 3,023 (16.0%, for lipophilic statins) 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. Among the metabolites studied, effect estimates derived from different datasets agreed convincingly ($P = 1.67 \times 10^{-11}$ to 1.0×10^{-318} for pairwise correlation tests; Supplementary Fig. 1 and Supplementary Table 5). Figure 1 shows the associations of model 1 for the top 15 drugs associated with the largest number of metabolites. These 15 drugs can be divided into five groups: (1) six antihypertensives (selective beta-blockers, angiotensin II antagonists, ACE inhibitors, high- and low-ceiling diuretics and potassium-sparing agents); (2) two glucose-lowering drugs (metformin and sulfonamide-urea derivatives); (3) two lipid-modifying drugs (lipophilic and hydrophilic statins); (4) three other cardiovascular-related drugs (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 cardiometabolic-related drugs which can, in large part, be explained by the fact that the numbers of users were large and the current metabolome spectrum contains mainly lipids correlated with each other (Extended Data Fig. 1).

Effects of body mass index, smoking and co-treatment as major confounders. Next, we studied the potential confounding effect of body mass index (BMI) and smoking. In total, 1,640 of the 2,087 significant associations (78.6%) in model 1 remained significant after additional adjustment for BMI and smoking in model 2 (Extended Data Fig. 2). The drug group for which the evidence for association was most dramatically impacted by adjustment for BMI and smoking was SSRIs: 59 of the initial 65 SSRI–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 had additionally excluded the confounding of other drugs by additional adjustment for co-treatments in model 3 (Extended Data Fig. 3), 1,071 significant associations remained to be investigated. For five out of six antihypertensives in the top 15 drugs (Fig. 2), associations with low-density lipoprotein (LDL) and intermediate-density lipoprotein (IDL) particles were explained by co-treatments. Notably, statin use was correlated with antihypertensives and associated with LDL and IDL particles, which led to a false discovery association of LDL and IDL particles and antihypertensives. Most antihypertensive associations disappeared after adjusting for co-treatment, including statins, except for 15.4% (4/26) of selective beta-blockers and 100% of 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 (Extended Data Fig. 4) and both drugs were associated with similar circulating metabolites—that is, there were 85 metabolites associated with metformin in model 2, 59 of which 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 metabolites (Fig. 2). These results were confirmed by our sensitivity analysis from subsamples of patients who were administered one drug only: all significant associations in the sensitivity analyses remained significant in model 3, the model with co-treatment adjusted (Extended Data Fig. 5).

Examples of applications of the atlas. *Effect of indicated disease: drug–metabolite associations explained by indication.* First, we tested whether indicated diseases causally related to drug-related metabolites using the genetic risk score of the disease as an instrumental variable in Mendelian randomization (MR) (Supplementary Tables 6 and 7). Second, we associated drug-related metabolites with the indicated disease in individuals who were not receiving treatment—that is, the on-target-treatment-naive population (Supplementary Table 4). For instance, in the current study, metformin use was found to be associated with increasing alanine but we also know that type 2 diabetes (causally by MR) increases alanine levels in

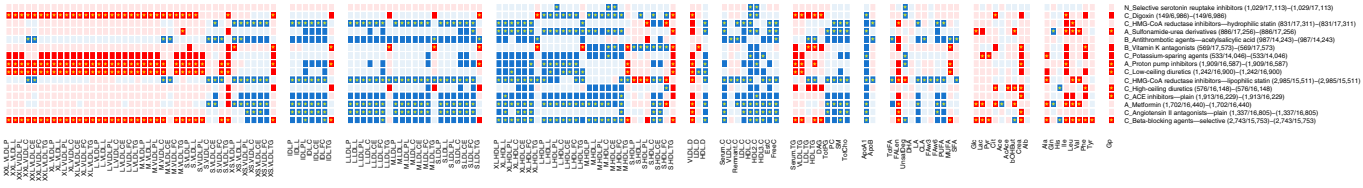


Fig. 2 | Drug–metabolite associations in model 2 versus model 3. The top 15 drugs associated with the largest number of metabolites in the baseline model (model 1) are ordered and shown. The first letter of the ATC code is shown preceding the drug name, to identify different categories. Sample sizes of drug users and non-users in regression model 2 (age, sex, BMI and smoking adjusted) and model 3 (age, sex, BMI, smoking and co-treatment adjusted) are shown following drug names, respectively. Dark red, positive significant associations in model 2 ($P < 1.9 \times 10^{-5}$); light red, positive nonsignificant associations in model 2 ($P \geq 1.9 \times 10^{-5}$); dark blue, negatively significant associations in model 2 ($P < 1.9 \times 10^{-5}$); light blue, negatively nonsignificant associations in model 2 ($P \geq 1.9 \times 10^{-5}$). Asterisks in boxes denote that neither direction nor significance status was different between models 2 and 3 (*P* value threshold is multiple testing-corrected per drug; see Supplementary Table 4). Two-tailed tests were used.

the blood[4]. 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 exclusion of all metformin users, type 2 diabetes was still associated with increasing alanine levels ($\beta = 0.42$, $P = 8.3 \times 10^{-19}$). Integration of the findings on drug–metabolite and disease–metabolite associations suggests that alanine levels in blood are most probably raised by the effect of type 2 diabetes rather than by that of metformin.

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 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 associations with metformin and 100% with sulfonamide-urea derivatives, based on a nominal significance level in disease–metabolite associations ($P < 0.05$, Fig. 3). With such a strict exclusion of effect of the indicated disease, we still found that acetate was negatively associated with metformin effect, and there was no evidence that the relationship had 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 to those of a longitudinal study published by Wurtz and co-workers[17]. In their paper, changes in metabolite concentrations in blood (two time points per individual) were compared between 716 patients who began statin therapy during follow-up and 4,874 persistent non-users[17]. In total, 48 metabolites from that study[17] overlapped with ours, in which metabolite and statin use were assessed simultaneously in 3,023 individuals prescribed lipophilic statins and 15,850 non-users, providing a cross-sectional snapshot. Twenty-nine (60%) metabolites showed consistently significant results between the two studies (Fig. 4a). We further checked 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 because the protective T allele results in low-functioning HMG-CoA reductase, which is one of the pharmacologically targeted effects of statins[17,37]. Figure 4a shows that 20 of 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 cholesterols and lipoprotein particles (50.0%). Meanwhile, 15 of the 19 metabolites (80%) that were inconsistently associated with statins between our study and the previous study[17] 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 lipophilic statins (Fig. 4b and Supplementary Table 8). Twenty-five of these are novel and complement the findings of the above-mentioned study by Wurtz and co-workers[17]. The novel metabolites emerging, by association with rs12916-T in our cross-sectional analyses, involved very small to medium VLDL particles, IDL particles, LDL particles and total phosphatidylcholine and other cholines.

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 for co-treatment (Fig. 5a), involving small to extremely large VLDL, large HDL and triglyceride particles, mono-unsaturated fatty acids, isoleucine, creatinine and glycoprotein acetyls (mainly a1-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 defined daily dose in PPI users and metabolites (Extended Data Fig. 6).

PPIs are often administered to patients with cirrhosis, and in these patients they are associated with infections and worsening prognosis[38]. We next investigated, via Rotterdam Study ($n = 3,436$), whether PPI-associated metabolites are also associated with liver function, including biochemical variables of liver function tests and hepatic steatosis. Figure 5a,b shows high consistency for 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, for hepatic steatosis is 98.2% (54/55), for gamma-glutamyl transferase (GGT) 80.0% (44/55) and for alanine transaminase (ALT) 81.8% (45/55; positively associated); 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 (Fig. 5b and Supplementary Table 10).

We then studied PPI-associated metabolites in relation to microbial diversity and the abundance of microbiota that are pharmacologically driven by PPI use in a population[39–43]. We found that 94.5% (52/55) of the metabolites associated with PPIs are also associated with gut microbial (alpha-) diversity, in a meta-analyses of 2,305 participants that did not use antibiotics (Fig. 5c and Supplementary Table 11). Of the 92 gut microbiota with which

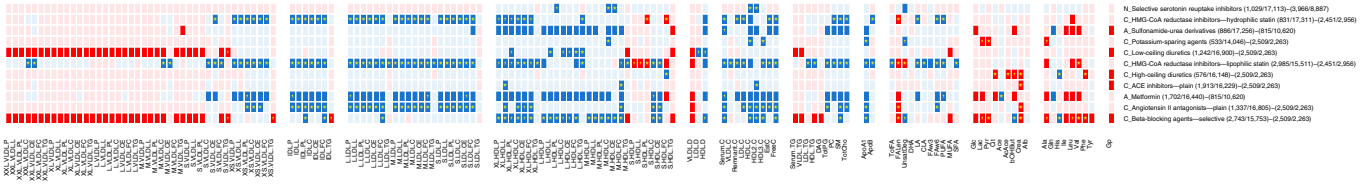


Fig. 3 | Drug–metabolite associations in model 3 versus significance after disentangling the indicated disease/endophenotype effect. The top 15 drugs associated with the largest number of metabolites in the baseline model (model 1) are ordered and shown. The first letter of the ATC code is shown preceding the drug name, to identify the different categories. Sample sizes of drug users and non-users in regression model 3 (age, sex, BMI, smoking and co-treatment adjusted) and those of cases and controls in disease–metabolite associations are shown following drug names, respectively. Dark red, positive significant associations in model 3; light red, positive nonsignificant associations in model 3; dark blue, negatively significant associations in model 3; light blue, negatively nonsignificant associations in model 3. Asterisks in boxes denote significant associations confirmed after disentangling the disease/endophenotype effect ($P < 0.05$ is the threshold for the significance in disease–metabolite associations). Quote marks in boxes denote associations confirmed after disentangling the disease/endophenotype effect ($P <$ the threshold after multiple testing-corrected per disease is the threshold for the significance in disease-metabolite associations; see Supplementary Table 4). Two-tailed tests were used.

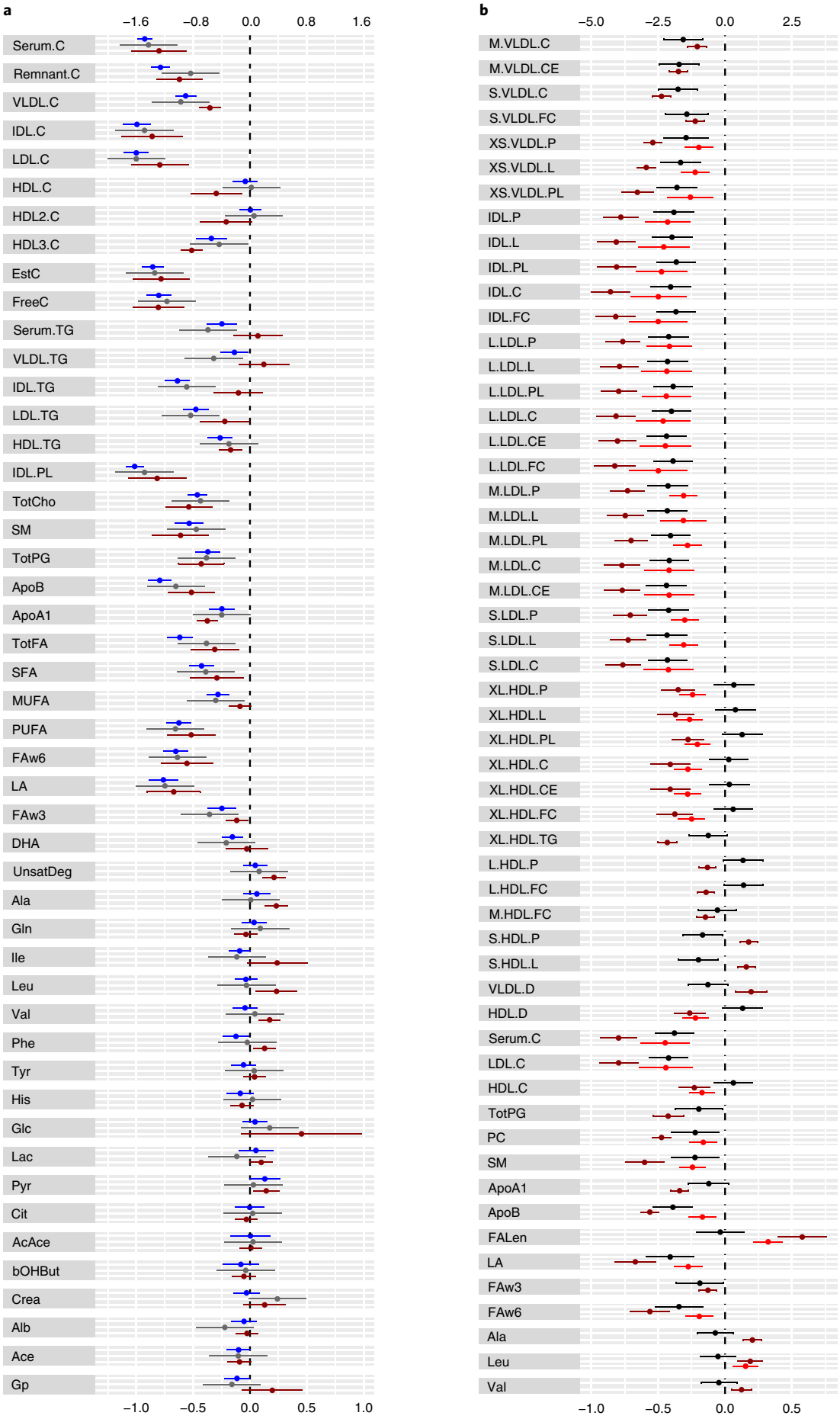
abundances were associated with PPI use[39], 45 were available for testing of the association with metabolites (Supplementary Table 12). We found that three common microbiota (phylum Tenericutes, class Mollicutes and family Ruminococcaceae) showing reduced abundance in PPI users had a metabolite association pattern consistent with that of PPI–metabolite association, but in the opposite direction (Fig. 5d and Supplementary Table 13). The genus *Scardovia* showed an increased abundance in the gut of patients using PPIs. Although this genus showed a metabolite association pattern similar to that of PPIs, only the association with glycoprotein reached statistical significance after adjusting for multiple testing.

Discussion

We performed a comprehensive analysis of the interaction between 87 commonly prescribed drugs and as many as 150 circulating metabolites measured by ^{1}H-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. We also demonstrated three examples of applications of the atlas, disentangling disease (for example, type 2 diabetes) and therapy (for example, metformin) effects, aligning longitudinal and genetic analysis with our large-scale cross-sectional findings and, ultimately, linking PPI–metabolite interactions to gut microbiome abundance and liver function.

Although many of the metabolites cluster strongly in populations (Extended Data Fig. 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, and this is especially true for VLDL and HDL particles. This is consistent with previous studies on the role of lipid particle profiles and diseases[4,6,31–34,44,45], and is also true for amino acids. In Rotterdam Study, histon clusters strongly with leucine, valine and isoleucine (in correlation tests, $P = 3.3 \times 10^{-23}$). However, histone is negatively associated with selective beta-blocker use (Fig. 2) while leucine, valine and isoleucine are positively associated. We showed that BMI is a major confounder of associations with SSRIs. The high proportion of elimination in SSRI–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 neither dieting nor weight gain is part of the diagnostic criteria for depression[46]. After adjustment for co-treatment, the similar significant association patterns between different drugs (for example, angiotensin II antagonists and metformin) may imply that drug–metabolite associations are independently associated with a similar shift in metabolism, but this is true only 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 (for example, statins, antihypertensives and metformin). This type of confounding was further addressed by investigating whether drug–metabolite associations are related to the pathology (for example diabetes, hypertension, dyslipidemia) that indicated prescription. As a typical metabolic disorder, evidence shows that type 2 diabetes explains a substantial association between glucose lowering and drug–metabolite associations. 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 a testable hypothesis for future trials. We further showed that, to some extent, statin–metabolite associations in a large-scale, cross-sectional study can mimic those of the longitudinal effect of statin administration, which

Fig. 4 | Comparison of statin–metabolite associations between cross-sectional, longitudinal and genetic studies. a, Comparison of statin–metabolite associations among the current cross-sectional study, a longitudinal study by Wurtz and co-workers[17] and a genetic study. Results in the longitudinal study ($n = 716/4,874$) are shown as Wurtz and co-workers' study in both s.d.-scaled metabolite concentration units (upper *x* axis) and relative to the lowering effect on LDL cholesterol (lower *x* axis). The results of metabolite–rs12916-T associations ($n = 27,914$) are shown as effect estimate per s.d. and relative to the lowering effect on LDL cholesterol (lower *x* axis). **b**, 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 units, lower *x* axis), and the results of metabolite–rs12916-T associations are represented as fivefold the effect estimate (standardized metabolite concentration units, upper *x* axis). Error bars, 95% confidence intervals, were statistically corrected for multiple testing; this means that, if the error bar crosses the zero line, the association is not significant at the multiple testing significance level. Statistical data were extracted from the previous longitudinal study by Wurtz and co-workers[17]. Two-tailed tests were used.


a
−1.6
−0.8
0.0
0.8
1.6
Serum.C
Remnant.C
VLDL.C
IDL.C
LDL.C
HDL.C
HDL2.C
HDL3.C
EstC
FreeC
Serum.TG
VLDL.TG
IDL.TG
LDL.TG
HDL.TG
IDL.PL
TotCho
SM
TotPG
ApoB
ApoA1
TotFA
SFA
MUFA
PUFA
FAw6
LA
FAw3
DHA
UnsatDeg
Ala
Gln
Ile
Leu
Val
Phe
Tyr
His
Glc
Lac
Pyr
Cit
AcAce
bOHBut
Crea
Alb
Ace
Gp
−1.0
−0.5
0.0
0.5
1.0
b
−5.0
−2.5
0.0
2.5
M.VLDL.C
M.VLDL.CE
S.VLDL.C
S.VLDL.FC
XS.VLDL.P
XS.VLDL.L
XS.VLDL.PL
IDL.P
IDL.L
IDL.PL
IDL.C
IDL.FC
L.LDL.P
L.LDL.L
L.LDL.PL
L.LDL.C
L.LDL.CE
L.LDL.FC
M.LDL.P
M.LDL.L
M.LDL.PL
M.LDL.C
M.LDL.CE
S.LDL.P
S.LDL.L
S.LDL.C
XL.HDL.P
XL.HDL.L
XL.HDL.PL
XL.HDL.C
XL.HDL.CE
XL.HDL.FC
XL.HDL.TG
L.HDL.P
L.HDL.FC
M.HDL.FC
S.HDL.P
S.HDL.L
VLDL.D
HDL.D
Serum.C
LDL.C
HDL.C
TotPG
PC
SM
ApoA1
ApoB
FALen
LA
FAw3
FAw6
Ala
Leu
Val
−1.0
−0.5
0.0
0.5

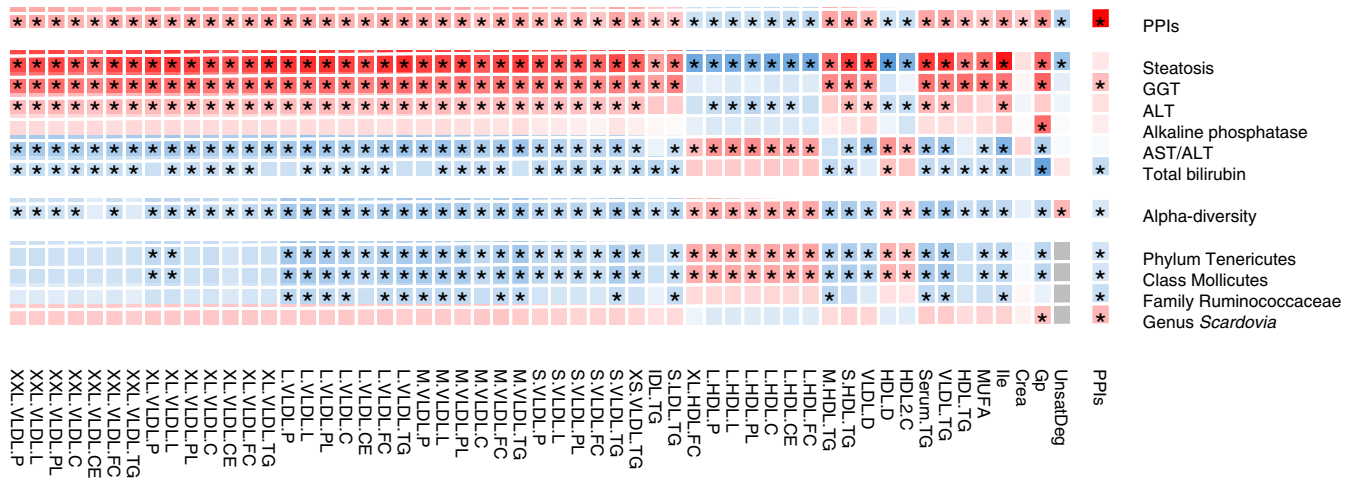


Fig. 5 | Integrated data of PPIs, metabolites, liver function measurements and gut microbiome. Significant results after integrating the directions of association among PPIs, metabolites, liver function measurements and gut microbiome by linear regression. Red, positive association; blue, negative association; the depth of color represents the value of effect estimate per s.e.m: range from 0 (white) to 15 (red) or -15 (blue). Gray, associations were not performed. Asterisks denote significance of associations. Two-tailed tests were used.

are preferred from a methodological perspective. This premise is strengthened by the fact that both studies are benchmarked by MR. These findings suggest that the atlas yields informative associations that could 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 in composition of the gut microbiome has been long recognized[39,41,47]. Interestingly, one recent study[48] 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 highly consistent with our findings. Focusing on oral bacteria, levels of the genus *Scardovia* were found to be increased in the gut microbiome of PPI users[39]. This raises the hypothesis that, due to PPI-related changes in gastric acid secretion, these microbiotas are reaching the gastrointestinal tract, very similar to the mechanism described in mice[40] and in a study of the human gut microbiome in patients with liver cirrhosis[49]. *Scardovia* was most strongly and significantly associated with acute-phase glycoprotein, which is an intriguing metabolite from a clinical and epidemiological perspective as it is synthesized in the liver[50] and associated with a wide spectrum of incident diseases[51], including cardiovascular disease[52], type 2 diabetes[53], cognition[6] and all-cause mortality[36]. A key question in future studies is to what extent glycoprotein plays a mediating role in the relation between gut microbiome and morbidity. Our analysis validated previous findings—that the human gut microbiome is altered in patients with liver cirrhosis[49] and that withdrawal of PPIs in these patients decreases oral-origin taxa[38]—in a general-population study with a very low prevalence (<3%) of severe diseases such as advanced liver or kidney disease. 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 required to examine this hypothesis.

Another point of interest is that the experimental study on the effect of PPIs on the gut microbiome in patients with cirrhosis was based on omeprazole[38]. When we compared the various drugs included in the PPI category, we found that omeprazole is indeed associated with those metabolites identified in the drug category analysis (Extended Data Fig. 7). However, we also found that other drugs such as lansoprazole are even more strongly and significantly associated, while the association with rabeprazole and esomeprazole is less strong and nonsignificant. These are interesting findings that should be followed up.

This comprehensive drug–metabolite atlas provides a basis for future exploration of drug–metabolite interactions, using either our omics-based approach or other, future, (un)targeted experimental and longitudinal pharmaceutical research. Our study includes examples of how to use the atlas that can be extended to other settings. We have limited the atlas to the most common drugs, but it could be expanded in the future to include rare drugs following the generation of such data for this platform in larger cohorts, such as UK Biobank. These 'mega-cohorts' would also facilitate systematic study of the interaction of multiple drug intake with sufficient statistical power. On the other hand, the current atlas could be a starting point for future research focusing on a limited number of drugs with metabolomics, to check drug interactions. Another future challenge is to extend the atlas to a wider range of metabolites measured using other platforms (for example, mass spectrometry) and tissues (for example, urine). The use of MR is a strength of the current study, because it enables us to disentangle the effect of drugs and indicated diseases. However, we are not always able to capture strong instruments for the MR test, which may reduce the power of our analyses when aiming to exclude disease effects. Since our knowledge of the gene-mimicking effects of drugs and diseases is rapidly expanding, we are optimistic that more powerful genetic instrumental variables will be identified in the near future, opening windows of opportunity into MR analyses in both pharmacometabolomic research and clinical trials.

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

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-019-0722-x.

Received: 22 May 2019; Accepted: 27 November 2019;
Published online: 13 January 2020

References

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**, 11 (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* https://doi.org/10.1093/database/baw083 (2016).
24. Voora, D. & Shah, S. H. Pharmacometabolomics meets genetics: a 'natural' clinical trial of statin effects. *J. Am. Coll. Cardiol.* **67**, 1211–1213 (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 Transl. Sci.* **1**, 170–179 (2016).
27. US Food and Drug Administration. *22 Case studies where phase 2 and phase 3 trials had divergent results* https://www.fda.gov/about-fda/reports/22-case-studies-where-phase-2-and-phase-3-trials-had-divergent-results (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* **62**, 2298–2309 (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. *Diabetologia* **62**, 2298–2309 (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. *J. Biol. Chemi.* **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).

[1]Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands. [2]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [3]Department of Bioanalysis, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium. [4]Department of Biological Psychology, Amsterdam University Medical Center, Vrije Universiteit, Amsterdam, the Netherlands. [5]Amsterdam Public Health Research Institute, Amsterdam, the Netherlands. [6]Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands. [7]Department of Pediatrics, University Medical Center Groningen, Groningen, the Netherlands. [8]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. [9]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands. [10]Department of Clinical Chemistry, Laboratory Genetic Metabolic Disease, Amsterdam University Medical Center, Amsterdam, the Netherlands. [11]Department of Biomedical Data Sciences, section of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [12]Department of Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, the Netherlands. [13]Leiden Computational Biology Center, Leiden University Medical Center, Leiden, the Netherlands. [14]Department of Epidemiology and Biostatistics, Amsterdam University Medical Center, Vrije Universiteit, Amsterdam, the Netherlands. [15]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands. [16]Division of Human Nutrition and Health, Wageningen University, Wageningen, the Netherlands. [17]Department of Internal Medicine, Maastricht University, Maastricht, the Netherlands. [18]School for Cardiovascular Diseases, Maastricht University, Maastricht, the Netherlands. [19]Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands. [20]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands. [21]Department of Internal Medicine, Erasmus MC, University Medical Center, Rotterdam, the Netherlands. [22]Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Center, Rotterdam, the Netherlands. [23]Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands. [24]Department of General Practice and Elderly Care Medicine, Amsterdam University Medical Center, Vrije Universiteit, Amsterdam, the Netherlands. [25]Department of Epidemiology, Maastricht University, Maastricht, the Netherlands. [26]Maastricht Center for Systems Biology, Maastricht University, Maastricht, the Netherlands. [27]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands. [28]Leiden Academic Center for Drug Research, Leiden University, Leiden, the Netherlands. [29]Netherlands Metabolomics Center, Leiden, the Netherlands. [30]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands. [31]Section of Statistical Multi-omics, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK. [32]Inspectorate of Healthcare, The Hague, the Netherlands. *e-mail: jun.liu@ndph.ox.ac.uk; cornelia.vanduijn@ndph.ox.ac.uk

Methods

Study population. The research was performed within BBMRI-NL. The study included 18,873 individuals from 12 datasets of 10 Dutch cohorts who had metabolites measured by Nightingale Health, drug information based on the Anatomical Therapeutic Chemical (ATC) Classification and clinical phenotypes that allowed us to control for confounders. These cohorts included Rotterdam Study (RS), 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 examples from the application atlas, we additionally involved the Netherlands Epidemiology of Obesity Study (NEO, $n=6,603$)[64], which is an obese cohort but adjusted for BMI in type 2 diabetes–metabolite associations by inverse probability weighting on BMI to make the results comparable to the Dutch general population. Cohort descriptions, specific data processing and ethical compliance can be found in Supplementary Table 3. All studies were approved by the respective institutional review boards' local research ethics committees, and all participants provided written informed consent to the original study.

Metabolite measurements. The present study included 150 absolute-value-based metabolites measured by high-throughput ^{1}H-NMR metabolomics (Brainshake/Nightingale Health). Details on metabolites are given in Supplementary Table 2. These include quantitative molecular data on 14 lipoprotein subclasses, apolipoproteins A-I and B, multiple cholesterol and triglyceride measures, albumin, various fatty acids and on numerous low-molecular-weight metabolites, including amino acids, glycolysis-related measures and ketone bodies. The 14 lipoprotein subclasses include IDL, six VLDL subclasses, three LDL subclasses and four HDL subclasses based on particle diameter. 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 variation for metabolites ranged between 0.3 and 19.5% (mean 4.5%), and most values are comparable to clinical chemistry assays[11,65].

The blood samples obtained from the cohorts were collated at Leiden University Medical Center and were shipped to and analyzed by Nightingale Health as part of a national initiative. A standardized protocol of metabolite measurement was applied to all 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. Metabolite values that were suggested as being 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[67]. Quality control was unified and included in-depth evaluation of the consistency of findings across datasets, a metabolite correlation matric and principal component analysis (PCA) of cohorts of varying population structure. Pearson's correlation test was used to check pairwise correlation of the overall estimated 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 biochemical pathways (Extended Data Fig. 1 and Supplementary Table 14). The effect of population structure on metabolite clustering was checked by PCA using joint data from four cohorts markedly differing in population: (1) one population-based study, Rotterdam Study[54], (2) one family-based, ERF[62], (3) one disease-based, TMS[61]—which includes only patients with type 2 diabetes in the current dataset—and (4) a case-control, Alpha Omega Cohort[60], including patients with cardiovascular disease and non-disease controls (details are given in Supplementary Table 3). The obvious difference between Alpha Omega Cohort and TMS underscores the fact that meta-analysis should be performed, rather than joint analysis, with pooled data (Extended Data Fig. 8): the fixed-effect meta-analysis assumes a similar effect and structure among cohorts, while the random-effect meta-analysis allows for high heterogeneity.

Because some distributions of metabolites were skewed, we transformed the metabolite values in each cohort to normal distribution. We first added a value of 1 to all metabolites before performing natural logarithm transformation, to include samples labeled zero with metabolite levels below the detectable value; we then scaled these transformed values to s.d. units.

Drug categories. Drug information was classified by ATC codes in each cohort. In brief, drug information by cohort was obtained from either pharmacy records or questionnaires during the interview. Details on drug administration for each cohort can be found in Supplementary Table 3. We used drug category rather than the individual compound in all analyses. We merged drugs with similar chemical, pharmacodynamics, pharmacokinetics and/or therapeutic characteristics into one category. Regarding ATC codes used for combinations of active ingredients, we categorized these separately where possible. We excluded categories with five or fewer users in each cohort, or <20 users in total, from all cohorts. We thus ended up with 87 drug categories (Supplementary Table 1). Drug categorization was confirmed by two experienced pharmacologists, L. L. and B.H.C.S. Throughout the following text, the term drug category is referred to simply as drug. Individuals for whom both metabolite and drug information was available were included in the analysis.

Statistical analysis. All statistical analyses were performed using R statistical software, using two-tailed testing.

Drug–metabolite associations. 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 individual cohorts. Specific family relationships were 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 in pathology[68], and BMI ($kg\,m^{-2}$), which is a major determinant of circulating metabolites that captures the effects of diet and physical activity[69] (model 2). Meta-analysis was performed with either the inverse variance-weighted fixed-effect model (no heterogeneity between cohorts) or the maximum-likelihood random-effect model (significant heterogeneity between cohorts). The degree of heterogeneity was based on Cochran's Q-test. The P value threshold for both Cochran's Q-test and the meta-analysis was Bonferroni corrected with 30 independent equivalents of the 150 metabolites and 87 drugs tested ($P<1.9\times10^{-5}$). Matrix spectral decomposition was used to calculate the number of independent equivalents[70] in the largest population-based dataset, RS Dataset 1. R-package metafor was used for the meta-analysis[71].

Effects of co-treatment—drugs prescribed simultaneously. We next checked the potential confounding of drugs prescribed simultaneously (model 3) in each significant drug–metabolite pair. A co-treatment matrix with Spearman's correlation was made separately in the two population-based cohorts (Rotterdam Study and LifeLines DEEP, $n=6,631$) 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 simultaneously; Extended Data Fig. 4 and Supplementary Table 15) and (2) this drug and the target drug were associated with the target metabolite in the same direction. We used Bonferroni P value correction with the 85 drugs available in the co-treatment matrix ($P<5.9\times10^{-4}$). We then performed the same regression analysis as above in each dataset (12 datasets) and meta-analyzed, with age, sex, BMI, smoking and all available confounding co-treatments as covariates in each significant drug–metabolite pair (model 3). Sensitivity analysis was performed in the subsamples of patients administered one drug only (one-drug-users) and all-treatment-naive 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.

Checking the effect of indicated disease on metabolites with MR. We further focused on drugs in the top 15 drug lists that had the largest number of related metabolites and 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 that uses the effect of genetic variants in determining an exposure, and tests its association with the outcome under study based on the assumption that the genetic variant is inherited independently of the confounding variables[72]. 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 more than five genetic single-nucleotide polymorphisms (SNPs) and explaining >1% of variance in exposure were taken forward. For type 2 diabetes we analyzed the results from our previous well-organized MR research[4], from which 16 metabolites were found to be associated with either metformin or sulfonamide-urea derivatives. In brief, this research was a two-sample, bi-directional 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 SNPs in the instrumental variables[4]. We concurrently used the results of backward MR to check the association of the genetic score of type 2 diabetes and metabolites.

For hypertension and depression, we performed two-sample MR based on previous Genome-wide Association Study (GWAS) results for blood pressure[73] ($n=317,754$), major depression[74] ($n=135,458$ cases and $n=344,901$ controls) and NMR metabolite GWAS[11] ($n=24,925$). Among the 123 metabolites associated with antihypertensives, 96 were available on which to perform MR for systolic and diastolic blood pressure. We also performed MR on major depression with six metabolites associated with SSRIs. We did not perform MR for dyslipidemia over statin-associated metabolites, because most of the latter are lipoproteins that are included in the definition of dyslipidemia.

The R package TwoSampleMR was used for two-sample MR tests[75]. Genetic loci of major depression were extracted from a previous paper[74] because the original

GWAS was not available. The default pipeline in the TwoSampleMR package[75] was used. In brief, the genetic score was based on the top genetic determinant SNPs ($P < 5 \times 10^{-8}$) with linkage disequilibrium $R^2 < 0.001$ within 10,000-base pair clumping distance. Proxy SNPs were searched for if SNPs were not available in the metabolite GWAS ($R^2 > 0.8$). Palindromic SNPs with minor allele frequency <0.3 were excluded, resulting in 161 independent SNPs for systolic blood pressure ($R^2 = 2.6\%$), 174 for diastolic blood pressure ($R^2 = 2.8\%$) and 40 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 results[75]. We used the Bonferroni *P* value threshold by correcting the independent equivalents of the number of tests per disease: $P < 2.3 \times 10^{-3}$ for blood pressure and $P < 0.025$ for depression.

Effect of indicated disease on indicated disease–metabolite associations. We associated drug-related metabolites with the indicated disease in patients not receiving the drug under study—that is, the on-target-treatment-naive population. This was focused on type 2 diabetes, dyslipidemia, hypertension and depression. Type 2 diabetes analyses were performed based on Rotterdam Study and NEO. Type 2 diabetes was defined as fasting glucose ≥7.0 mmol l^{-1}, excluding cases of patients using glucose-lowering drugs from the analysis ($n = 815$ cases and $n = 10,619$ non-diabetic controls in the meta-analysis). We performed a regression model with type 2 diabetes status as an independent variable and 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 87 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^{-1}, LDL-C ≥160 mg dl^{-1}, HDL-C <40 mg dl^{-1} or triglyceride ≥200 mg dl^{-1}[76] ($n = 2,451$ cases and $n = 2,956$ controls in the meta-analysis). We excluded subjects prescribed lipid-modifying drugs and adjusted for age, sex, BMI and smoking in the model. The associations of 123 antihypertensive-related metabolites and hypertension were analyzed. Hypertension was defined as either systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90 mmHg ($n = 2,506$ cases and $n = 2,263$ controls in the meta-analysis). We excluded subjects prescribed antihypertensives and adjusted for age, sex, BMI, smoking and lipid-modifying drugs in the model.

In regard to depression, we tested associations between the six SSRI-related metabolites and depressed mood in participants not prescribed an antidepressant drug (ATC code, N06A)[77]. Depressed mood was measured by either diagnostic interviews or validated depression questionnaires ($n = 3,966$ cases and $n = 8,887$ controls in the meta-analysis). Detailed definitions of cases and controls in cohorts are given in our previous publication[77]. We adjusted for age, sex, fasting status, lipid-modifying drugs 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 antihypertensive-related metabolites in the non-hypertension population ($n = 2,263$) in ERF and Rotterdam Study. The non-diabetes population was defined as individuals with fasting glucose ≤6.9 mg dl^{-1} and not prescribed anti-diabetic treatment; the non-hypertension population was defined as individuals whose systolic blood pressure was <140 mmHg, diastolic blood pressure <90 mmHg and not prescribed 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 ($P < 0.05$).

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 a previous longitudinal study by Wurtz and co-workers, which also quantified ^{1}H-NMR metabolic profiles in blood samples but focused on the change in metabolite concentrations at two time points—baseline and follow-up[17]. Because the longitudinal study adjusted only 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 and hydrophilic statins were similar in the current study, we used the results of the former, which had the largest sample size for the comparison. The results of MR analysis and the association of rs12916-T and metabolites[17] were also used in the comparison[17].

We then compared significant statin–metabolite associations in the current cross-sectional study with those of rs12916-T and metabolites. We used the GWAS results of NMR metabolites from our previous paper, which included 24,925 individuals not prescribed lipid-modifying medication[11]. This resulted in 55 metabolites being included in the comparison.

PPIs, circulating metabolites and liver function. We studied biochemical variables used in liver function tests—that is, ALT, AST, GGT, AST/ALT, total bilirubin and alkaline phosphatase, and hepatic steatosis. The liver function test used automatic enzymatic procedures (Roche Diagnostics)[78]. Abdominal ultrasonography was performed by a certified and experienced technician (P. Taimr) on a Hitachi HI VISION 900. Images were stored digitally and re-evaluated by a single hepatologist with >10 years of experience in ultrasonography. The diagnosis of steatosis was determined by the ultrasound technician according to the protocol of Hamaguchi et al.[79].

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. Covariates included age, sex, BMI, smoking, lipid-modifying drugs, PPIs and daily alcohol intake calculated from questionnaires. The *P* value threshold was Bonferroni corrected with 10 independent equivalents of 55 PPI-related metabolites and 6 independent equivalents of the 7 liver function measurements ($P < 8.3 \times 10^{-4}$). We further checked the association of PPI use and liver function measurements by linear regression, with adjustment for age, sex, BMI, smoking and daily alcohol intake ($P < 8.3 \times 10^{-3}$).

PPIs, circulating metabolites and gut microbiome. We extracted the associations of PPIs with gut microbiota and (alpha-) diversity from a previous paper by Imhann and co-workers[39]. Age, sex, BMI, antibiotic use and sequence read depth were corrected in the association analysis[39]. In total, 92 bacterial taxa abundances—assessed by tag sequencing of the 16 S rRNA gene[58] 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, false discovery rate <0.05). Forty-five of the 92 bacterial taxa abundance and alpha-diversity factors were also tested for association with metabolites as measured by Nightingale Health in our previous study[80]. In brief, this included 2,309 individuals not prescribed 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 ($P < 3.3 \times 10^{-4}$). The *P* value threshold for alpha-diversity was 5.0×10^{-3}.

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

Data availability

All summary statistics of the meta-analysis, and those utilized in compilation of the figures, are made available through the Supplementary tables. In regard 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 included in this study are available through the 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 are required 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 in regard to overlap with other ongoing projects before access is granted. Data on medication used in the current study are available through the individual studies on 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 was used in the current study.

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. Preprint at *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. *Details and Considerations of the UK Biobank GWAS* (Neale Lab, accessed 10 December 2018); http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas
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**, e34408 (2018).
76. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). 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. *Biol. Psychiatr.* (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. Vojinovic, D. et al. Relationship between gut microbiota and circulating metabolites in population-based cohorts. *Nat. Commun.* **10**, 5813 (2019).

Acknowledgements

We acknowledge all participants included in the cohorts. We also acknowledge the BBMRI Metabolomics Consortium (see Supplementary Information, http://www.bbmri.nl/omics-metabolomics/) funded by BBMRI-NL, a research infrastructure financed by the Dutch government through the 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), the Memorabel program (project no. 733050814), Netherlands X-omics Research Infrastructure and U01-AG061359 NIA. The full list of funding information for each cohort can be found in the cohort acknowledgements below. J.L., C.M.v.D. and A.D. benefitted from exchange grants from the Personalized pREvention of Chronic DIseases consortium (no. H2020-MSCA-RISE-2014). A.D. is supported by the Dutch Science Organization (ZonMW-VENI, grant no. 2015). L.C. is supported by a joint PhD fellowship from China Scholarship Council (no. 201708320268) and University of Groningen. M.G.N. is supported by Royal Netherlands Academy of Science Professor Award (no. PAH/6635) to D.I.B. D.O.M.-K. is supported by the ZonMW-VENI (grant no. 916.14.023). B.H.C.S. received a grant from TransQST (no. 116030-2; IMI2). D.R. is funded by an Erasmus MC mRACE grant (Profiling of the human gut microbiome). Cohort acknowledgements, 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 supervision of the laboratory work and P. Snijders for his help in data collection. ERF was supported by the Consortium for Systems Biology (NCSB), both within the framework of the Netherlands Genomics Initiative (NGI)/NWO). The ERF study as a part of European Special Populations Research Network (EUROSPAN) was supported by European Commission FP6 STRP, grant no. 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement no. HEALTH-F4-2007-201413 by the European Commission under the program 'Quality of Life and Management of the Living Resources' of the 5th Framework Programme (no. QLG2-CT-2002-01254), as well as the FP7 project EUROHEADPAIN (no. 602633). High-throughput analysis of ERF data was supported by a joint grant from NWO and the Russian Foundation for Basic Research (NWO-RFBR no. 047.017.043). High-throughput metabolomics measurements in the ERF study were supported by BBMRI-NL. Rotterdam Study: we thank the study subjects, the staff from Rotterdam Study and the participating pharmacists and general practitioners. The Rotterdam Study is supported by Erasmus MC and Erasmus University Rotterdam; by 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 from Boehringer Ingelheim and Novartis, and unrestricted grants from AstraZeneca and Chiesi. We are also grateful to nurse ultrasonographist, Mrs. van Wijngaarden for performing 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. We thank N. El Faquir and J. Verkroost-Van Heemst for their help in sample collection and registration, and P. van der Wal, K. Arabe, H. Razawy and K. Singh Asra for their help in DNA isolation and sequencing. Furthermore, we thank J. Raes and J. Wang (KU Leuven, Belgium) for their guidance in 16S rRNA profiling and dataset generation. NTR: we are grateful to all twins and their relatives for their continued participation. Funding was obtained from NWO and MagW/ZonMW (grant nos. 904-61-090, 985-10-002, 904-61-193,480-04-004, 400-05-717, Addiction-31160008, Middelgroot-911-09-032 and Spinozapremie 56-464-14192); BBMRI-NL (no. 184.021.007); VU University's Institute for Health and Care Research (no. EMGO1); Neuroscience Campus Amsterdam (NCA); the European Community's Seventh Framework Program (no. FP7/2007-2013); ENGAGE (no. HEALTH-F4-2007-201413); and the European Science Council (ERCAdvanced, no. 230374). M.G.N. is supported by the 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 no. R01AG054628 02 S. NESDA: the infrastructure for the NESDA study (www.nesda.nl) was funded through the Geestkracht program of ZonMw (grant no. 10-000-1002) and through 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 and Rob Giel Onderzoekscentrum). LLS: we thank all participants. This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem, no. IGE05007), the Centre for Medical Systems Biology and the Netherlands Consortium for Healthy Ageing (grant no. 050-060-810), all within the framework of NWO by the BBMRI Metabolomics Consortium funded by BBMRI-NL (NWO, grant nos. 184.021.007 and 184033111). LifeLines DEEP: we thank participants and staff of the LifeLines DEEP cohort for their collaboration. We thank J. Dekens, M. Platteel, A. Maatman and J. Arends for management and technical support. This project was funded by the Netherlands Heart Foundation (IN-CONTROL CVON, grant no. 2012-03 to A.Z. and J.F.; by NWO (nos. NWO-VIDI 864.13.013 to J.F. and NWO-VIDI 016.178.056 to A.Z.; and by the European Research Council Starting Grant no. 715772 to A.Z., who also holds a Rosalind Franklin Fellowship from the University of Groningen. Hoorn DCS: we thank participants of this study and research staff of the Diabetes Care System West-Friesland. High-throughput metabolomics measurements in the DCS study were supported by BBMRI-NL 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 was received from several sources including the Dutch Federation of University Medical Centres, health insurers, NWO, ZonMw, the Dutch Diabetes Foundation, the European Foundation for the Study of Diabetes, International Diabetes Federation, the European Innovative Medicine Initiative and the European Union. Alpha Omega Cohort: the Alpha Omega Cohort is registered with ClinicalTrials.gov. (identifier: NCT03192410). It was funded by the Netherlands Heart Foundation (grant no. 200T401) and the National Institutes of Health (NIH, grant no. 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 for the Alpha Omega Cohort were supported by BBMRI-NL. 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 no. 31 O.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). LUMINA: the LUMINA study is funded by grants obtained from ZonMw (no. 90700217) and VIDI (ZonMw, no. 91711319) (to G.M.T.); by NCSB and the Centre for Medical System Biology (CMSB),

both within the framework of the Netherlands Genomics Initiative (NGI)/NWO (to A.M.J.M.v.d.M.) and by the FP7 EU project EUROHEADPAIN (grant no. 602633) (to A.M.J.M.v.d.M. and G.M.T.). NEO: the authors of the NEO study thank all individuals who participated, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group—P. van Beelen, P. Noordijk and I. de Jonge—for coordination, laboratory and data management of the study. Genotyping in the NEO study was supported by Centre National de Génotypage (Paris, France), headed by J.-F. Deleuze. This study was supported by the participating departments, the Division and the Board of Directors of Leiden University Medical Center and by Leiden University, Research Profile Area Vascular and Regenerative Medicine. The study was also supported by the Netherlands Cardiovascular Research Initiative: an initiative with the support of the Dutch Heart Foundation (no. CVON2014-02 ENERGISE).

Author contributions

A.D., J.F., J.M.G., L.L., J.L., M.N., L.M.t'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.W.J.H.P., F.R., P.E.S., C.D.A.S., B.H.C.S., L.M.t'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.t'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.G.N., R.C.S., L.M.H., C.S.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 the 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.t'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.

Competing interests

D.O.M.-K. is a part-time clinical research consultant for Metabolon, Inc. All other authors have nothing to disclose. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Additional information

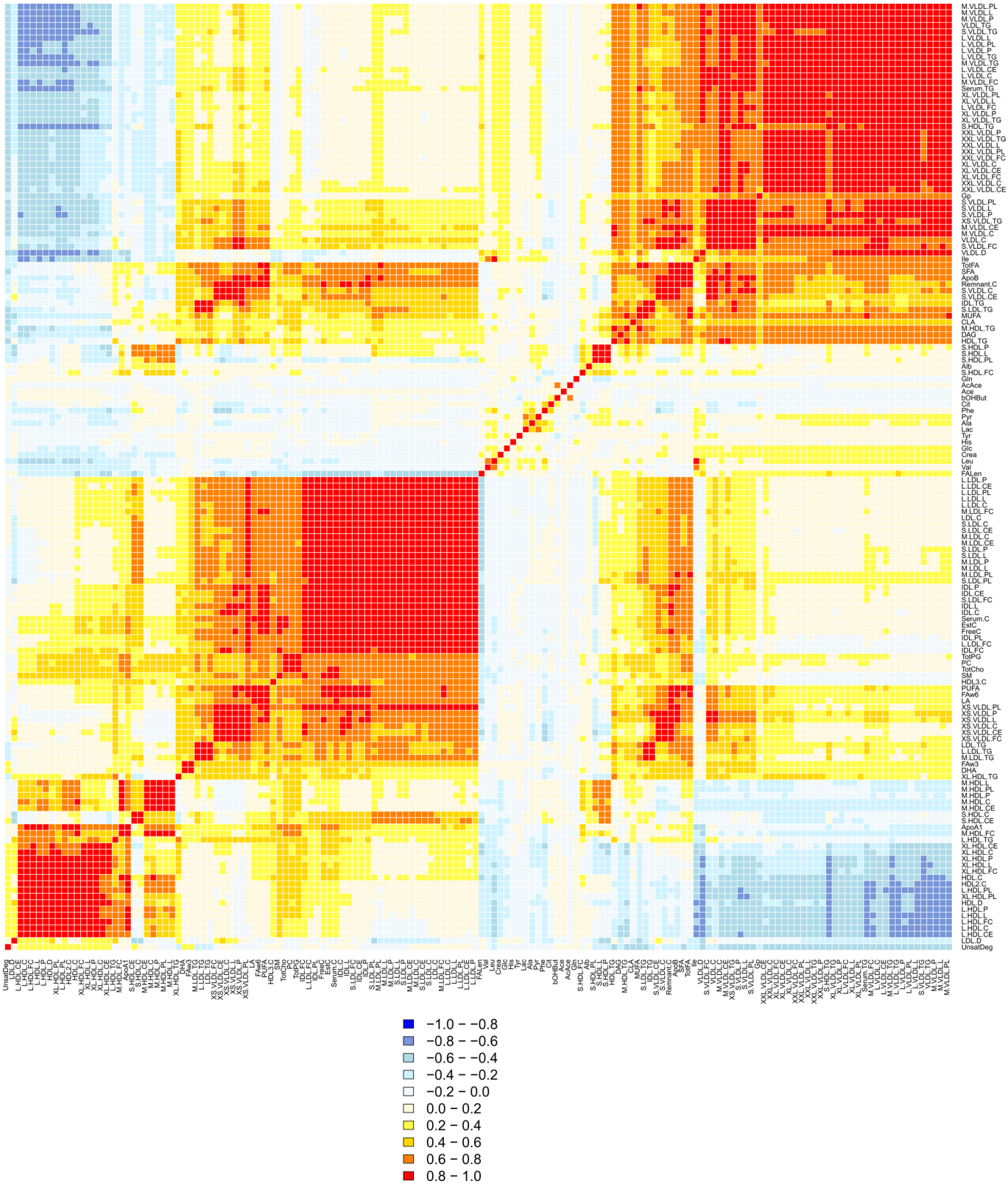
Extended data is available for this paper at https://doi.org/10.1038/s41591-019-0722-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-019-0722-x.

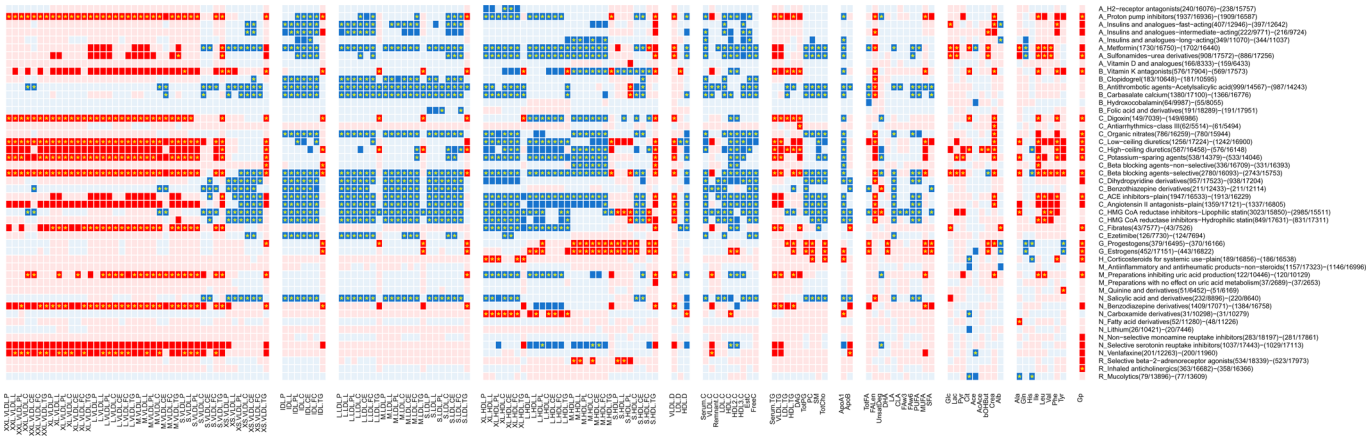
Correspondence and requests for materials should be addressed to J.L. or C.M.v.D.

Peer review information Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

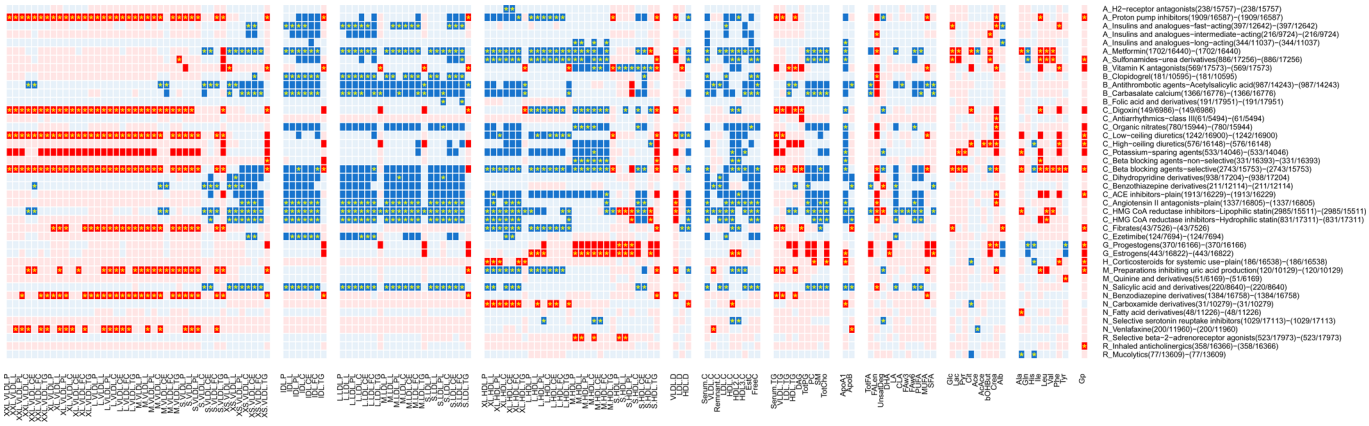
Reprints and permissions information is available at www.nature.com/reprints.



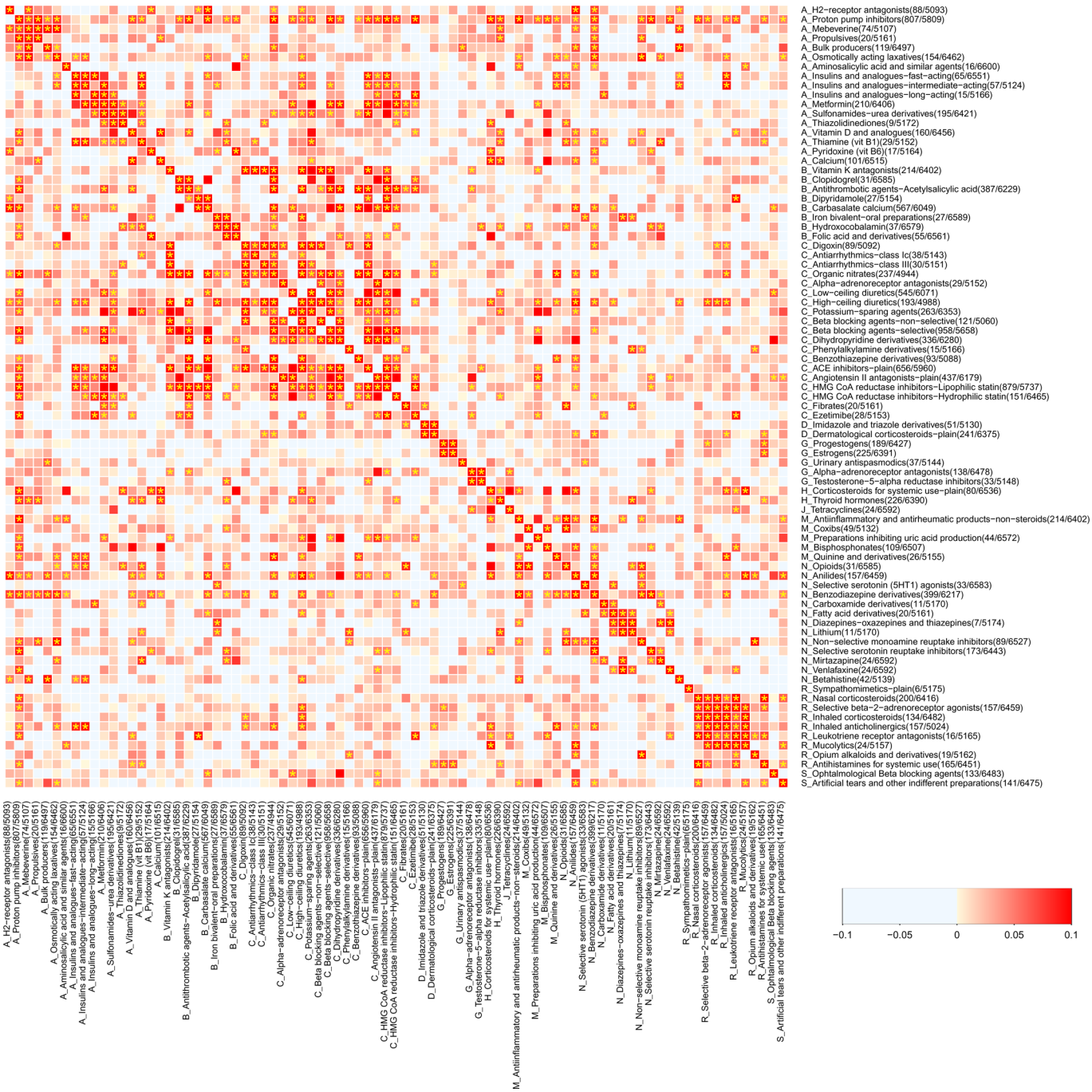
Extended Data Fig. 1 | Correlation between metabolites in Rotterdam Study. The correlation matrix of metabolites were performed by Pearson's correlation (n = 5,191). The hierarchical cluster analysis was used in the clustering. Color in the boxes, correlation coefficient.



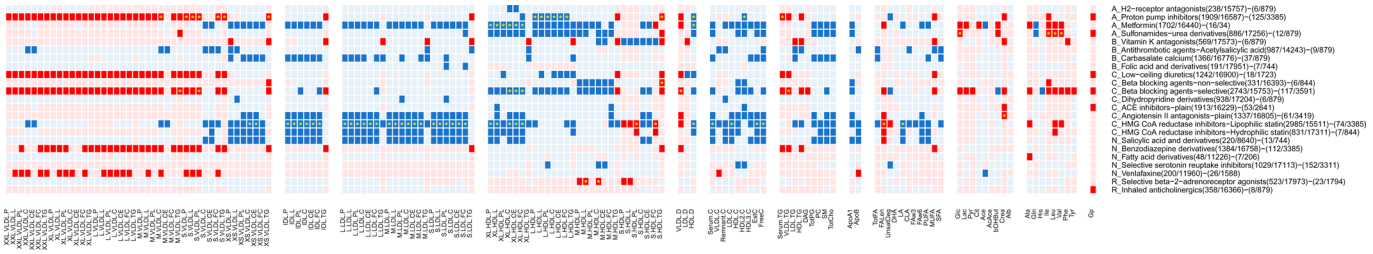
Extended Data Fig. 2 | Drug–metabolite associations in model 1 versus model 2. The drugs with at least one significant metabolite association in baseline model (model 1) by linear regression are shown. The first letter of the ATC code precedes the drug name, to identify different categories. Sample sizes of the drug users and non-users in model 1 (age and sex adjusted) and model 2 (age, sex, BMI and smoking adjusted) are shown following drug names, respectively. Dark red, positive significant associations in model 1 (P-value $< 1.9 \times 10^{-5}$); light red, positive nonsignificant associations in model 1 (P-value $\geq 1.9 \times 10^{-5}$); dark blue, negatively significant associations in model 1 (P-value $< 1.9 \times 10^{-5}$); light blue, negatively nonsignificant associations in model 1 (P-value $\geq 1.9 \times 10^{-5}$). Asterisks in boxes denote that neither direction nor significance status were different between model 1 and model 2 (P-value $< 1.9 \times 10^{-5}$). Two-tailed tests were used.



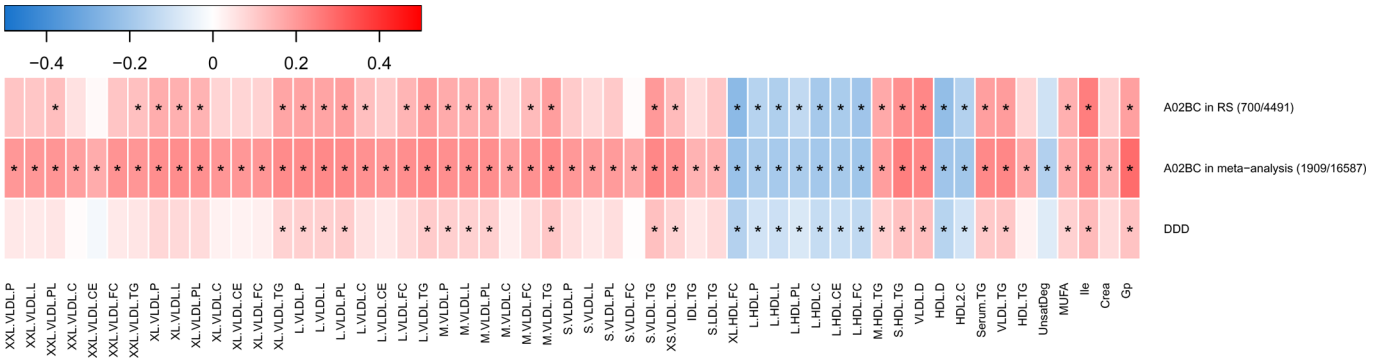
Extended Data Fig. 3 | Drug–metabolite associations in model 2 versus model 3. The drugs with at least one significant metabolite association in model 2 (age, sex, BMI and smoking adjusted) by linear regression are shown. The first letter of the ATC code is shown preceding the drug name, to identify different categories. Sample sizes of the drug users and non-users in model 2 and model 3 (age, sex, BMI, smoking and co-treatment adjusted) are shown following drug names, respectively. Dark red, positive significant associations in model 2 (P-value $< 1.9 \times 10^{-5}$); light red, positive nonsignificant associations in model 2 (P-value $\geq 1.9 \times 10^{-5}$) dark blue, negatively significant associations in in model 2 (P-value $< 1.9 \times 10^{-5}$); light blue, negatively nonsignificant associations in in model 2 (P-value $\geq 1.9 \times 10^{-5}$). Asterisks in boxes denote that neither direction nor significance status was different between model 2 and model 3 (P-value threshold is multiple testing-corrected per drug; See Supplementary Table 4). Two-tailed tests were used.



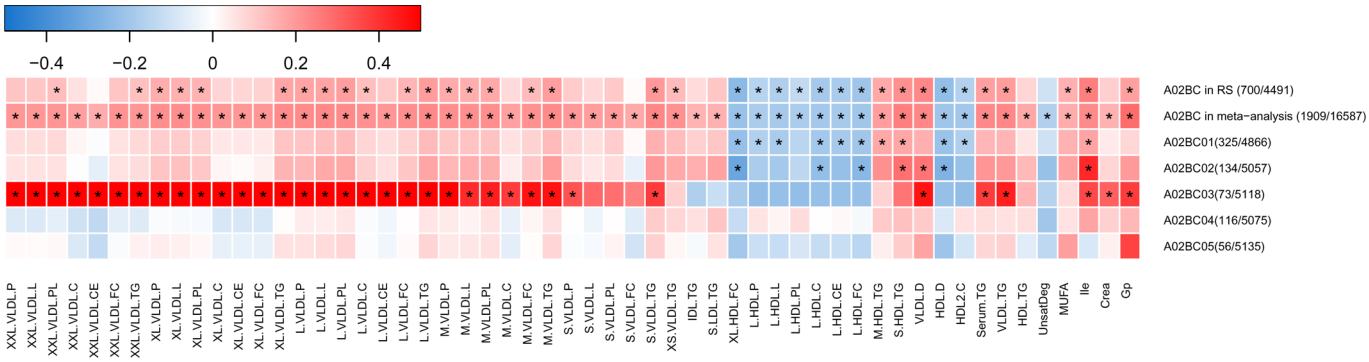
Extended Data Fig. 4 | Correlation between drugs. The correlation matrix of metabolites were performed by Spearman's correlation (n = 6,631). The first letter of the ATC code is shown preceding the drug name, to identify different categories. Sample size of the drug users and non-users is shown following drug names. The depth of the color refers to the correlation coefficients. Asterisks in boxes denote the positively significant correlations (P-value $< 5.9 \times 10^{-4}$). Two-tailed tests were used.



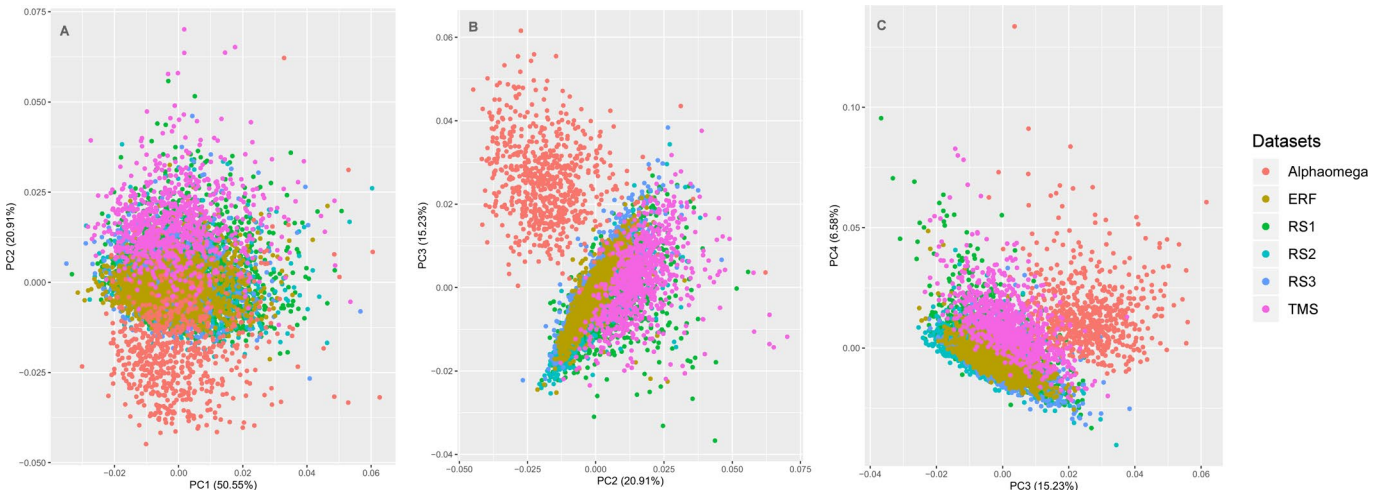
Extended Data Fig. 5 | Drug–metabolite Associations in model 3 versus single drug test. The first letter of the ATC code is shown preceding the drug name, to identify different categories. Single drug test: Association analysis (linear regression) in the sub-samples of patients who use one drug only (one-drug-users) and all-treatment-naive controls. Sample size of the drug users and non-users in model 3 (age, sex, BMI, smoking and co-treatment adjusted) and the single drug test are shown following drug names, respectively. Dark 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; dark 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. Asterisks in boxes denote that 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 Fig. 6 | Association of PPI/dosage and the PPI-related metabolites. The association of dosage of PPI and metabolites were tested by linear regression in Rotterdam Study (n = 700). The PPI-related metabolites were selected in model 3. DDD, defined daily dose of PPI. (/), sample size of user/non-user. Red, positive association, blue, negative association. The depth of the color refers to the association estimates. Asterisks in boxes denote significance after correcting for multiple test (P-value < 0.004). Two-tailed tests were used.



Extended Data Fig. 7 | Association of specific PPI drugs and the PPI-related metabolites. The association of PPI drugs (A02BC) and metabolites were tested by linear regression in Rotterdam Study. The PPI-related metabolites were selected in model 3. A02BC01, omeprazole; A02BC02, pantoprazole; A02BC03, lansoprazole; A02BC04, rabeprazole; A02BC05, esomeprazole. (/), sample size of user/non-user. Red, positive association; blue: negative association. The depth of the color refers to the association estimates. Asterisks in boxes denote significance after correcting for multiple test (P-value < 0.004). Two-tailed tests were used.



Extended Data Fig. 8 | 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.