Plasma metabolites to profile pathways in non-communicable disease multimorbidity

Maik Pietzner (PhD)[1], Isobel D. Stewart (PhD)[1], Johannes Raffler (PhD)[2], Kay-Tee Khaw (FRCP)[3], Gregory A. Michelotti (PhD)[4], Gabi Kastenmüller (PhD)[1,2], Nicholas J. Wareham (FRCP)[1], Claudia Langenberg (MD)[1,5,6*]

[1]MRC, Epidemiology Unit, University of Cambridge, Cambridge, UK
[2]Institute of Computational Biology, Helmholtz Zentrum München, Germany
[3]Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom
[4]Metabolon Inc, Durham, North Carolina USA
[5]Health Data Research UK, Wellcome Genome Campus and University of Cambridge, UK
[6]Computational Medicine, Berlin Institute of Health (BIH), Charité University Medicine, Berlin, Germany

*****Correspondence to:
Dr Claudia Langenberg
MRC Epidemiology Unit
Box 285
Institute of Metabolic Science
Cambridge Biomedical Campus
University of Cambridge
Cambridge CB2 0QQ, UK
E-mail: claudia.langenberg@mrc-epid.cam.ac.uk
Telephone: +44 1223 769216

ABSTRACT

Multimorbidity, the simultaneous presence of multiple chronic conditions, is an increasing global health problem and research into its determinants is of high priority. We used baseline untargeted plasma metabolomics profiling covering >1,000 metabolites as a comprehensive read out of human physiology to characterise pathways associated with and across 27 incident non-communicable diseases (NCDs) assessed using electronic health record hospitalisation and cancer registry data from over 11,000 participants (219,415 person-years). We identified 420 metabolites shared between at least 2 NCDs, representing 65.5% of all 640 significant metabolite-disease associations. We integrated baseline data on over 50 diverse clinical risk factors and characteristics to identify actionable shared pathways represented by those metabolites. Our study highlights liver and kidney function, lipid and glucose metabolism, low-grade inflammation, surrogates of gut microbial diversity, and specific health-related behaviours as antecedents of common NCD multimorbidity with potential for early prevention. We integrated results into an open-access webserver (https://omicscience.org/apps/mwasdisease/) to facilitate future research and meta-analyses.

INTRODUCTION

Deep molecular profiling of human blood has the potential to identify novel pathways to disease, improve risk prediction and to enable stratified prevention and management[1]. Prospective studies have shown the promise of deep phenotypic profiling for precision medicine[2,3] but these were very small scale and focused on single diseases[4,5]. Many pathways are shared across different diseases and one in four patients now presents with two or more chronic conditions at the same time, referred to as multimorbidity[6,7]. The incidence of non-communicable disease (NCD) multimorbidity is increasing not only in high-income[8,9] but also in middle and low-income countries[7,10], which poses major challenges for health care systems globally.

The co-occurrence of conditions, such as type 2 diabetes (T2D) and cardiovascular diseases, is common and previous work has shown a high degree of interconnectivity with other diseases [11]. The lack of horizontal integration between specialities delivering care for patients with co-existing diseases means that multimorbidity is more likely to be seen as a random assortment of individual conditions. There is now a call by public health authorities and policy makers for a shift to recognising multimorbidity as an accumulation of largely predictable clusters of disease in the same person[12]. However, the knowledge about shared aetiologies of less obviously related diseases is sparse. Molecular profiling has the potential to simultaneously and systematically identify pathways across many different incident diseases assessed objectively and at scale. Research into the determinants of NCD multimorbidity is a high priority[12], but, to our knowledge, investigations of in-depth molecular profiles in large prospective cohorts with comprehensive, long-term clinical follow-up have not been previously undertaken. Detailed information on modifiable factors that underlie and drive shared risk, which is required to establish actionable insights for prevention and management of multimorbidity[13], is also lacking.

The human blood metabolome provides a comprehensive read out of human physiology obtained through untargeted assessment of hundreds of small circulating molecules, which reflect influences and interactions of genetics, lifestyle, environment, medical treatment, and microbial activity[14]. We investigated associations between baseline levels of 1,014 metabolites assessed through untargeted profiling of plasma samples and the onset of 27 NCDs, all-cause mortality, and NCD multimorbidity (Extended Data Figure 1). Clinical outcomes were assessed using electronic health record hospitalisation and cancer registry data in over 11,000 participants (219,415 person-years of follow-up) of the European Prospective Investigation of Cancer (EPIC)-Norfolk study[15].

We systematically analysed and established a comprehensive catalogue of risk factor–metabolite–disease associations to address unanswered questions related to the shared aetiology and drivers of multiple chronic conditions and multimorbidity. We sought to characterise 1) pathways at baseline shared across multiple incident conditions, to identify those that predispose individuals to multimorbidity; 2) which of the identified metabolite-disease associations are driven by modifiable clinical and other risk factors, to identify targets of interventions; and 3) metabolites most strongly associated with the onset of NCD multimorbidity. We share our results through an open-access webserver (https://omicscience.org/apps/mwasdisease/) to maximise the use of this resource considerably augmenting existing efforts[16].

RESULTS

We used data from the EPIC-Norfolk cohort, which includes 25,639 middle-aged participants from the general population of Norfolk in Eastern England[15]. A quasi-random subsample of 11,966 participants (mean age of 60 years [s.d.: 9 years], 53.7% females) was selected for metabolomic profiling using the Metabolon HD4 platform and detailed characteristics of participants and metabolites can be found in Supplemental Tables 1-3.

Small molecule profiles of incident diseases

Plasma levels of 458 metabolites were significantly associated with at least one incident disease or all-cause mortality representing 1,226 associations in total (trait-wise Bonferroni cut-off for significance accounting for the number of metabolites: $p<4.95x10^{-5}$; **Extended Data Fig. 2**). All-cause mortality was associated with the majority of those metabolites (n=268) followed by incident T2D (n=214), chronic obstructive pulmonary disease (COPD) (n=142), coronary heart disease (CHD) (n=127), heart failure (n=110), renal disease (n=110), peripheral arterial disease (PAD) (n=95), lung cancer (n=43), liver disease (n=39), atrial fibrillation (AF) (n=27), abdominal aortic aneurysms (AAA) (n=21), and asthma (n=16). We observed only few associations with incident colon cancer (n=5), cataract (n=5), cerebral stroke (n=2), stomach cancer (n=1), and Parkinson's disease (n=1). The five most significant associations for each of the incident diseases as well as all-cause mortality are shown in Extended Data Figure 3. The number of metabolites associated with each disease outcome was partly explained by the number of cases for each disease and hence the power to detect an association (**Extended Data Fig. 4**). Specifically, incident T2D, COPD, PAD, and lung cancer were associated with more metabolites than expected based on the overall relationship between the number of cases and the number of associated metabolites in the present study (**Extended Data Fig. 4**). The opposite was the case for incident cerebral stroke, eye diseases or skin cancers, among others.

We observed highly correlated effect sizes (r>0.9 for most analyses) while testing for an effect of delayed diagnosis of patients in various sensitivity analysis, including logistic regression models and exclusion of participants with any event up to five years after baseline examinations (**Extended Data Fig. 5**). This, however, might not exclude the possibility that effect estimates obtained in the present study could underestimate the effect for conditions usually defined in primary care settings, such as fractures or cataracts.

We identified 54 metabolite–outcome associations with suggestive evidence (p<0.001) for differing effect sizes between men and women (**Extended Data Fig. 6**) of which seven passed the more stringent Bonferroni-corrected threshold, including larger effect sizes in women for orotidine, erythornate, and three unknown compounds with incident CHD. We provide sex-specific effect estimates along with p-values for sex-interaction effects for all metabolite–outcome associations in a webserver published along with this study (https://omicscience.org/apps/mwasdisease/).

Two-thirds of associated metabolites are shared among diseases

A total of 420 (65.6%) metabolites were associated with at least two different diseases or all-cause mortality (p<0.001 see **Methods**, **Fig. 1**) and 220 (34.6%) metabolites were specifically associated with one disease only (**Fig. 2**). We observed high connectivity among cardiometabolic and respiratory diseases including CHD, heart failure, T2D, cerebral stroke, PAD, renal and liver diseases, COPD, and lung cancer across different biochemical classes of metabolites (**Fig. 2**). Plasma levels of the non-classical carbohydrate N-acetylneuraminate were positively associated with 14, partly unrelated, diseases, including incident stomach, oesophageal, and lung cancer as well as major cardiovascular events and metabolic diseases (**Fig. 2**). Highly pleiotropic metabolites, i.e. those associated with multiple diseases, showed wide biochemical and biological diversity (**Fig. 2**), and included N-acetylated amino acids (e.g. N-acetylphenylalanine), surrogate markers of smoking (e.g. cotinine), modified nucleotides (e.g. pseudouridine), glycerophsopholipids (e.g. 1-palmitoyl-2-oleoyl-GPC), catabolites of vitamin C (e.g. threonate), products of microbial metabolism (e.g. indolepropionate), sulphated steroids (e.g. epiandrosterone sulfate), haem degradation products (e.g. bilirubin (E,E)), proteinogenic amino acids (e.g. serine), and several compounds of yet unknown identity (e.g. X-11429).

We identified some metabolites with shared associations among seemingly unrelated diseases. Plasma levels of the unknown compound X-11305 were inversely associated with the risk of colon cancer, heart failure, PAD, COPD, and mortality. In another example, plasma levels of maltose were positively associated with stomach cancer, T2D, heart failure, CHD, PAD, venous thrombosis, COPD, and mortality.

The vast majority (93%) of metabolites associated with multiple outcomes showed consistent effect directions across all significantly associated diseases, i.e. being either positively or inversely associated with all diseases. Exceptions included N-acetylmethionine which was inversely associated with incident T2D and liver diseases but positively with incident AAA, heart failure, PAD, renal diseases, COPD, and mortality; and the unknown compound X – 23997 which was inversely associated with prostate cancer but positively with Parkinson's disease.

In-depth exploration of these and other examples, along with additional results, is possible via our webserver (https://omicscience.org/apps/mwasdisease/)**.**

Integration of diverse traits at baseline identifies actionable antecedents

To put the identified small molecule profiles into context and identify actionable antecedents, i.e. possible targets for intervention or management, we quantified the explained variance for each metabolite using information on more than 50 diverse participant baseline characteristics. Prevalent conditions, anthropometric and lifestyle markers, as well as comprehensive clinical chemistry markers (**Extended Data Fig. 7A**) were included in the analysis. Almost every measured metabolite (972 out of 1,014) was significantly associated ($p<4.93x10^{-5}$) with at least one trait in cross-sectional analyses (**Extended Data Fig. 7B**).

To identify dependencies among specific risk factors, metabolites and diseases of interest, we utilised a formal mediation analysis framework. To match triplets among risk factors, metabolites and outcomes, we ran Cox models for 21 baseline characteristics that were selected based on clinical utility and to minimize redundancy (**Extended Data Fig. 8**). Out of 6,364 possible paths (significant and directionally consistent triangles between risk factor–metabolite–disease, **Methods**), 1,084 (17.0%) had a significant indirect effect ($p<7.8x10^{-6}$) indicating a relationship between a risk factor and a metabolite with respect to a specific disease. We thereby identified common antecedents, i.e. exposures associated with multiple metabolites and outcomes, such as obesity (waist-to-hip ratio or BMI), inflammation (fibrinogen), measures of liver (liver enzyme levels) and kidney function (uric acid and creatinine), blood lipids, systolic blood pressure, smoking behaviour, and glucose homeostasis (**Fig. 3A**). The median proportion mediated was 15.7% (IQR: 11.0% - 26.6%; **Extended Data Fig. 9 and Supplemental Table 4**) and effects largely mediated by metabolites appeared to be exposure specific, e.g. N-formylmethionine was estimated to mediate 47.3% of the effect of uric acid or creatinine on renal disease, CHD, and mortality on average (**Fig. 3B**). We identified a few metabolites possibly mediating the associations of multiple exposures (n≥10) on multiple outcomes (n≥5), including X - 12117 (**Fig. 3C**), C-glycosyltryptophan, N-acetylneuraminate, N-acetylglucoseamine, mannose, 1-palmitoyl-2-oleoyl-GPE (16:0/18:1), and X – 11429, representing antecedents such as kidney function, inflammation as well as glucose and lipid metabolism.

We note that for some exposures metabolite associations superseded exposure associations as indicated by complete attenuation of risk factor associations or a proportion of mediated effect larger than 100%, e.g. metabolites such as C-glycosyltryptophan or pseudouridine might be better markers to judge the risk associated with kidney function decline on all-cause mortality. Further, X - 12117 almost completely mediated the increased risk associated with BMI on all-cause mortality (**Supplemental Table 4**).

To validate the effect of identified antecedents, we included those (i.e. body mass index, waist-to-hip ratio, smoking behaviour, serum uric acid concentrations, total triglycerides, HDL-cholesterol, random glucose, serum alkaline phosphatase concentrations, serum vitamin C concentrations, systolic blood pressure, and plasma fibrinogen concentrations) as additional covariates to the initial Cox regression models. Consequently, the number of associated metabolites more than halved (361 compared to 640 with $p<0.001$, **Supplemental Table 5**) and the proportion of uniquely associated metabolites increased to 56.2% (203 out of 361).

Metabolites specifically associated with diseases

A total of 79 metabolites (**Supplemental Table 6**) showed evidence of being uniquely associated with incident T2D (n=36), all-cause mortality (n=21), COPD (n=10), CHD (n=10), or liver disease (n=2). The metabolite with the strongest association was gamma-glutamylglycine, for which one standard deviation (SD) increase in plasma levels was associated with a 37% lower risk for incident T2D (hazard ratio [HR] per SD increase in metabolite levels: 0.63; 95%-CI: 0.58 – 0.68; $p<1.6\times10^{-28}$). Formation of gamma-glutamyl amino acids is facilitated at the plasma membrane by gamma-glutamyl transpeptidase activity and contributes to amino acid influx and formation of the essential antioxidant glutathione[17]. Our cross-disease comparison revealed two distinct subgroups of gamma-glutamyl peptides. In addition to gamma-glutamylglycine, gamma-glutamylthreonine and gamma-glutamyltyrosine were uniquely associated with incident T2D (**Supplemental Table 5**) whereas gamma-glutamylglutamine or gamma-glutamylisoleucine were associated with multiple phenotypes, including incident T2D, and have been previously suggested as markers of liver injury[18]. Such systematic investigations can pinpoint disease-characterizing perturbations in amino acid flux.

Other examples of uniquely associated metabolites included plasma levels of 7-methylxanthine (HR for COPD: 1.24 [1.14; 1.34], $p<1.9\times10^{-8}$), 1-palmitoyl-2-stearoyl-GPC (16:0/18:0) (HR for CHD: 1.12 [1.07; 1.17], $p<6.6\times10^{-7}$), and 2-palmitoleoylglycerol (16:1) (HR for liver disease: 1.28 [1.14; 1.43], $p<2.0\times10^{-5}$).

From multiple outcome associations to NCD multimorbidity

We identified 1,858 (32.6%) participants who developed multiple chronic conditions during follow-up and Figure 4 displays a detailed composition of disease counts.

Plasma levels of 30 metabolites were significantly associated ($p<4.93\times10^{-5}$) with the risk of NCD multimorbidity (defined as developing ≥2 chronic conditions during follow-up) (**Fig. 4, 5 and Supplemental Table 7**). Odds ratios ranged between 1.29 (cotinine; 95%-CI: 1.16 – 1.42) and 0.82 (beta-cryptoxanthin, 95%-CI: 0.77 – 0.87) per one SD increase in metabolites levels and were comparable to those from other baseline characteristics such as C-reactive protein [1.28 (1.20; 1.37)] or the waist-to-hip ratio [1.27 (1.15; 1.40)] (**Supplemental Table 8**). The majority of metabolites that were associated with NCD multimorbidity were also associated with multiple chronic conditions in disease-wise Cox models (Pearson correlation coefficient: 0.41, $p<2.2\times10^{-16}$).

To identify common traditional clinical measures that are antecedents of NCD multimorbity, we first clustered the 30 multimorbidity-associated metabolites to account for their correlated structure and derived nine different clusters (**Extended Data Fig. 10**). From each of the clusters we chose the metabolite with the largest effect size as a representative. Some antecedents were immediately apparent, including smoking behaviour *via* cotinine, lipoprotein metabolism *via* 1-stearoyl-2-meadoyl-GPC, kidney function *via* C-glycosyltryptophan, and vitamin C metabolism *via* cysteine sulfinic acid, all indicated by a large amount (>10%) of variance explained in metabolite levels through those risk factors (**Fig. 6**). Plasma levels of N-acetylphenylalanine were again best explained by surrogate markers of kidney function but seem to reflect body composition as well, given that waist-to-hip ratio explained 5.7% of its variance. Further, haem degradation which is tightly linked to sufficient iron supply might be the most likely explanation for the pattern seen with bilirubin (Z,Z).

We identified other potential novel antecedents of NCD multimorbidity, such as plasma levels of 3-phenylpropionate and indolepropionate, since variation in plasma levels of these metabolites were only partly explained by traditional clinical measures.

Possible biochemical pathways related to the onset of NCD multimorbidity

Metabolomics profiling allows for comprehensive characterisation of pathways shared among multiple diseases and contributing to NCD multimorbidity in conjunction with established risk factors. Prominent associations for N-acetylated amino acids, in particular N-acetylalanine, were consistently present in all analyses performed and variance in plasma levels was best explained by estimated baseline glomerular filtration rate (inversely associated). Expression of aminoacylase 1, the most abundant aminoacylase that catabolises N-acetylated amino acids, is highest in the cytosol of tubular cells of the kidneys[19]. Impaired kidney function over and above a reduced glomerular filtration rate, indicated by altered aminoacylase activity, is likely to be a major disease driver, emphasizing the importance of kidney function and management of kidney disease for the prevention of NCD multimorbidity. Associations with N-acetylated amino acids were not limited to major cardiovascular events - for which chronic kidney disease is a known independent risk factor[20] - but also included lung cancer, COPD, T2D, and liver disease.

Inflammation or so-called inflammaeging[21] has been suggested to be an important risk factor for diverse diseases and we observed a related molecular signatures among the metabolites associated with multiple outcomes. N-acetylneuraminate and N-acetylglucosamine are part of the glycocalyx surrounding the apical membrane of epithelial cells contributing to vascular integrity by regulating permeability[22]. Shedding in response to inflammatory stimuli[23] of the glycocalyx leads to higher concentrations of its components, like N-acetylneuraminate, in the circulation. A functional role of N-acetylneuraminate during myocardial infarction has been suggested and pharmacological suppression of the producing enzyme neuramidase-1 using influenza medication was shown to preserve cardiomyocytes from injury during infarction[24]. It remains to be established whether N-acetylneuraminate has a functional role in mediating the effect of low-grade inflammation on the risk of chronic conditions such as cardiovascular and pulmonary diseases, including lung cancer and T2D.

Our results highlight putative novel antecedents of NCD multimorbidity, including 3-phenylpropionate (hydrocinnamic acid) and indolepropionate, plasma level of which were only weakly explained by established risk factors. Both metabolites have previously been linked to greater diversity of the gut microbiome as measured by the Shannon index[25]. Circulating levels in blood might therefore act as an indirect readout for the relative abundance of species such as *Clostridium* in the gut[26]. Cross-sectional studies have shown a variety of associations between the abundance of microbial species in the gut and several prevalent chronic conditions[27,28]. The microbial-derived metabolite trimethylamine-N-oxide[29] has been shown to be a candidate mediator for the adverse effect of red meat consumption on CVD risk and was associated with an increased risk of heart failure and mortality in our study. However, high red meat consumption explained only little (0.2%) in the variance of trimethylamine-N-oxide plasma levels compared with markers of kidney function (3.2%).

The aetiology of gut dysbiosis remains to be established, but a diet poor in fibre has been suggested to contribute to overgrowth of harmful species, such as *Clostridium* or *Bacteroides*, diminishing overall diversity and production of microbial metabolites beneficial for the host, such as short-chain fatty acids[30]. The ability to characterise individual disease trajectories in-depth using microbial profiling along with other high-resolution ‘omics’ data has been demonstrated in a small pioneering study of around 100 individuals at high risk for metabolic diseases[2,4]. Here we show that plasma levels of surrogates of microbial diversity are inversely associated with several common severe incident NCDs, including T2D, renal diseases, heart failure, CHD, asthma, COPD, lung cancer, and all-cause mortality as well as multimorbidity using objectively ascertained outcomes from a long-term prospective population-based study. We cannot, however, exclude that other factors related to diet not investigated in the present study, such as a healthier lifestyle, might have contributed to our observations.

DISCUSSION

Multimorbidity is becoming the rule rather than the exception in clinical practice and identification of shared disease mechanisms and modifiable drivers is high priority[31]. Through systematic, data-driven integration of the metabolome and phenome with near-complete follow-up using externally derived electronic health record data for 27 major diseases and all-cause mortality, we identify common and possibly actionable antecedents related to the onset of multiple NCDs and multimorbidity. In-depth molecular profiling together with detailed baseline characterisation of participants highlights mediating pathways through characterisation of triangles of clinical risk factor-metabolite-disease links.

We identified obesity, smoking, impaired glucose homeostasis, low-grade inflammation, lipoprotein metabolism, liver and kidney function as common actionable antecedents of NCD multimorbidity, i.e. there are already established treatment or prevention strategies to attenuate the associated disease risk. These common risk factors account for the majority of premature deaths worldwide[32], and our results now highlight their central role for the potential prevention and management of multimorbidity in health care systems, together with previous studies[33,34].

Patients at greatest risk for multimorbidity are those with a pre-existing chronic condition. Effective prevention strategies focused on multimorbidity need to be anchored within primary care and secondary prevention efforts[35]. Our data-driven approach suggests that a focus on monitoring of kidney and liver function and glycaemic control, together with weight loss and smoking cessation support, are essential for the prevention and management of multimorbidity among middle aged and older individuals with chronic conditions.

The diverse nature of the antecedents identified in the current study, including the gut microbiome, calls for the consideration of a broad and novel range of risk factors in the care of patients with chronic conditions who are at risk of multimorbidity, which may go beyond the single-disease focus of specialist care[36]. Linkage of the molecular patterns or antecedents that we have identified with the incidence of specific subtypes of multimorbidity[37], i.e. clusters of more frequently co-occurring diseases, can help to inform successful prevention and intervention strategies managed in general practice. Further, integration of molecular pathways shared across multiple diseases, as identified in the present study, can guide identification of subtypes of multimorbidity by investigating how those molecules or pathways associate with or even determine co-occurrence of seemingly unrelated diseases, for instance guided by comorbidity networks[38,39], in independent studies.

We found sparse evidence for discordant directions of associations of specific metabolites across different diseases, which suggests that intervening on identified shared pathways has potential to convey benefit in a consistent way and to not increase the risk of developing other conditions.

Our systematic comparison across NCDs allowed us to untangle associations among closely related molecules, such as a liver-function independent association between certain gamma-glutamyl amino acids and incident T2D. To our knowledge, we provide the most comprehensive catalogue of risk factor–metabolite associations reported to date, which helped us to contextualise our findings and can inform future metabolomics studies. Our data-driven and hypothesis-free approach allowed us to challenge current concepts of the most important host factors explaining variation in plasma levels of microbial metabolites, for instance estimated glomerular filtration explained more variance in plasma levels of trimethylamine-N-oxide compared with high meat intake. Our mediation approach to triangulate risk factors, metabolites, and diseases does not prove causality and strong correlations among metabolites and risk factors make it almost impossible to pinpoint the true underlying relation from observational data and complementary methods, for instance incorporating genetic techniques, might help to identify key mechanisms.

We have generated an easily accessible web application to enable the interrogation of these results in an interactive way and have provided an intuitive graphical representation of the results. The web application allows the identification of factors explaining the variance of specific plasma metabolites of interest and the query of individual disease summary statistics for future meta-analyses and power calculations, specifically for some of the less common outcomes. It also enables comparison with diseases not studied for the purpose of this analysis, and may help other investigators to prioritise metabolomics approaches, for example lipidomics, for in-depth investigation of specific diseases in new studies.

To our knowledge, this is the first study integrating comprehensive metabolomic and phenotypic profiling with detailed assessment of multiple incident diseases at the same time. Our study distinguishes by having near-complete follow-up of 219,415 person-years, which maximises power and minimises selection bias. Application of Cox models was an appropriate for most of the investigated metabolite – endpoint associations but we cannot completely rule out the possibility that some relationships might be better modelled with other statistical strategies. Despite being the largest study of its kind to date and having long-term follow-up, we were unable to provide coverage of rare and infectious diseases as well as the less severe spectrum of the diseases included, which would be better covered by inclusion of primary care data. Large-scale biobank studies with hundreds of thousands of participants linked with electronic health records from primary care, such as UK Biobank, could provide such opportunities in the future, especially if they cover not only metabolomics as a comprehensive snapshot of human physiology, but other 'omics' data (e.g. proteomics[40]) that provide distinct and complementary information to extend the findings from the present study.

ACKNOWLEDGEMENTS

We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research. The EPIC-Norfolk study (https://doi.org/10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). Metabolite measurements in the EPIC-Norfolk study were supported by the MRC Cambridge Initiative in Metabolic Science (MR/L00002/1) and the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement no. 115372. MP was supported by a fellowship of the German Research Foundation (DFG 1446/2-1). JR is supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant no. 01ZX1912D).

AUTHOR CONTRIBUTIONS

M.P. and C.L. designed the analysis and drafted the manuscript. M.P. and I.D.S. analysed the data. J.R. and G.K. designed and implemented the webserver. K.K. and N.J.W. are PIs of the EPIC-Norfolk cohort. G.A.M. advised on metabolite mapping across batches and provided annotations for retired unknown compounds. All authors contributed to the interpretation of results and critically reviewed the manuscript.

COMPETING INTEREST

G.A.M. is an employee of Metabolon Inc. All other authors declare no competing interest.

REFERENCES

1. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat. Rev. Genet.* **19**, 299–310 (2018).
2. Zhou, W. *et al.* Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* **569**, 663–671 (2019).
3. Alpert, A. *et al.* A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat. Med.* **25**, 487–495 (2019).
4. Schüssler-Fiorenza Rose, S. M. *et al.* A longitudinal big data approach for precision health. *Nat. Med.* **25**, 792–804 (2019).
5. Hoyles, L. *et al.* Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. *Nat. Med.* **24**, 1070–1080 (2018).
6. Barnett, K. *et al.* Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* **380**, 37–43 (2012).
7. Yao, S.-S. *et al.* The prevalence and patterns of multimorbidity among community-dwelling older adults in China: a cross-sectional study. *Lancet* **392**, S84 (2018).
8. Lebenbaum, M., Zaric, G. S., Thind, A. & Sarma, S. Trends in obesity and multimorbidity in Canada. *Prev. Med. (Baltim).* **116**, 173–179 (2018).
9. van Oostrom, S. H. *et al.* Time Trends in Prevalence of Chronic Diseases and Multimorbidity Not Only due to Aging: Data from General Practices and Health Surveys. *PLoS One* **11**, e0160264 (2016).
10. Hussin, N. M. *et al.* Incidence and predictors of multimorbidity among a multiethnic population in Malaysia: a community-based longitudinal study. *Aging Clin. Exp. Res.* **31**, 215–224 (2019).
11. Guasch-Ferré, M. *et al.* Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care* **39**, 833–846 (2016).
12. Whitty, C. J. M. *et al.* Rising to the challenge of multimorbidity. *The BMJ* vol. 368 (2020).
13. Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing. *Nature* **561**, 45–56 (2018).
14. Nicholson, J. K. *et al.* Metabolic phenotyping in clinical and surgical environments. *Nature* **491**, 384–392 (2012).
15. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br. J. Cancer* **80 Suppl 1**, 95–103 (1999).
16. Liu, J. *et al.* Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug–metabolite atlas. *Nat. Med.* **26**, 110–117 (2020).
17. Griffith, O. W. & Meister, A. *Glutathione: Interorgan translocation, turnover, and metabolism*. vol. 76 (1979).
18. Soga, T. *et al.* Serum metabolomics reveals γ-glutamyl dipeptides as biomarkers for discrimination among different forms of liver disease. *J. Hepatol.* **55**, 896–905 (2011).
19. Sommer, A. *et al.* The molecular basis of aminoacylase 1 deficiency. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1812**, 685–690 (2011).
20. Gansevoort, R. T. *et al.* Chronic kidney disease and cardiovascular risk: Epidemiology, mechanisms, and prevention. *Lancet* **382**, 339–352 (2013).
21. Ferrucci, L. & Fabbri, E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat. Rev. Cardiol.* **15**, 505–522 (2018).
22. Varki, A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* vol. 446 1023–1029 (2007).
23. Jourde-Chiche, N. *et al.* Endothelium structure and function in kidney health and disease. *Nat. Rev. Nephrol.* **15**, 87–108 (2019).
24. Zhang, L. *et al.* Functional Metabolomics Characterizes a Key Role for N-Acetylneuraminic Acid in Coronary Artery Diseases. *Circulation* **137**, 1374–1390 (2018).
25. Pedersen, H. K. *et al.* Human gut microbes impact host serum metabolome and insulin

sensitivity. *Nature* **535**, 376–381 (2016).

26. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food components. *Eur. J. Nutr.* **57**, 1–24 (2018).
27. Martin, T. *et al.* Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat. Commun.* **9**, 1–8 (2018).
28. Aulchenko, Y. S. *et al.* Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science (80-.).* **352**, 565–569 (2016).
29. Heianza, Y., Ma, W., Manson, J. A. E., Rexrode, K. M. & Qi, L. Gut microbiota metabolites and risk of major adverse cardiovascular disease events and death: A systematic review and meta-analysis of prospective studies. *J. Am. Heart Assoc.* **6**, (2017).
30. Canfora, E. E., Meex, R. C. R., Venema, K. & Blaak, E. E. Gut microbial metabolites in obesity, NAFLD and T2DM. *Nat. Rev. Endocrinol.* 1 (2019) doi:10.1038/s41574-019-0156-z.
31. Multimorbidity: a priority for global health research. *Acad. Med. Sci.* (2018).
32. Stanaway, J. D. *et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990â€"2017: a systematic analysis for the Global Burden of Disease S*. www.thelancet.com (2018) doi:10.1016/S0140-6736(18)32225-6.
33. Wikström, K., Lindström, J., Harald, K., Peltonen, M. & Laatikainen, T. Clinical and lifestyle-related risk factors for incident multimorbidity: 10-year follow-up of Finnish population-based cohorts 1982-2012. *Eur. J. Intern. Med.* **26**, 211–216 (2015).
34. Freisling, H. *et al.* Lifestyle factors and risk of multimorbidity of cancer and cardiometabolic diseases: a multinational cohort study. *Christina C. Dahm* **12**,.
35. Smith, S. M., Wallace, E., O'Dowd, T. & Fortin, M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. *Cochrane Database of Systematic Reviews* vol. 2016 (2016).
36. Tinetti, M. E., Fried, T. R. & Boyd, C. M. Designing health care for the most common chronic condition - Multimorbidity. *JAMA - J. Am. Med. Assoc.* **307**, 2493–2494 (2012).
37. Busija, L., Lim, K., Szoeke, C., Sanders, K. M. & Mccabe, M. P. Do replicable profiles of multimorbidity exist? Systematic review and synthesis. *Eur. J. Epidemiol.* **34**, 1025–1053 (2019).
38. Jensen, A. B. *et al.* Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat. Commun.* **5**, 4022 (2014).
39. Marx, P. *et al.* Comorbidities in the diseasome are more apparent than real: What Bayesian filtering reveals about the comorbidities of depression. *PLoS Comput. Biol.* **13**, e1005487 (2017).
40. Williams, S. A. *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).

FIGURE LEGENDS

Figure 1 Connectivity between incident diseases established based on associated metabolites. The outer ring illustrates the number of metabolites associated with each individual disease - each disease fragment is split to represent associations with at least one other disease (coloured) or associations specific to that disease (grey). Lines across the circle connecting two outcomes illustrate the number of metabolites associated with both outcomes, where line width is proportional to the number of metabolites. Outer ring fragments in white indicate there were no associations with this disease and are proportional to half the size of at least one associated metabolite. Metabolite-disease associations are based on Cox proportional hazard models with age as the underlying time scale adjusting for sex. A p-value <0.001 was considered significant accounting for 28 diseases tested for each metabolite. Graphs were grouped and coloured according to biochemical entities, e.g., the graph *Amino acid* contains only metabolite associations originating from amino acid related compounds. Numbers in brackets indicate: number of uniquely associated metabolites and total number of associated metabolites. AAA = abdominal aortic aneurysms; PAD = peripheral atrial disease; COPD = chronic obstructive pulmonary disease

Figure 2 Brick plot showing the ranking of metabolites based on the number of associated incident endpoints. Metabolite-disease associations are based on Cox proportional hazard models with age as the underlying time scale adjusting for sex. A p-value <0.001 was considered significant accounting for 28 diseases tested for each metabolite. The x-axis displays the rank of each metabolite according to the number of associated metabolites, counting inverse associations as negative numbers to ease representation of the results. The y-axis counts the number of associated metabolites, whereby positive numbers indicate positive associations and negative numbers indicate inverse associations. Colours of each box indicate the associated endpoint. Selected metabolites with multiple associated endpoints have been annotated. *Metabolites were annotated based on *in-silico* prediction. An interactive version of this figure is available on our webserver (https://omicscience.org/apps/mwasdisease/).

Figure 3 Summary of mediation analysis. A) Bar chart showing for each exposure the number of putative mediating metabolites (coloured bar indicating composition of metabolite species) and number of associated incident outcomes (shaded bar). Only exposures with at least one associated incident outcome are listed and have been sorted by the number of outcomes. B) For each metabolite the number of source exposures is plotted against the median proportion mediated by the metabolite. Dot sizes indicate the number of associated outcomes for which the metabolite mediated at least

some percent of the effect of an exposure. C) Detailed listing for the effect estimated to be significantly mediated by X-12117 from the exposures on the left on the risk for a disease listed on the right.

Figure 4 Percentage of each disease acquired during follow-up. Counts are normalized to the total number of diseases each participant developed. Only participants without any of these diseases at baseline were included (N=5,699). COPD = Chronic obstructive pulmonary disease

Figure 5 Metabolites associated with multimorbidity. Odds ratios and 95%-confidence intervals (Cis) from logistic regression analysis with plasma metabolites as the exposure and a binary NCD multimorbidity variable (onset of two or more diseases during follow-up) as the outcome adjusting for age and sex. Metabolites were ordered by association strength and direction (from left to right). Colouring indicates association direction (red – positively; blue – inversely) and statistical significance correcting for multiple testing (darker colours, $p<4.93x10^{-5}$). The size of the dots indicates the number of associated diseases in disease-specific Cox models. *Metabolites were annotated based on *in-silico* prediction.

Figure 6 Variance explained in plasma levels of selected metabolites associated with multimorbidity. Amount of variance explained by risk factors and other continuous traits on selected metabolites which are representative of metabolites associated with incident NCD multimorbidity (see main text). Solid colours indicate positive associations with metabolite levels whereas shading indicates inverse associations. The column on the far right indicates the maximum amount of variance for any metabolite by each risk factor. [1] 1,5-anhydroglucitol (1,5-AG); [2] X - 14662; [3] creatinine; [4] 2-hydroxyhippurate (salicylurate); [5] X - 21364; [6] X - 23291; [7] X - 12063; [8] cotinine; [9] o-cresol sulfate; [10] X - 24293; [11] 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)*; [12] 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*; [13] cholesterol; [14] palmitoyl-linoleoyl-glycerol (16:0/18:2) *; [15] 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*; [16] 1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1); [17] atenolol; [18] glycerol; [19] glucose; [20] N-acetylmethionine; [21] cysteine-glutathione disulfide; [22] retinol (Vitamin A); [23] choline phosphate; [24] serine; [25] N-acetylneuraminate; [26] citrate; [27] gamma-glutamylglutamine; [28] threonate; [29] perfluorooctanesulfonic acid (PFOS); [30] bilirubin (Z,Z); [31] betaine; [32] urate; [33] thyroxine; *Metabolites were annotated based on *in-silico* prediction.

METHODS

Study Cohort

The EPIC-Norfolk study is a cohort of 25,639 middle-aged individuals from the general population of Norfolk in Eastern England[15], which is a component of the European Prospective Investigation into Cancer and Nutrition (EPIC). The EPIC-Norfolk study was approved by the Norfolk Research Ethics Committee (ref. 05/Q0101/191) and all participants gave their written consent before entering the study.

All participants were flagged for mortality at the UK Office of National Statistics, and vital status was ascertained for the entire cohort. Death certificates were coded by trained nosologists according to the International Classification of Diseases (ICD), 10th revision. Hospitalisation data were obtained using National Health Service numbers through linkage with the East Norfolk Health Authority (ENCORE) database, which contains information on all hospital contacts throughout England and Wales. Participants were identified as having experienced an event if the corresponding ICD-10 code was registered on the death certificate (as the underlying cause of death or as a contributing factor), or as the cause of hospitalisation (**Supplemental Table 2**). Since the long-term follow-up of EPIC-Norfolk comprised the ICD-9 and ICD-10 coding system, codes were consolidated. The current study is based on follow-up to 31st March 2016. Information on lifestyle factors and medical history was obtained from questionnaires as has been reported previously[15]. Supplemental Table 2 summarises the methods for all characteristics investigated in the present study.

Metabolite Measurements

We used non-fasted plasma samples stored in liquid nitrogen since baseline in 1993-97 from a total of 11,966 men and women from the EPIC-Norfolk prospective cohort to perform untargeted metabolomic measurements using the Discovery HD4® platform (Metabolon, Inc., Durham, USA). Measurements were undertaken in two sub-cohorts of 5,989 and 5,977 participants, respectively, quasi-randomly selected from the full cohort following the exclusion of a type 2 diabetes case-cohort. We note that comparing effect estimates from Cox models from the sample used in the present study and the type 2 diabetes cohort were strongly correlated (Pearson's r=0.85). In total, 1,015 metabolites were measured in both sub-cohorts, of which 1,014 were included in statistical analyses as they were present in at least 10 cases for at least one of the outcomes under investigation. Those metabolites cover a broad spectrum of chemical entities, including lipids, amino acids or nucleotides, that is, products of human metabolism but also substances of exogenous origin like drugs or markers of nutrition and lifestyle. Due to this broad coverage and the hypothesis-free nature of the approach several metabolites are of yet unknown identity and referred to by an X followed by a unique number.

Plasma samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. Plasma proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

Several types of controls were analysed in concert with the experimental samples: a pool of well-characterized human plasma served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analysed sample, allowed instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability as determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples was 10%. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analysed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analysed using acidic positive ion conditions; however, it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analysed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analysed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MS^n scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z.

Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (*m/z)*, and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis. Library matches for each compound were checked for each sample and corrected if necessary. All named compounds fulfil tier 1 or tier 2 (indicated by a star) criteria according to the metabolomics reporting standards outlined in Sumner et al. [41].

Peaks were quantified using area-under-the-curve. We performed runday normalization to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected in run-day blocks by registering the medians to equal one (1.00) and normalizing each data point proportionately.

Prior to statistical analyses, metabolite levels were transformed using the natural logarithm and values at the tail of the distribution, defined by mean ± 5*SD, were replaced by the respective lower/upper bound. Metabolite measures were then rescaled to a mean of zero and standard deviation of one. Processing steps were performed for each of the two batches separately. To achieve comparable estimates, all continuous cross-sectional traits at baseline (Supplemental Table 2) were processed in the same way as the metabolome data except for log-transformation for most of the traits.

Statistical Analyses

Cox proportional hazard models and multiple testing correction

We first used Cox proportional hazards models to estimate hazard ratios for the association of metabolite levels (log-transformed and standardized) with each incident disease, with age as the underlying time scale adjusting for sex unless otherwise noted. In case of prostate (males), endometrial, ovarian, and breast cancer (females) only participants of that specific sex were included in the analyses. Cox models were constructed separately for each sub-cohort and the associations were meta-analysed using the R package *metafor*. Participants who reported diseases at baseline or who had incident cancer within the first six months of follow-up were excluded from the analyses for that specific disease. All participants who reported a previous diagnosis of cancer at baseline were excluded from all cancer analyses. A modifying effect of sex was tested by inclusion of an interaction term in the Cox models. For each metabolite – endpoint model separately, we excluded participants with missing values in any of the two variables.

We applied a two-stage approach to define first shared and subsequently disease specific associations. To increase power to detect shared associations with rare outcomes, such as stomach cancer, we applied a threshold of p<0.001 (accounting for 28 outcomes per metabolite). We report significant associations for each outcome based on a stringent Bonferroni threshold accounting for the number of metabolites tested (p<0.05/1,014) and declared a metabolite to be specifically associated with a disease, if the association passed the more stringent threshold (p<0.05/1,014) and was not associated with any other outcome at a liberal level of significance (p<0.001).

We used logistic regression models to test for possible misspecifications of time-to-event data. To test whether participants already diseased but yet undiagnosed at baseline might have influenced effect estimates we 1) rerun all Cox models while subsequently excluding participants experiencing any event within the first five years in one-year steps and 2) excluding all participants who have died within the first five years of follow-up (n=469).

Linear regression analysis and variance decomposition for baseline characteristics

We assessed the relevance of clinical risk factors and traits measured at baseline in two ways: 1) we used linear regression models to test for an association between traits as exposure and metabolite levels as outcome adjusting for age, sex, fasting time and, time of blood sampling, and 2) obtaining the variance in metabolite levels explained by each trait using variance partitioning as implemented in the R package *variancePartition*.

Extended Cox proportional hazard models and mediation analyses

We evaluated the effects of confounders in longitudinal analyses using two different approaches. Firstly, following the establishment of metabolite – disease onset and risk factor – metabolite associations, we performed formal mediation analysis assuming a linear dependency among risk factor – metabolite – diseases onset to test for a possible role of metabolites in mediating the association between risk factors and diseases[42]. We used Cox models to identify significant risk factor – disease onset associations in our data ($p<0.01$) and tested only those triangles with consistent association directions along the putative path (n=6,364). We computed the proportion of effect mediated from the risk factor through the metabolite (indirect effect of the risk factor) as the quotient between the indirect and total effect of the risk factor on the disease. An indirect effect with a p-value<$8.8\text{x}10^{-6}$ was considered significant to account for the number of tests. The proposed relationship might not hold true for every tested association and in particular mediation analysis is not suited to distinguish mediation from confounding. However, by using this approach we were able to link and quantify the effects of risk factors on the presented metabolite – disease onset associations. None of the presented significant findings imply causality but from an aetiological perspective such analysis can provide hints on putative disease pathways which would otherwise been missed using a resolute prediction framework. We then used a set of most common exposures as additional covariates in multivariable adjusted Cox models to test for the persistence of associations, including body mass index, waist-to-hip ratio, smoking behaviour, serum uric acid concentrations, total triglycerides, HDL-cholesterol, random glucose, serum alkaline phosphatase concentrations, serum vitamin C concentrations, systolic blood pressure, and plasma fibrinogen concentrations. Due to missing availability of confounder data for some individuals the total number of included individuals included in this analysis dropped to a maximum of 9,427.

Logistic regression models for multimorbidity

We defined NCD multimorbidity as developing two or more ICD-10-coded diseases during follow-up and logistic regression models were used to test for an association between metabolite levels and this binary outcome. To avoid confounding by diseases present at baseline we excluded all participants reporting at least one of the diseases under investigation at baseline, leaving 5699 participants to be included in these analyses. Models were adjusted for age and sex.

We used hierarchical clustering analysis (with complete linkage) to group metabolites based on absolute Pearson correlations as measure of similarity. The number of clusters was determined using silhouette coefficients.

Figures were created using the basic plot functions of R as well as the R package *circlize*. All statistical analyses were done using R version 3.5.1 (R Foundation for statistical computing, Vienna, Austria).

DATA AVAILABILITY

We provide open access to all summary statistics for academic use through an interactive webserver.

EPIC-Norfolk data can be requested by bona fide researchers for specified scientific purposes via the study website (https://www.mrc-epid.cam.ac.uk/research/studies/epic-norfolk/). Data will either be shared through an institutional data sharing agreement or arrangements will be made for analyses to be conducted remotely without the necessity for data transfer.

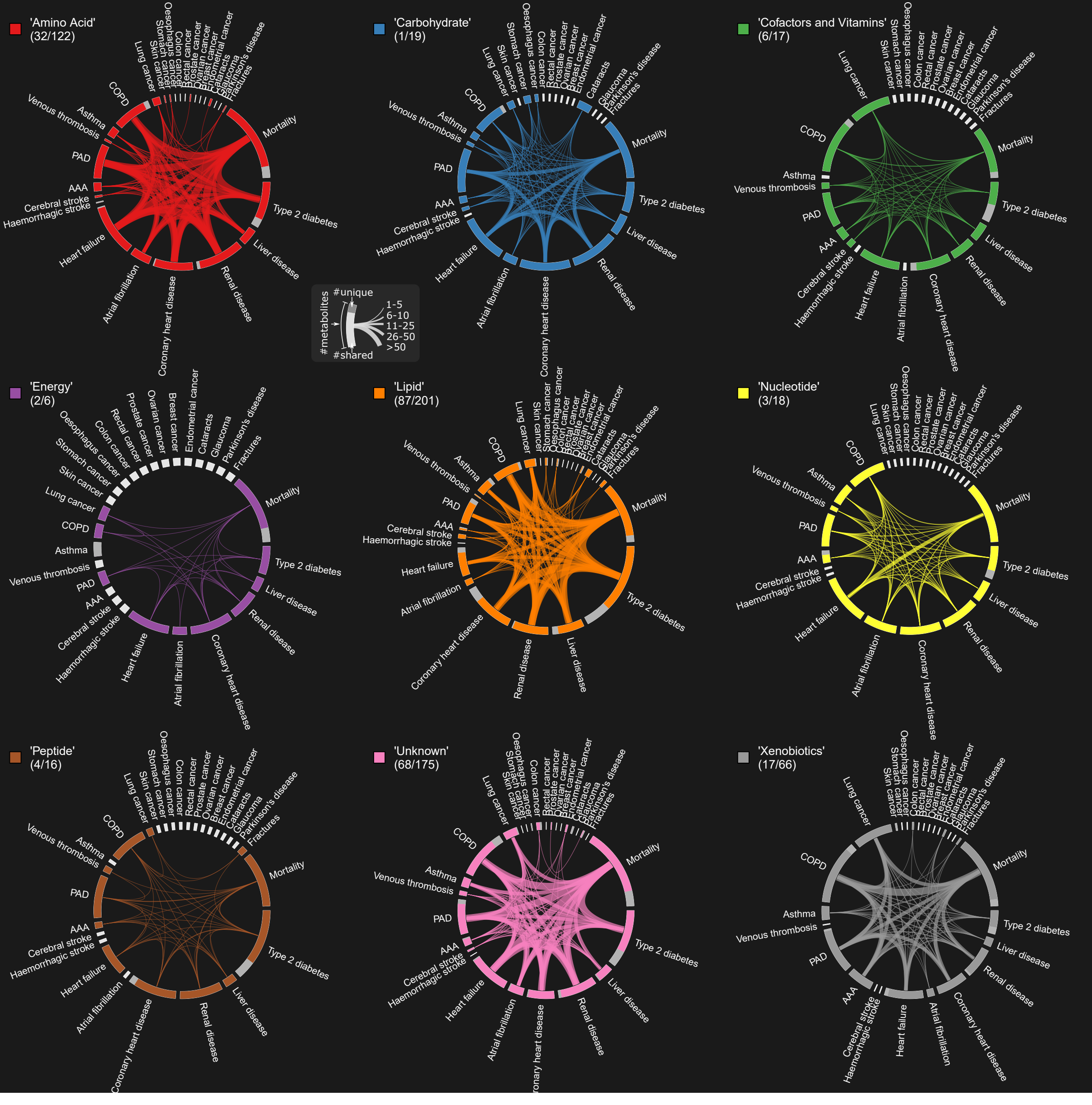
CODE AVAILABILITY

Any code used in the present analysis is freely available to academic researchers upon request from the authors.

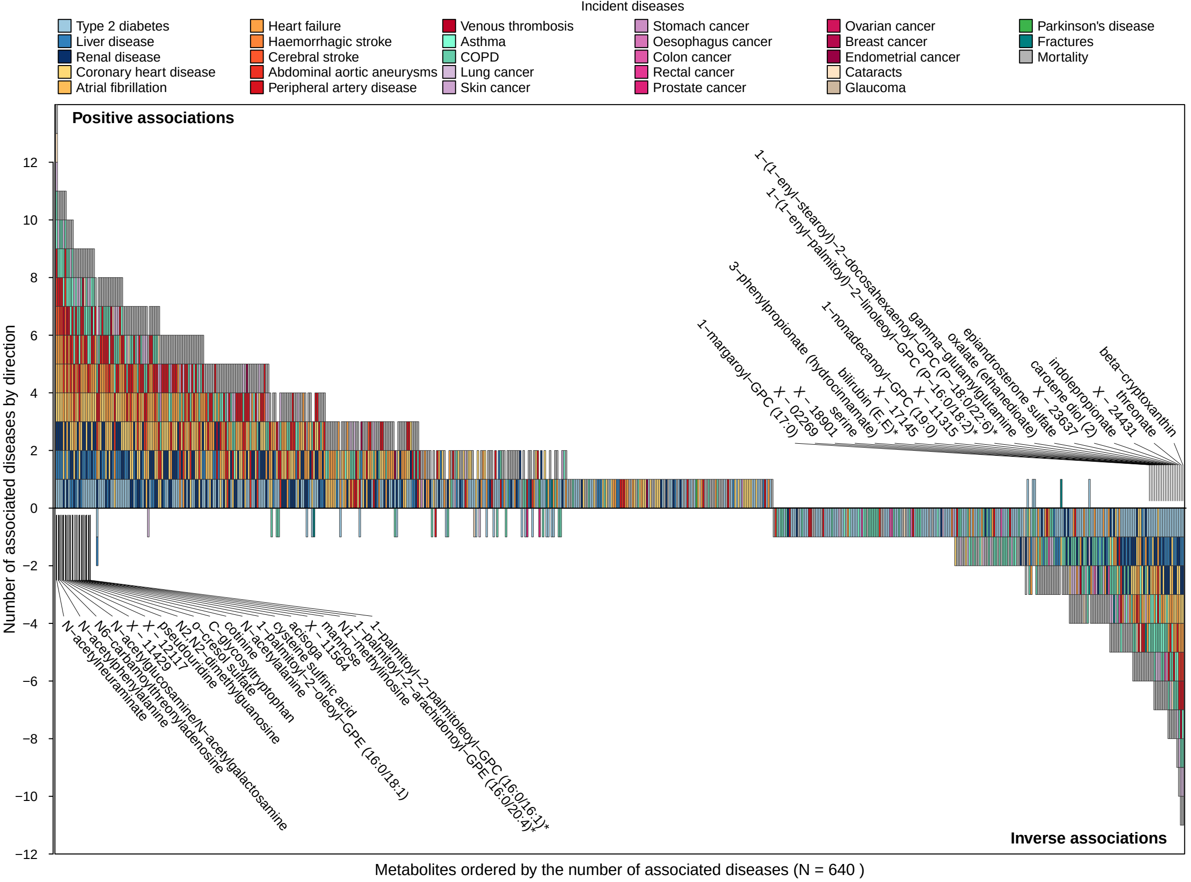
METHODS-ONLY REFERENCES

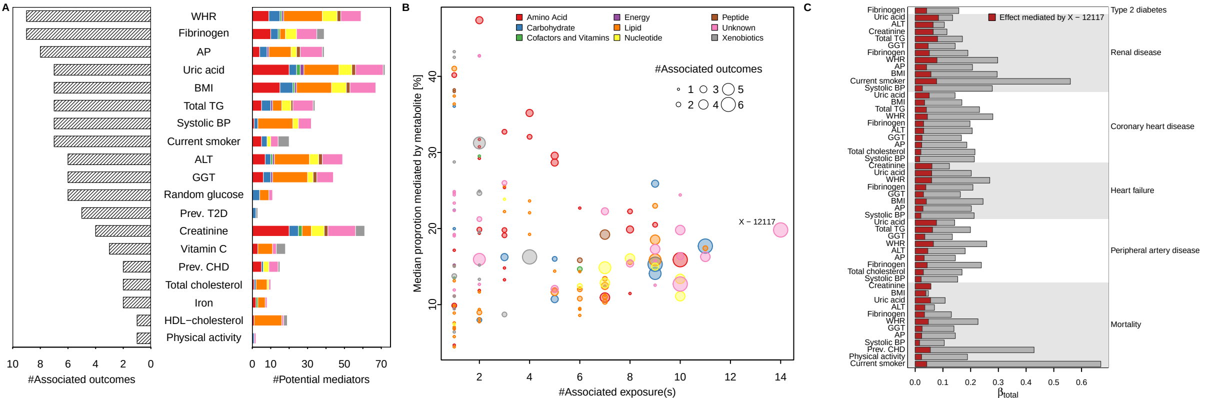
41. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI) NIH Public Access. *Metabolomics* **3**, 211–221 (2007).

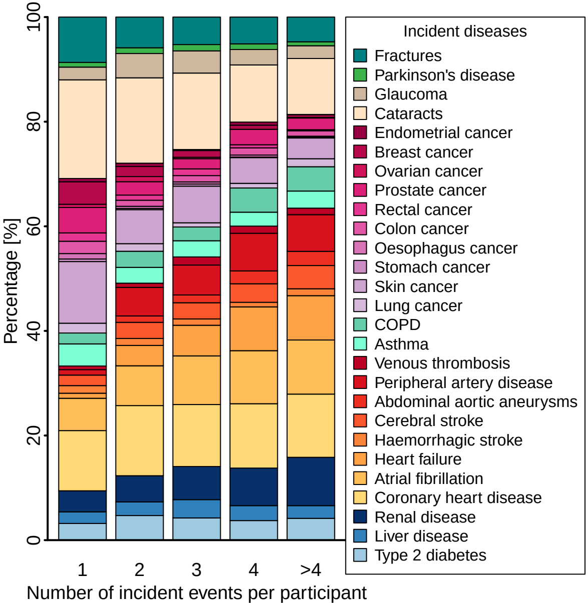
42. Huanga, Y. T., Yangc, H. I., Huang, Y.-T. & Yang, H.-I. Causal Mediation Analysis of Survival Outcome with Multiple Mediators. *Epidemiology* **28**, 370–378 (2017).

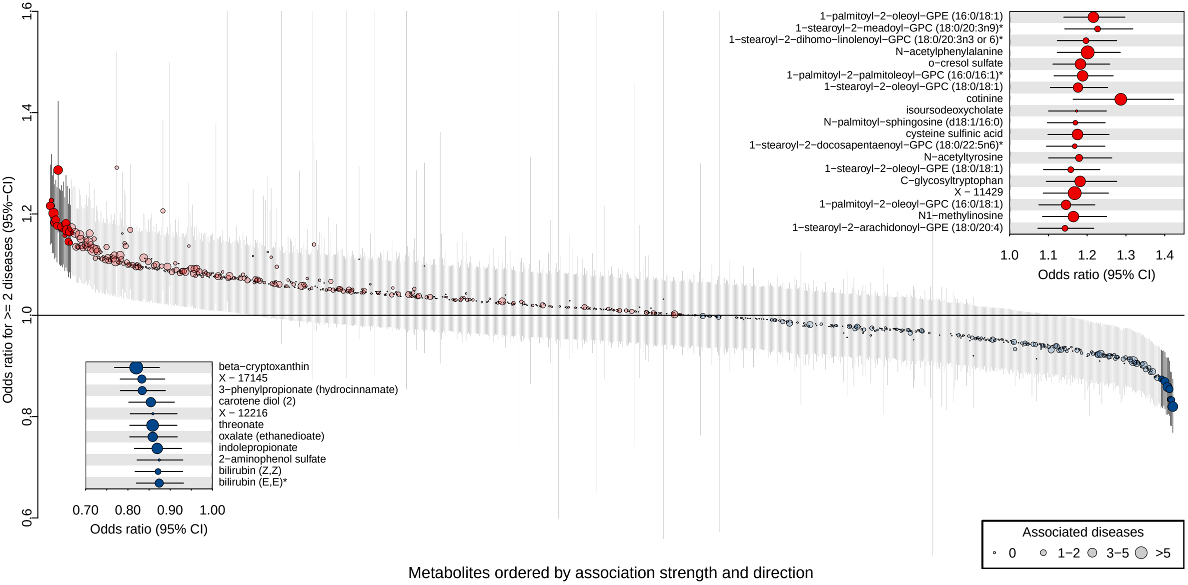


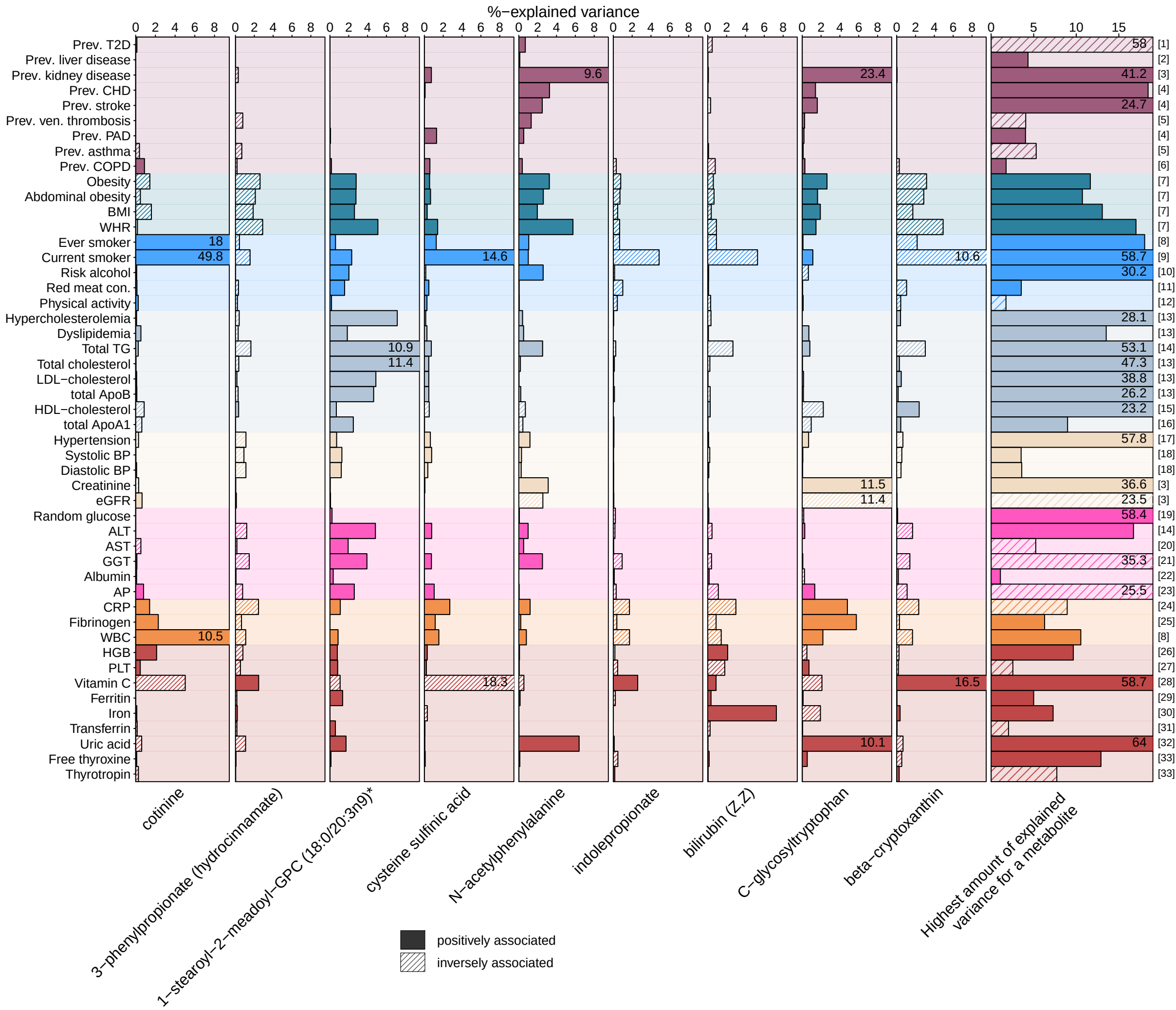
🟥 'Amino Acid'
(32/122)
🟦 'Carbohydrate'
(1/19)
🟩 'Cofactors and Vitamins'
(6/17)
#metabolites
#unique
#shared
1-5
6-10
11-25
26-50
>50
🟪 'Energy'
(2/6)
🟧 'Lipid'
(87/201)
🟨 'Nucleotide'
(3/18)
🟫 'Peptide'
(4/16)
'Unknown'
(68/175)
⬜ 'Xenobiotics'
(17/66)
Lung cancer
Skin cancer
Stomach cancer
Oesophagus cancer
Colon cancer
Rectal cancer
Prostate cancer
Ovarian cancer
Breast cancer
Endometrial cancer
Cataracts
Glaucoma
Parkinson's disease
Fractures
Mortality
Type 2 diabetes
Liver disease
Renal disease
Coronary heart disease
Atrial fibrillation
Heart failure
Haemorrhagic stroke
Cerebral stroke
AAA
PAD
Venous thrombosis
Asthma
COPD

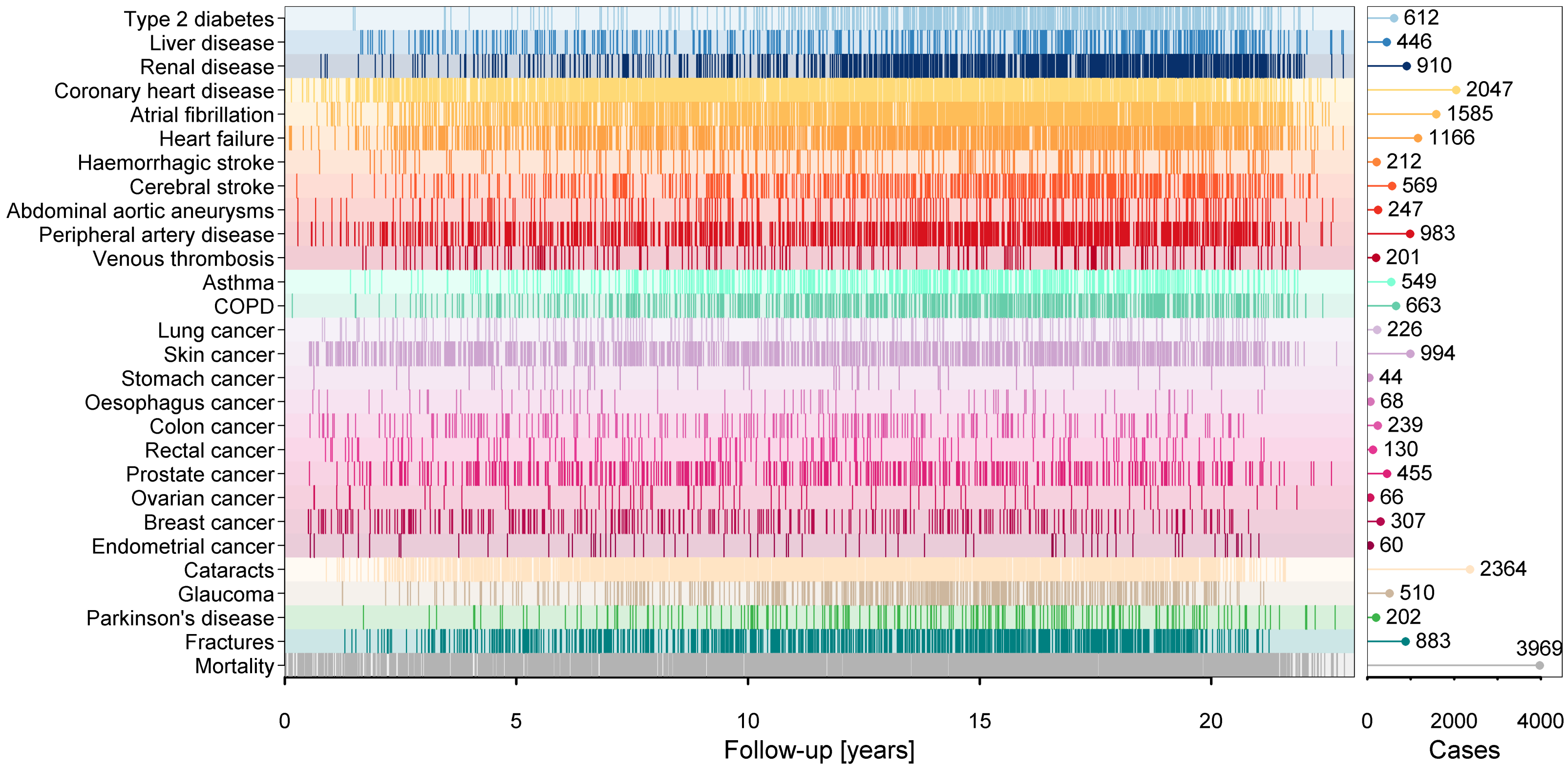

Incident diseases
Type 2 diabetes
Liver disease
Renal disease
Coronary heart disease
Atrial fibrillation
Heart failure
Haemorrhagic stroke
Cerebral stroke
Abdominal aortic aneurysms
Peripheral artery disease
Venous thrombosis
Asthma
COPD
Lung cancer
Skin cancer
Stomach cancer
Oesophagus cancer
Colon cancer
Rectal cancer
Prostate cancer
Ovarian cancer
Breast cancer
Endometrial cancer
Cataracts
Glaucoma
Parkinson's disease
Fractures
Mortality
Positive associations
Inverse associations
Number of associated diseases by direction
12
10
8
6
4
2
0
−2
−4
−6
−8
−10
−12
1−margaroyl−GPC (17:0)
X − 02269
X − 18901
serine
3−phenylpropionate (hydrocinnamate)
bilirubin (E,E)*
X − 17145
1−nonadecanoyl−GPC (19:0)
X − 11315
1−(1−enyl−palmitoyl)−2−linoleoyl−GPC (P−16:0/18:2)*
1−(1−enyl−stearoyl)−2−docosahexaenoyl−GPC (P−18:0/22:6)*
gamma−glutamylglutamine
oxalate (ethanedioate)
epiandrosterone sulfate
X − 23637
carotene diol (2)
indolepropionate
X − 24431
threonate
beta−cryptoxanthin
N−acetylneuraminate
N−acetylphenylalanine
N6−carbamoylthreonyladenosine
N−acetylglucosamine/N−acetylgalactosamine
X − 11429
X − 12117
pseudouridine
N2,N2−dimethylguanosine
o−cresol sulfate
C−glycosyltryptophan
cotinine
N−acetylalanine
1−palmitoyl−2−oleoyl−GPE (16:0/18:1)
cysteine sulfinic acid
acisoga
X − 11564
mannose
N1−methylinosine
1−palmitoyl−2−arachidonoyl−GPE (16:0/20:4)*
1−palmitoyl−2−palmitoleoyl−GPC (16:0/16:1)*
Metabolites ordered by the number of associated diseases (N = 640)

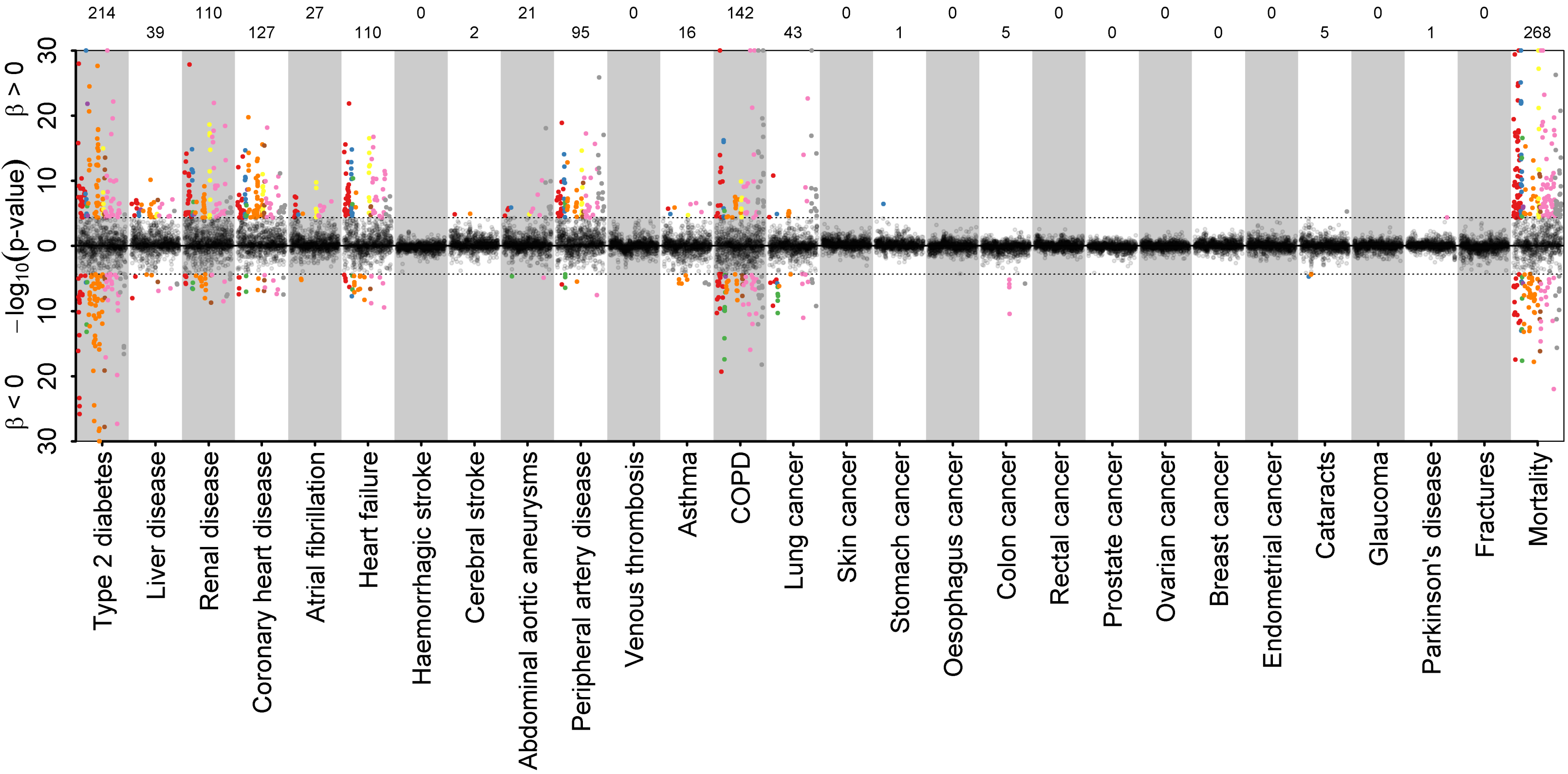

A
WHR
Fibrinogen
AP
Uric acid
BMI
Total TG
Systolic BP
Current smoker
ALT
GGT
Random glucose
Prev. T2D
Creatinine
Vitamin C
Prev. CHD
Total cholesterol
Iron
HDL−cholesterol
Physical activity
10 8 6 4 2 0
#Associated outcomes
0 10 30 50 70
#Potential mediators
B
Amino Acid
Carbohydrate
Cofactors and Vitamins
Energy
Lipid
Nucleotide
Peptide
Unknown
Xenobiotics
#Associated outcomes
1 3 5
2 4 6
X − 12117
Median proprotion mediated by metabolite [%]
10 20 30 40
2 4 6 8 10 12 14
#Associated exposure(s)
C
Effect mediated by X − 12117
Type 2 diabetes
Fibrinogen
Uric acid
ALT
Creatinine
Total TG
GGT
Fibrinogen
WHR
AP
BMI
Current smoker
Systolic BP
Renal disease
Uric acid
BMI
Total TG
WHR
Fibrinogen
ALT
GGT
AP
Total cholesterol
Systolic BP
Coronary heart disease
Creatinine
Uric acid
WHR
Fibrinogen
GGT
BMI
AP
Systolic BP
Heart failure
Uric acid
Total TG
GGT
WHR
ALT
AP
Fibrinogen
Total cholesterol
Systolic BP
Peripheral artery disease
Creatinine
BMI
Uric acid
ALT
Fibrinogen
WHR
GGT
AP
Systolic BP
Prev. CHD
Physical activity
Current smoker
Mortality
0.0 0.1 0.2 0.3 0.4 0.5 0.6
β_{total}

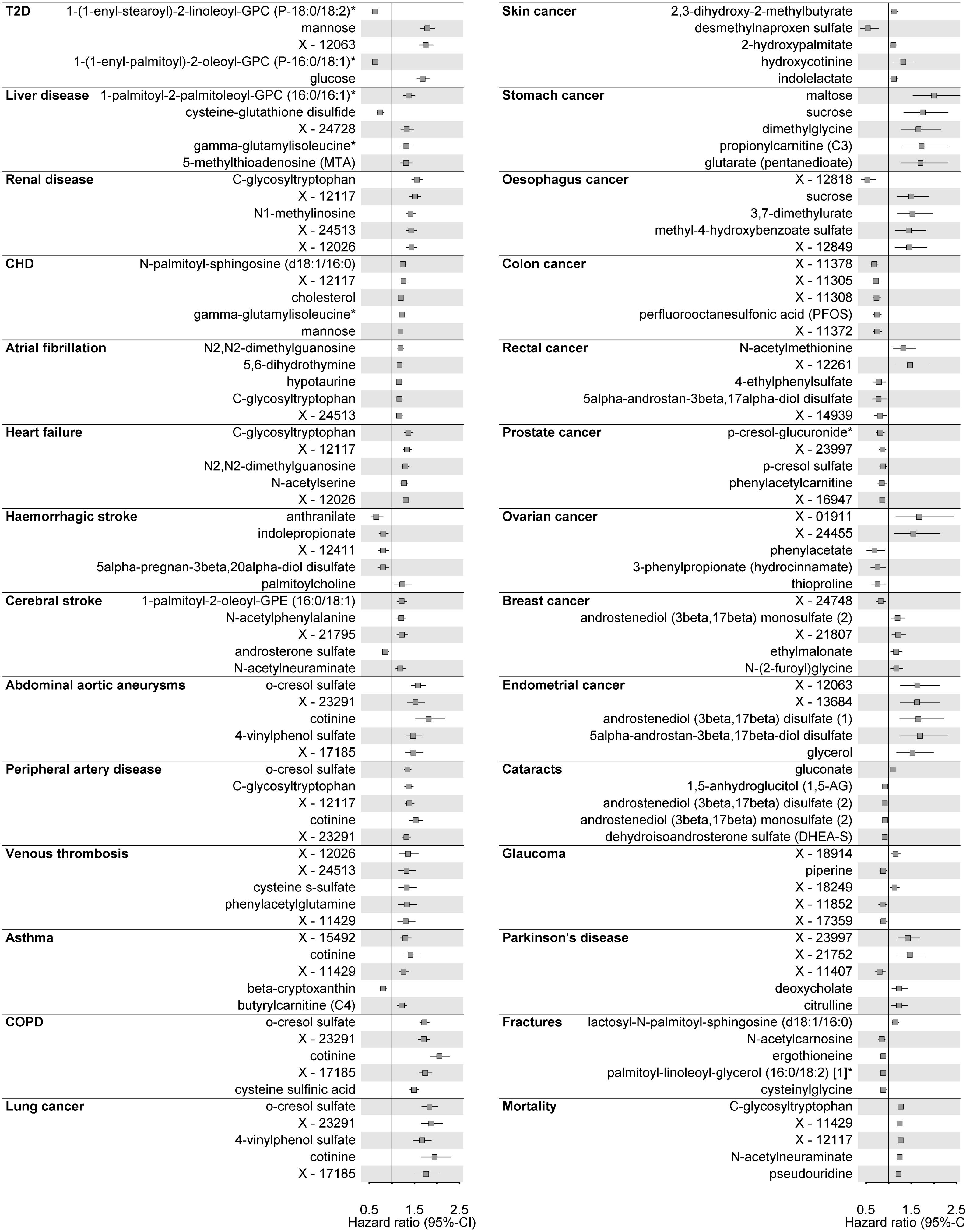

Incident diseases
Fractures
Parkinson's disease
Glaucoma
Cataracts
Endometrial cancer
Breast cancer
Ovarian cancer
Prostate cancer
Rectal cancer
Colon cancer
Oesophagus cancer
Stomach cancer
Skin cancer
Lung cancer
COPD
Asthma
Venous thrombosis
Peripheral artery disease
Abdominal aortic aneurysms
Cerebral stroke
Haemorrhagic stroke
Heart failure
Atrial fibrillation
Coronary heart disease
Renal disease
Liver disease
Type 2 diabetes
Percentage [%]
0
20
40
60
80
100
1
2
3
4
>4
Number of incident events per participant

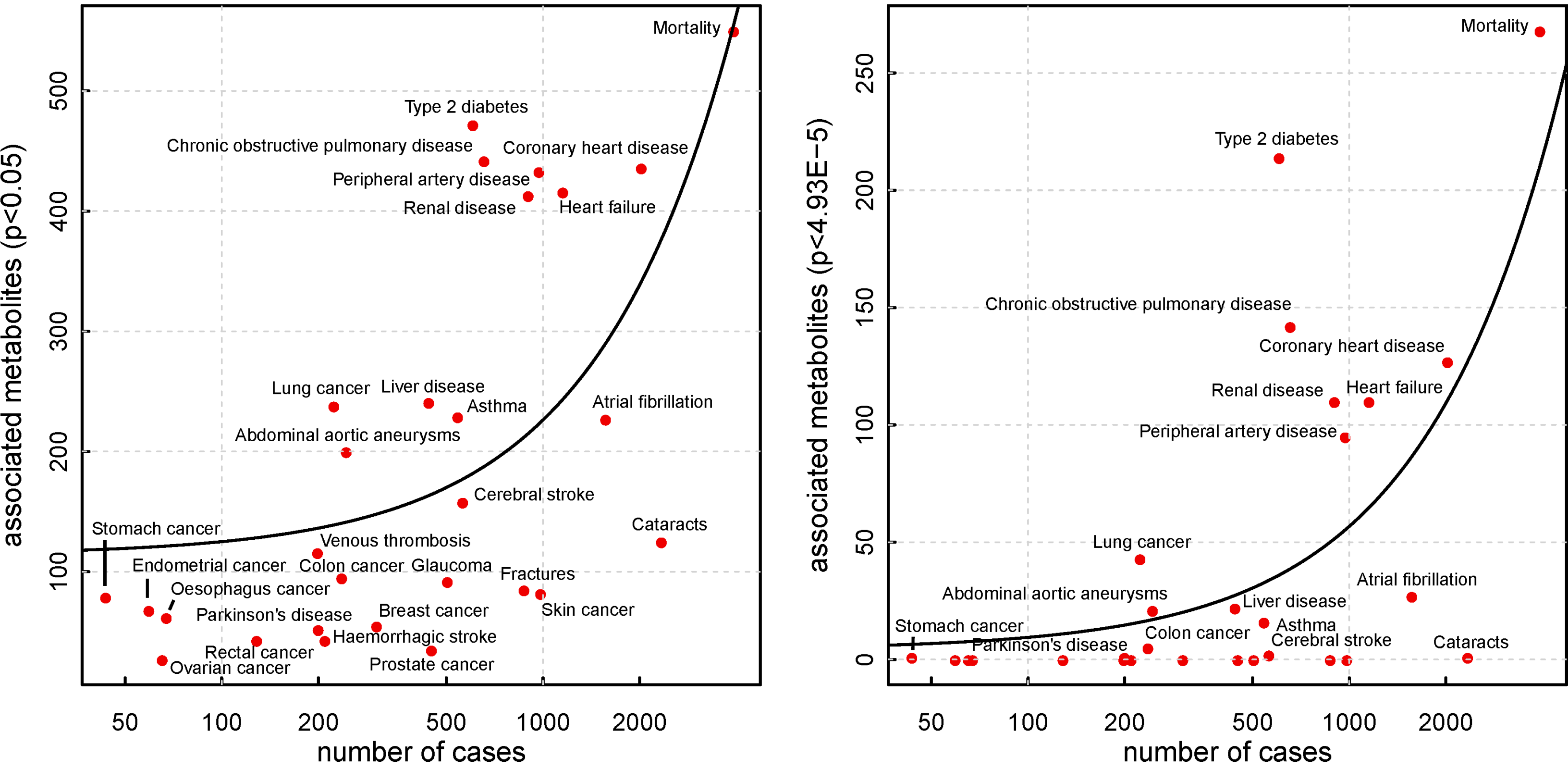

Odds ratio for >= 2 diseases (95%−CI)
1.6
1.4
1.2
1.0
0.8
0.6
Metabolites ordered by association strength and direction
1−palmitoyl−2−oleoyl−GPE (16:0/18:1)
1−stearoyl−2−meadoyl−GPC (18:0/20:3n9)*
1−stearoyl−2−dihomo−linolenoyl−GPC (18:0/20:3n3 or 6)*
N−acetylphenylalanine
o−cresol sulfate
1−palmitoyl−2−palmitoleoyl−GPC (16:0/16:1)*
1−stearoyl−2−oleoyl−GPC (18:0/18:1)
cotinine
isoursodeoxycholate
N−palmitoyl−sphingosine (d18:1/16:0)
cysteine sulfinic acid
1−stearoyl−2−docosapentaenoyl−GPC (18:0/22:5n6)*
N−acetyltyrosine
1−stearoyl−2−oleoyl−GPE (18:0/18:1)
C−glycosyltryptophan
X − 11429
1−palmitoyl−2−oleoyl−GPC (16:0/18:1)
N1−methylinosine
1−stearoyl−2−arachidonoyl−GPE (18:0/20:4)
1.0
1.1
1.2
1.3
1.4
Odds ratio (95% CI)
beta−cryptoxanthin
X − 17145
3−phenylpropionate (hydrocinnamate)
carotene diol (2)
X − 12216
threonate
oxalate (ethanedioate)
indolepropionate
2−aminophenol sulfate
bilirubin (Z,Z)
bilirubin (E,E)*
0.70
0.80
0.90
1.00
Odds ratio (95% CI)
Associated diseases
0
1−2
3−5
>5

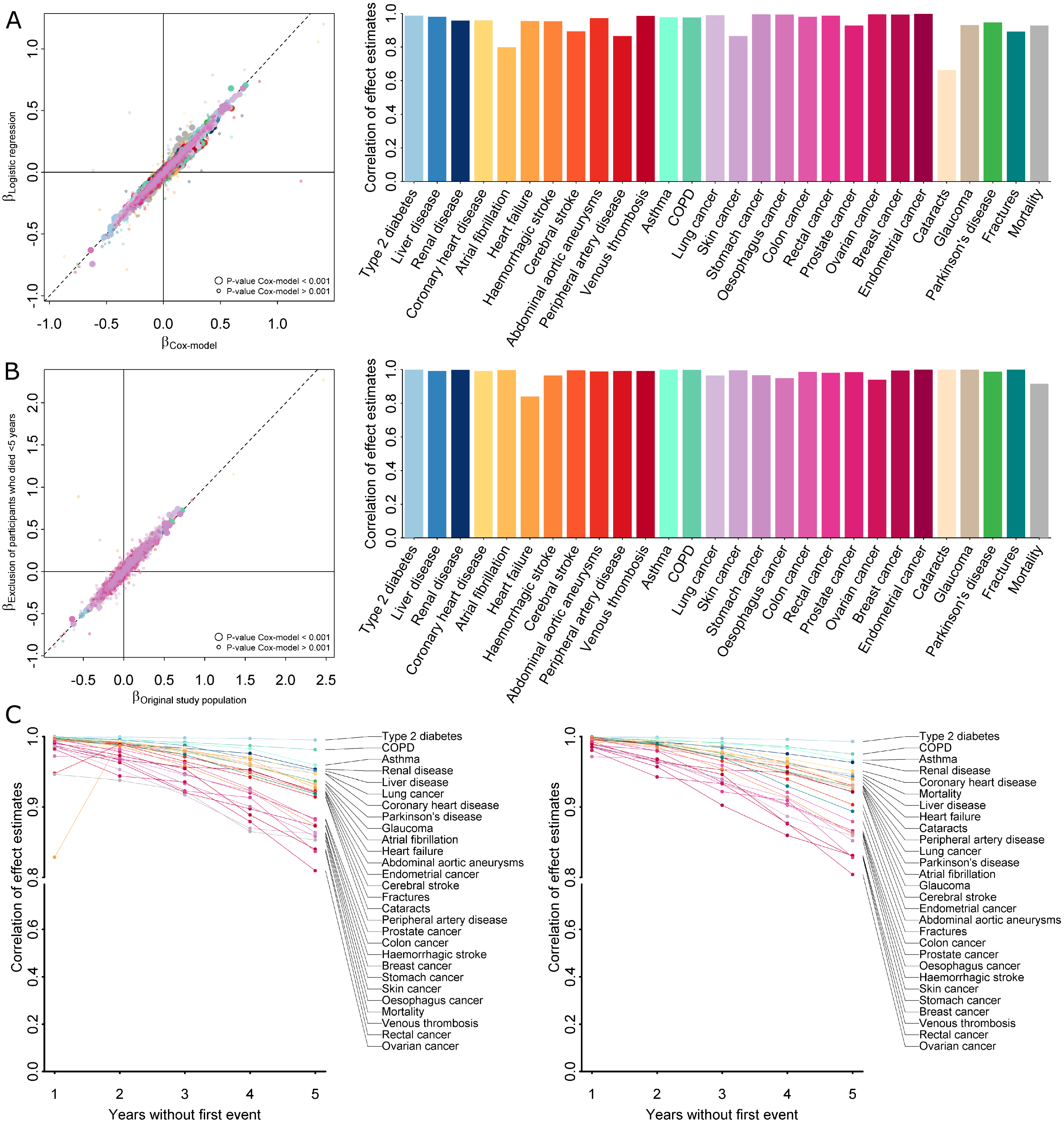

%−explained variance
0
2
4
6
8
0
5
10
15
Prev. T2D
Prev. liver disease
Prev. kidney disease
Prev. CHD
Prev. stroke
Prev. ven. thrombosis
Prev. PAD
Prev. asthma
Prev. COPD
Obesity
Abdominal obesity
BMI
WHR
Ever smoker
Current smoker
Risk alcohol
Red meat con.
Physical activity
Hypercholesterolemia
Dyslipidemia
Total TG
Total cholesterol
LDL−cholesterol
total ApoB
HDL−cholesterol
total ApoA1
Hypertension
Systolic BP
Diastolic BP
Creatinine
eGFR
Random glucose
ALT
AST
GGT
Albumin
AP
CRP
Fibrinogen
WBC
HGB
PLT
Vitamin C
Ferritin
Iron
Transferrin
Uric acid
Free thyroxine
Thyrotropin
18
49.8
10.5
10.9
11.4
14.6
18.3
9.6
23.4
11.5
11.4
10.1
10.6
16.5
58
41.2
24.7
58.7
30.2
28.1
53.1
47.3
38.8
26.2
23.2
57.8
36.6
23.5
58.4
35.3
25.5
58.7
64
[1]
[2]
[3]
[4]
[4]
[5]
[4]
[5]
[6]
[7]
[7]
[7]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[13]
[14]
[13]
[13]
[13]
[15]
[16]
[17]
[18]
[18]
[3]
[3]
[19]
[14]
[20]
[21]
[22]
[23]
[24]
[25]
[8]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[33]
cotinine
3−phenylpropionate (hydrocinnamate)
1−stearoyl−2−meadoyl−GPC (18:0/20:3n9)*
cysteine sulfinic acid
N−acetylphenylalanine
indolepropionate
bilirubin (Z,Z)
C−glycosyltryptophan
beta−cryptoxanthin
Highest amount of explained variance for a metabolite
positively associated
inversely associated

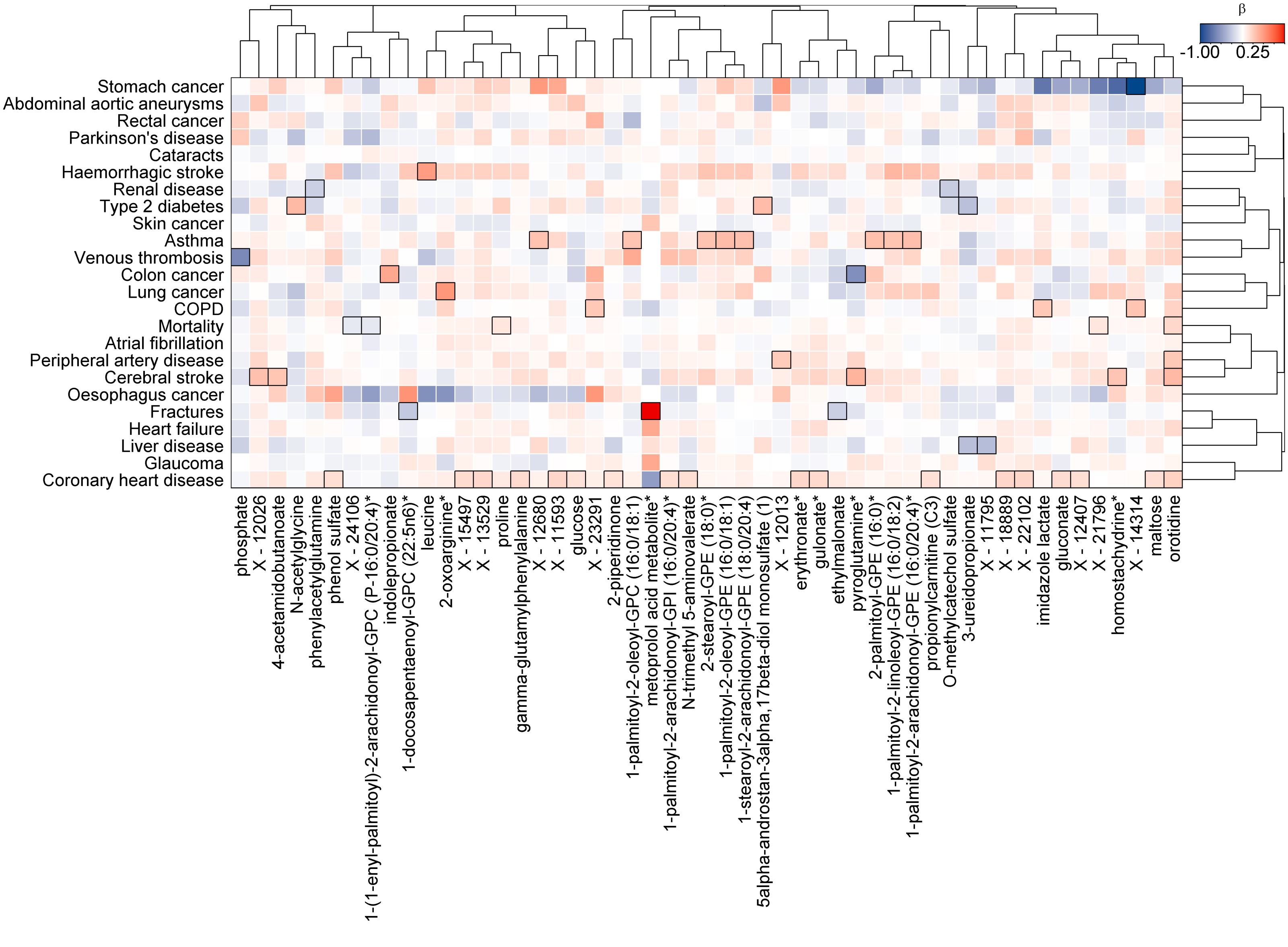

Type 2 diabetes
Liver disease
Renal disease
Coronary heart disease
Atrial fibrillation
Heart failure
Haemorrhagic stroke
Cerebral stroke
Abdominal aortic aneurysms
Peripheral artery disease
Venous thrombosis
Asthma
COPD
Lung cancer
Skin cancer
Stomach cancer
Oesophagus cancer
Colon cancer
Rectal cancer
Prostate cancer
Ovarian cancer
Breast cancer
Endometrial cancer
Cataracts
Glaucoma
Parkinson's disease
Fractures
Mortality
612
446
910
2047
1585
1166
212
569
247
983
201
549
663
226
994
44
68
239
130
455
66
307
60
2364
510
202
883
3969
0
5
10
15
20
Follow-up [years]
0
2000
4000
Cases

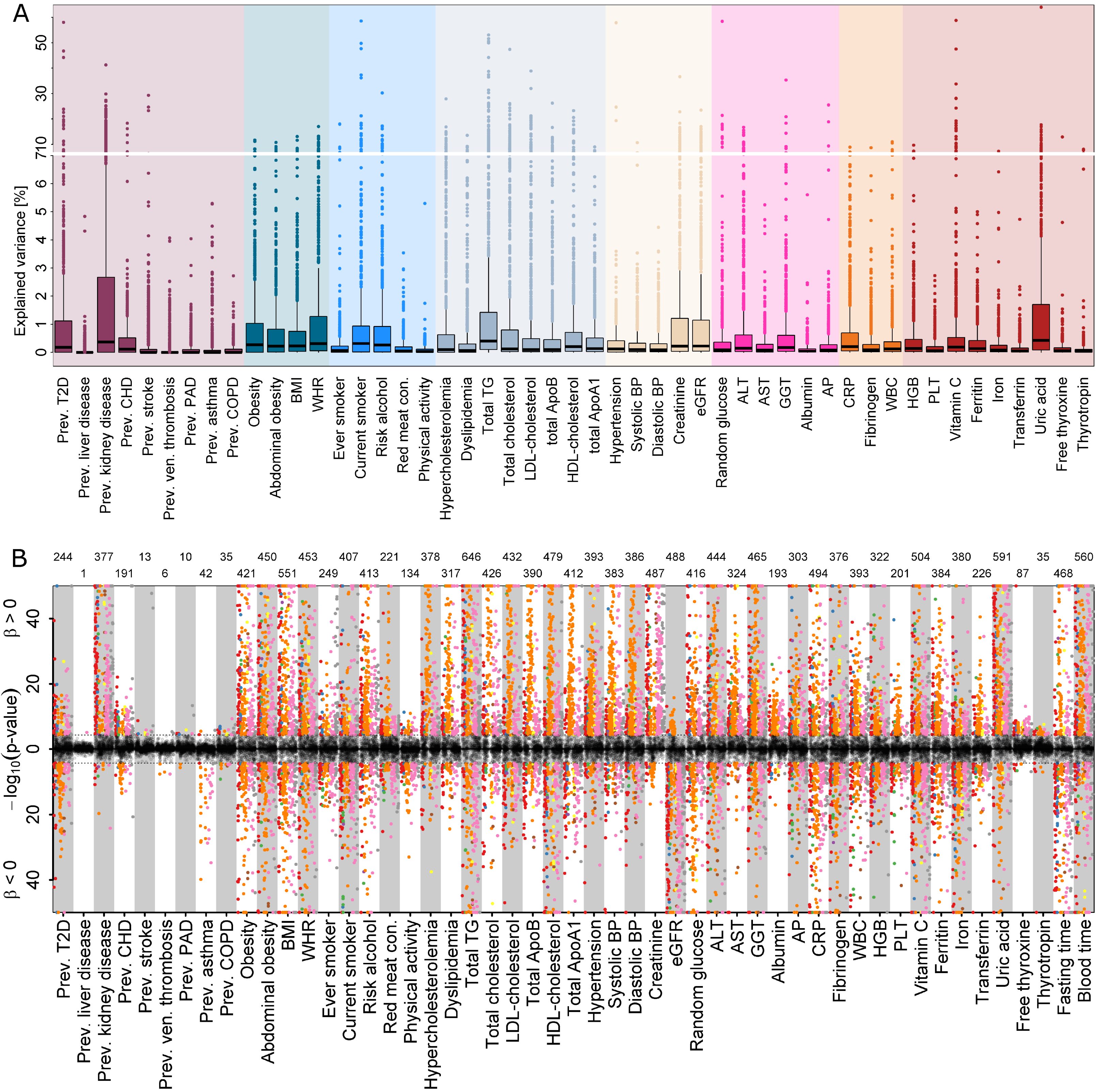

214
39
110
127
27
110
0
2
21
95
0
16
142
43
0
1
0
5
0
0
0
0
0
5
0
1
0
268
β > 0
β < 0
−log$_{10}$(p-value)
30
20
10
0
10
20
30
Type 2 diabetes
Liver disease
Renal disease
Coronary heart disease
Atrial fibrillation
Heart failure
Haemorrhagic stroke
Cerebral stroke
Abdominal aortic aneurysms
Peripheral artery disease
Venous thrombosis
Asthma
COPD
Lung cancer
Skin cancer
Stomach cancer
Oesophagus cancer
Colon cancer
Rectal cancer
Prostate cancer
Ovarian cancer
Breast cancer
Endometrial cancer
Cataracts
Glaucoma
Parkinson's disease
Fractures
Mortality

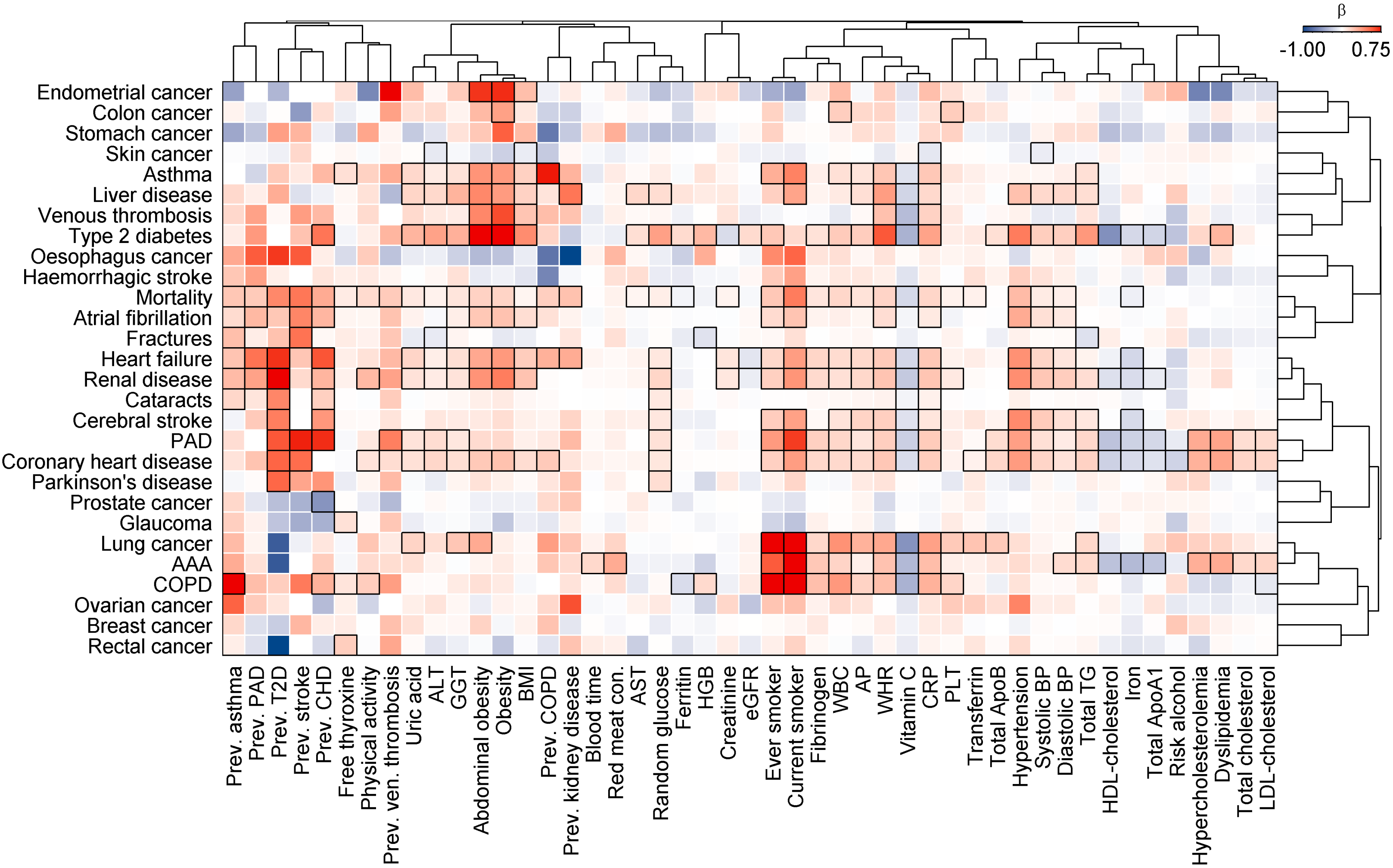

T2D
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*
mannose
X - 12063
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*
glucose
Liver disease
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*
cysteine-glutathione disulfide
X - 24728
gamma-glutamylisoleucine*
5-methylthioadenosine (MTA)
Renal disease
C-glycosyltryptophan
X - 12117
N1-methylinosine
X - 24513
X - 12026
CHD
N-palmitoyl-sphingosine (d18:1/16:0)
X - 12117
cholesterol
gamma-glutamylisoleucine*
mannose
Atrial fibrillation
N2,N2-dimethylguanosine
5,6-dihydrothymine
hypotaurine
C-glycosyltryptophan
X - 24513
Heart failure
C-glycosyltryptophan
X - 12117
N2,N2-dimethylguanosine
N-acetylserine
X - 12026
Haemorrhagic stroke
anthranilate
indolepropionate
X - 12411
5alpha-pregnan-3beta,20alpha-diol disulfate
palmitoylcholine
Cerebral stroke
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)
N-acetylphenylalanine
X - 21795
androsterone sulfate
N-acetylneuraminate
Abdominal aortic aneurysms
o-cresol sulfate
X - 23291
cotinine
4-vinylphenol sulfate
X - 17185
Peripheral artery disease
o-cresol sulfate
C-glycosyltryptophan
X - 12117
cotinine
X - 23291
Venous thrombosis
X - 12026
X - 24513
cysteine s-sulfate
phenylacetylglutamine
X - 11429
Asthma
X - 15492
cotinine
X - 11429
beta-cryptoxanthin
butyrylcarnitine (C4)
COPD
o-cresol sulfate
X - 23291
cotinine
X - 17185
cysteine sulfinic acid
Lung cancer
o-cresol sulfate
X - 23291
4-vinylphenol sulfate
cotinine
X - 17185
0.5
1.5
2.5
Hazard ratio (95%-CI)
Skin cancer
2,3-dihydroxy-2-methylbutyrate
desmethylnaproxen sulfate
2-hydroxypalmitate
hydroxycotinine
indolelactate
Stomach cancer
maltose
sucrose
dimethylglycine
propionylcarnitine (C3)
glutarate (pentanedioate)
Oesophagus cancer
X - 12818
sucrose
3,7-dimethylurate
methyl-4-hydroxybenzoate sulfate
X - 12849
Colon cancer
X - 11378
X - 11305
X - 11308
perfluorooctanesulfonic acid (PFOS)
X - 11372
Rectal cancer
N-acetylmethionine
X - 12261
4-ethylphenylsulfate
5alpha-androstan-3beta,17alpha-diol disulfate
X - 14939
Prostate cancer
p-cresol-glucuronide*
X - 23997
p-cresol sulfate
phenylacetylcarnitine
X - 16947
Ovarian cancer
X - 01911
X - 24455
phenylacetate
3-phenylpropionate (hydrocinnamate)
thioproline
Breast cancer
X - 24748
androstenediol (3beta,17beta) monosulfate (2)
X - 21807
ethylmalonate
N-(2-furoyl)glycine
Endometrial cancer
X - 12063
X - 13684
androstenediol (3beta,17beta) disulfate (1)
5alpha-androstan-3beta,17beta-diol disulfate
glycerol
Cataracts
gluconate
1,5-anhydroglucitol (1,5-AG)
androstenediol (3beta,17beta) disulfate (2)
androstenediol (3beta,17beta) monosulfate (2)
dehydroisoandrosterone sulfate (DHEA-S)
Glaucoma
X - 18914
piperine
X - 18249
X - 11852
X - 17359
Parkinson's disease
X - 23997
X - 21752
X - 11407
deoxycholate
citrulline
Fractures
lactosyl-N-palmitoyl-sphingosine (d18:1/16:0)
N-acetylcarnosine
ergothioneine
palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*
cysteinylglycine
Mortality
C-glycosyltryptophan
X - 11429
X - 12117
N-acetylneuraminate
pseudouridine
0.5
1.5
2.5
Hazard ratio (95%-CI)

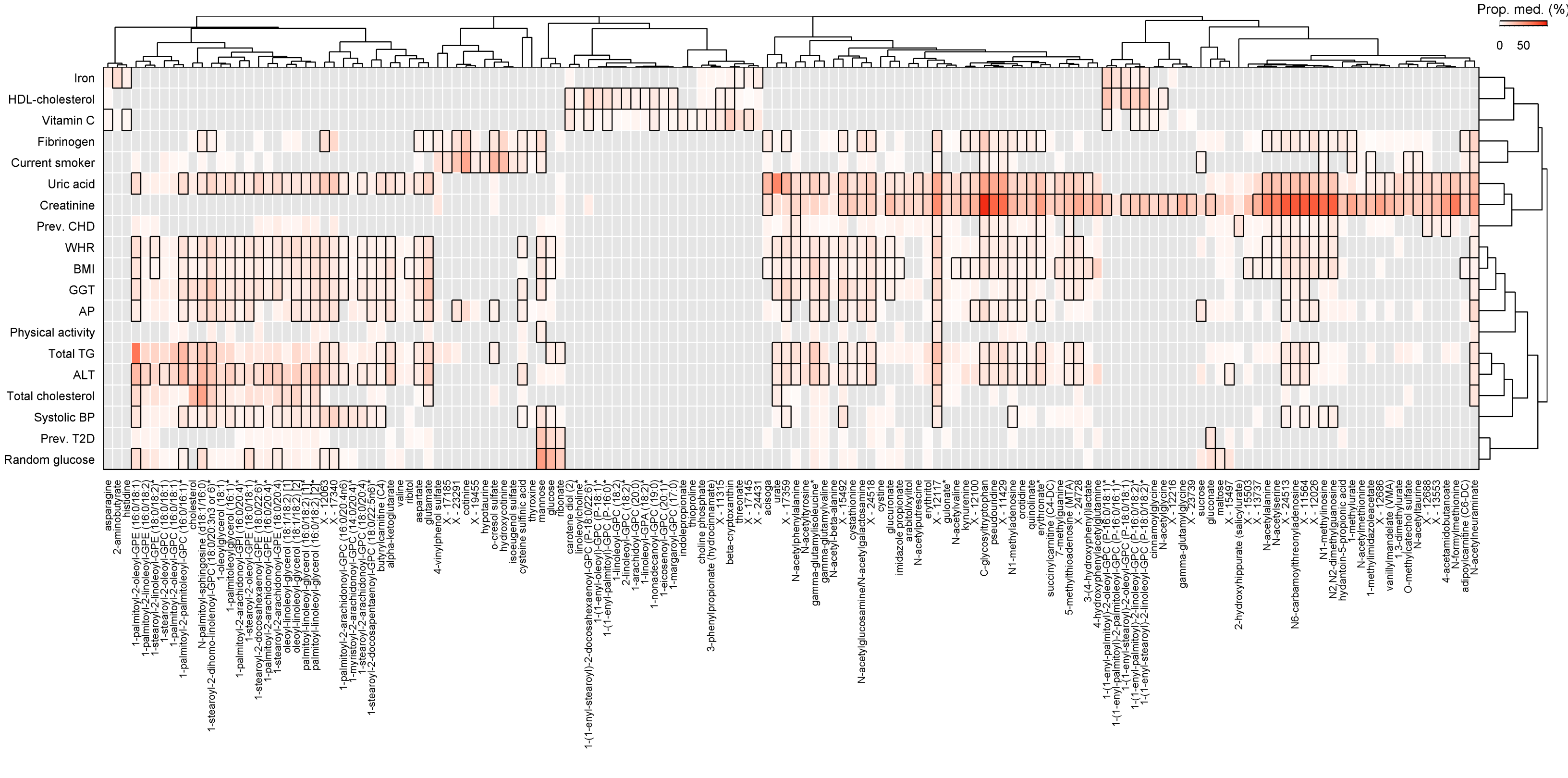

associated metabolites (p<0.05)
100
200
300
400
500
Mortality
Type 2 diabetes
Chronic obstructive pulmonary disease
Coronary heart disease
Peripheral artery disease
Renal disease
Heart failure
Lung cancer
Liver disease
Asthma
Atrial fibrillation
Abdominal aortic aneurysms
Cerebral stroke
Stomach cancer
Cataracts
Venous thrombosis
Endometrial cancer
Colon cancer
Glaucoma
Fractures
Oesophagus cancer
Skin cancer
Parkinson's disease
Breast cancer
Haemorrhagic stroke
Rectal cancer
Ovarian cancer
Prostate cancer
50
100
200
500
1000
2000
number of cases
associated metabolites (p<4.93E−5)
0
50
100
150
200
250
Mortality
Type 2 diabetes
Chronic obstructive pulmonary disease
Coronary heart disease
Renal disease
Heart failure
Peripheral artery disease
Lung cancer
Abdominal aortic aneurysms
Atrial fibrillation
Liver disease
Stomach cancer
Colon cancer
Asthma
Parkinson's disease
Cerebral stroke
Cataracts
50
100
200
500
1000
2000
number of cases

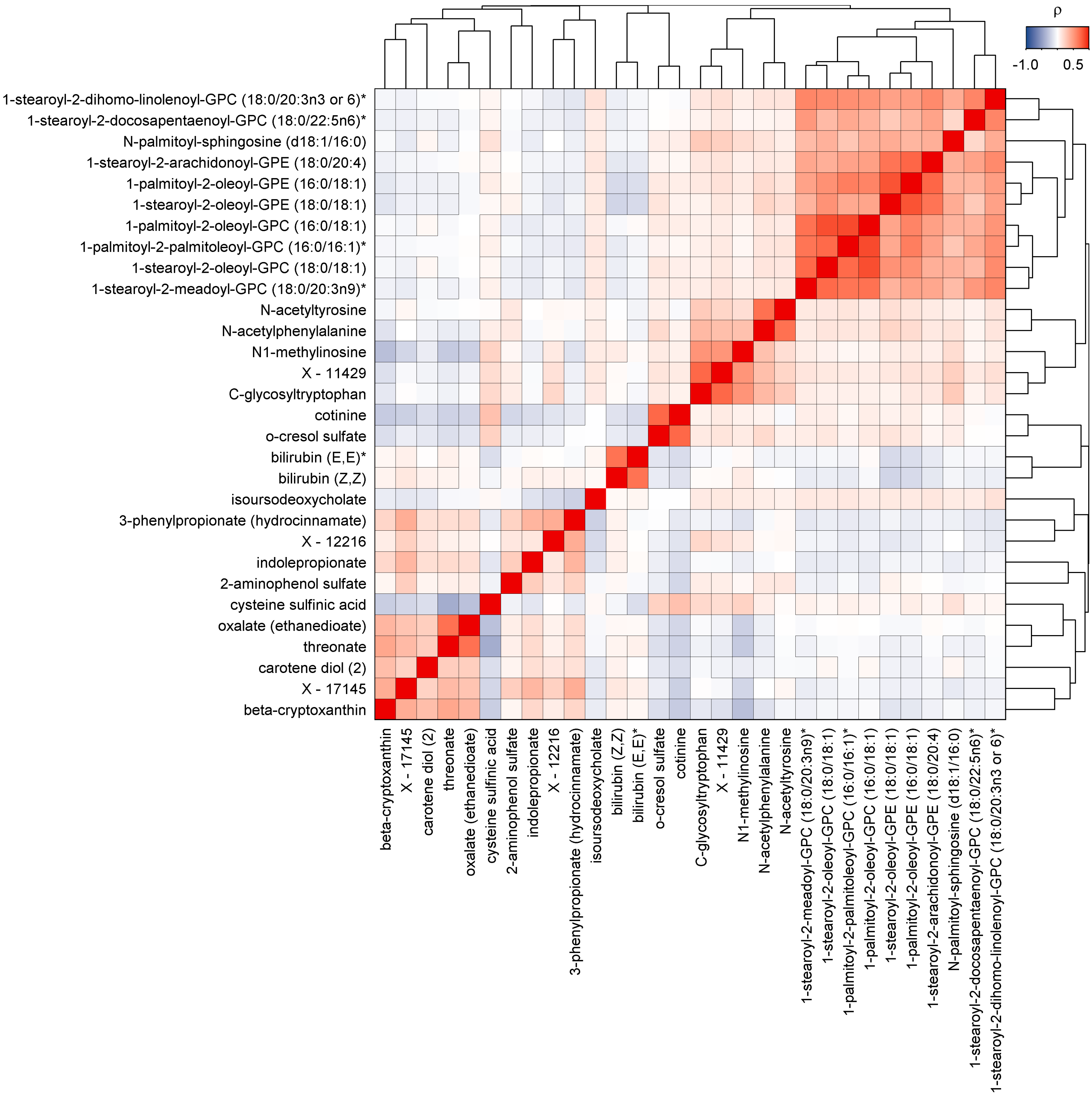

A
β Logistic regression
β Cox-model
-1.0
-0.5
0.0
0.5
1.0
○ P-value Cox-model < 0.001
○ P-value Cox-model > 0.001
Correlation of effect estimates
0.0
0.2
0.4
0.6
0.8
1.0
Type 2 diabetes
Liver disease
Renal disease
Coronary heart disease
Atrial fibrillation
Heart failure
Haemorrhagic stroke
Cerebral stroke
Abdominal aortic aneurysms
Peripheral artery disease
Venous thrombosis
Asthma
COPD
Lung cancer
Skin cancer
Stomach cancer
Oesophagus cancer
Colon cancer
Rectal cancer
Prostate cancer
Ovarian cancer
Breast cancer
Endometrial cancer
Cataracts
Glaucoma
Parkinson's disease
Fractures
Mortality
B
β Exclusion of participants who died <5 years
β Original study population
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
○ P-value Cox-model < 0.001
○ P-value Cox-model > 0.001
Correlation of effect estimates
0.0
0.2
0.4
0.6
0.8
1.0
Type 2 diabetes
Liver disease
Renal disease
Coronary heart disease
Atrial fibrillation
Heart failure
Haemorrhagic stroke
Cerebral stroke
Abdominal aortic aneurysms
Peripheral artery disease
Venous thrombosis
Asthma
COPD
Lung cancer
Skin cancer
Stomach cancer
Oesophagus cancer
Colon cancer
Rectal cancer
Prostate cancer
Ovarian cancer
Breast cancer
Endometrial cancer
Cataracts
Glaucoma
Parkinson's disease
Fractures
Mortality
C
Correlation of effect estimates
0.0
0.2
0.4
0.6
0.8
0.8
0.9
1.0
1
2
3
4
5
Years without first event
Type 2 diabetes
COPD
Asthma
Renal disease
Liver disease
Lung cancer
Coronary heart disease
Parkinson's disease
Glaucoma
Atrial fibrillation
Heart failure
Abdominal aortic aneurysms
Endometrial cancer
Cerebral stroke
Fractures
Cataracts
Peripheral artery disease
Prostate cancer
Colon cancer
Haemorrhagic stroke
Breast cancer
Stomach cancer
Skin cancer
Oesophagus cancer
Mortality
Venous thrombosis
Rectal cancer
Ovarian cancer
Correlation of effect estimates
0.0
0.2
0.4
0.6
0.8
0.8
0.9
1.0
1
2
3
4
5
Years without first event
Type 2 diabetes
COPD
Asthma
Renal disease
Coronary heart disease
Mortality
Liver disease
Heart failure
Cataracts
Peripheral artery disease
Lung cancer
Parkinson's disease
Atrial fibrillation
Glaucoma
Cerebral stroke
Endometrial cancer
Abdominal aortic aneurysms
Fractures
Colon cancer
Prostate cancer
Oesophagus cancer
Haemorrhagic stroke
Skin cancer
Stomach cancer
Breast cancer
Venous thrombosis
Rectal cancer
Ovarian cancer


β
-1.00
0.25
Stomach cancer
Abdominal aortic aneurysms
Rectal cancer
Parkinson's disease
Cataracts
Haemorrhagic stroke
Renal disease
Type 2 diabetes
Skin cancer
Asthma
Venous thrombosis
Colon cancer
Lung cancer
COPD
Mortality
Atrial fibrillation
Peripheral artery disease
Cerebral stroke
Oesophagus cancer
Fractures
Heart failure
Liver disease
Glaucoma
Coronary heart disease
phosphate
X - 12026
4-acetamidobutanoate
N-acetylglycine
phenylacetylglutamine
phenol sulfate
X - 24106
1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*
indolepropionate
1-docosapentaenoyl-GPC (22:5n6)*
leucine
2-oxoarginine*
X - 15497
X - 13529
proline
gamma-glutamylphenylalanine
X - 12680
X - 11593
glucose
X - 23291
2-piperidinone
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)
metoprolol acid metabolite*
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*
N-trimethyl 5-aminovalerate
2-stearoyl-GPE (18:0)*
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)
1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)
5alpha-androstan-3alpha,17beta-diol monosulfate (1)
X - 12013
erythronate*
gulonate*
ethylmalonate
pyroglutamine*
2-palmitoyl-GPE (16:0)*
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*
propionylcarnitine (C3)
O-methylcatechol sulfate
3-ureidopropionate
X - 11795
X - 18889
X - 22102
imidazole lactate
gluconate
X - 12407
X - 21796
homostachydrine*
X - 14314
maltose
orotidine


A
Explained variance [%]
0
1
2
3
4
5
6
710
30
50
Prev. T2D
Prev. liver disease
Prev. kidney disease
Prev. CHD
Prev. stroke
Prev. ven. thrombosis
Prev. PAD
Prev. asthma
Prev. COPD
Obesity
Abdominal obesity
BMI
WHR
Ever smoker
Current smoker
Risk alcohol
Red meat con.
Physical activity
Hypercholesterolemia
Dyslipidemia
Total TG
Total cholesterol
LDL-cholesterol
total ApoB
HDL-cholesterol
total ApoA1
Hypertension
Systolic BP
Diastolic BP
Creatinine
eGFR
Random glucose
ALT
AST
GGT
Albumin
AP
CRP
Fibrinogen
WBC
HGB
PLT
Vitamin C
Ferritin
Iron
Transferrin
Uric acid
Free thyroxine
Thyrotropin
B
244
1
377
191
13
6
10
42
35
450
421
453
551
407
249
221
413
378
134
646
317
432
426
479
390
393
412
386
383
488
487
444
416
465
324
303
193
376
494
322
393
504
201
380
384
591
226
35
87
560
468
β > 0
β < 0
$-\log_{10}$(p-value)
40
20
0
20
40
Prev. T2D
Prev. liver disease
Prev. kidney disease
Prev. CHD
Prev. stroke
Prev. ven. thrombosis
Prev. PAD
Prev. asthma
Prev. COPD
Obesity
Abdominal obesity
BMI
WHR
Ever smoker
Current smoker
Risk alcohol
Red meat con.
Physical activity
Hypercholesterolemia
Dyslipidemia
Total TG
Total cholesterol
LDL-cholesterol
Total ApoB
HDL-cholesterol
Total ApoA1
Hypertension
Systolic BP
Diastolic BP
Creatinine
eGFR
Random glucose
ALT
AST
GGT
Albumin
AP
CRP
Fibrinogen
WBC
HGB
PLT
Vitamin C
Ferritin
Iron
Transferrin
Uric acid
Free thyroxine
Thyrotropin
Fasting time
Blood time


β
-1.00
0.75
Endometrial cancer
Colon cancer
Stomach cancer
Skin cancer
Asthma
Liver disease
Venous thrombosis
Type 2 diabetes
Oesophagus cancer
Haemorrhagic stroke
Mortality
Atrial fibrillation
Fractures
Heart failure
Renal disease
Cataracts
Cerebral stroke
PAD
Coronary heart disease
Parkinson's disease
Prostate cancer
Glaucoma
Lung cancer
AAA
COPD
Ovarian cancer
Breast cancer
Rectal cancer
Prev. asthma
Prev. PAD
Prev. T2D
Prev. stroke
Prev. CHD
Free thyroxine
Physical activity
Prev. ven. thrombosis
Uric acid
ALT
GGT
Abdominal obesity
Obesity
BMI
Prev. COPD
Prev. kidney disease
Blood time
Red meat con.
AST
Random glucose
Ferritin
HGB
Creatinine
eGFR
Ever smoker
Current smoker
Fibrinogen
WBC
AP
WHR
Vitamin C
CRP
PLT
Transferrin
Total ApoB
Hypertension
Systolic BP
Diastolic BP
Total TG
HDL-cholesterol
Iron
Total ApoA1
Risk alcohol
Hypercholesterolemia
Dyslipidemia
Total cholesterol
LDL-cholesterol


Prop. med. (%)
0
50
Iron
HDL-cholesterol
Vitamin C
Fibrinogen
Current smoker
Uric acid
Creatinine
Prev. CHD
WHR
BMI
GGT
AP
Physical activity
Total TG
ALT
Total cholesterol
Systolic BP
Prev. T2D
Random glucose
asparagine
2-aminobutyrate
histidine
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*
1-stearoyl-2-oleoyl-GPC (18:0/18:1)
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*
cholesterol
N-palmitoyl-sphingosine (d18:1/16:0)
1-stearoyl-2-dihomo-linolenoyl-GPC (18:0/20:3n3 or 6)*
1-oleoylglycerol (18:1)
1-palmitoleoylglycerol (16:1)*
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*
1-stearoyl-2-oleoyl-GPE (18:0/18:1)
1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)*
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*
1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]
palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*
palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*
X - 12063
X - 17340
1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6)
1-myristoyl-2-arachidonoyl-GPC (14:0/20:4)*
1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)
1-stearoyl-2-docosapentaenoyl-GPC (18:0/22:5n6)*
butyrylcarnitine (C4)
alpha-ketoglutarate
valine
ribitol
aspartate
glutamate
4-vinylphenol sulfate
X - 17185
X - 23291
cotinine
X - 19455
hypotaurine
o-cresol sulfate
hydroxycotinine
isoeugenol sulfate
cysteine sulfinic acid
thyroxine
mannose
glucose
ribonate
carotene diol (2)
linoleoylcholine*
1-(1-enyl-stearoyl)-2-docosahexaenoyl-GPC (P-18:0/22:6)*
1-(1-enyl-oleoyl)-GPC (P-18:1)*
1-(1-enyl-palmitoyl)-GPC (P-16:0)*
1-linoleoyl-GPC (18:2)
2-linoleoyl-GPC (18:2)*
1-arachidoyl-GPC (20:0)
1-linoleoyl-GPA (18:2)*
1-nonadecanoyl-GPC (19:0)
1-eicosenoyl-GPC (20:1)*
1-margaroyl-GPC (17:0)
indolepropionate
thioproline
choline phosphate
3-phenylpropionate (hydrocinnamate)
X - 11315
beta-cryptoxanthin
threonate
X - 17145
X - 24431
acisoga
urate
X - 17359
N-acetylphenylalanine
N-acetyltyrosine
gamma-glutamylisoleucine*
gamma-glutamylvaline
N-acetyl-beta-alanine
X - 15492
cystathionine
N-acetylglucosamine/N-acetylgalactosamine
X - 24518
cystine
glucuronate
imidazole propionate
arabitol/xylitol
N-acetylputrescine
erythritol
X - 12117
gulonate*
N-acetylvaline
kynurenine
X - 12100
C-glycosyltryptophan
pseudouridine
X - 11429
N1-methyladenosine
orotidine
quinolinate
erythronate*
succinylcarnitine (C4-DC)
7-methylguanine
5-methylthioadenosine (MTA)
X - 24728
3-(4-hydroxyphenyl)lactate
4-hydroxyphenylacetylglutamine
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*
1-(1-enyl-palmitoyl)-2-palmitoleoyl-GPC (P-16:0/16:1)*
1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1)
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*
cinnamoylglycine
N-acetylglycine
X - 12216
gamma-glutamylglycine
X - 23739
sucrose
gluconate
maltose
X - 15497
2-hydroxyhippurate (salicylurate)
X - 15503
X - 13737
N-acetylalanine
N-acetylserine
X - 24513
N6-carbamoylthreonyladenosine
X - 11564
X - 12026
N1-methylinosine
N2,N2-dimethylguanosine
hydantoin-5-propionic acid
1-methylurate
N-acetylmethionine
1-methylimidazoleacetate
X - 12686
vanillylmandelate (VMA)
1,3-dimethylurate
O-methylcatechol sulfate
N-acetyltaurine
X - 12688
X - 13553
4-acetamidobutanoate
N-formylmethionine
adipoylcarnitine (C6-DC)
N-acetylneuraminate


ρ
-1.0
0.5
1-stearoyl-2-dihomo-linolenoyl-GPC (18:0/20:3n3 or 6)*
1-stearoyl-2-docosapentaenoyl-GPC (18:0/22:5n6)*
N-palmitoyl-sphingosine (d18:1/16:0)
1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)
1-stearoyl-2-oleoyl-GPE (18:0/18:1)
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*
1-stearoyl-2-oleoyl-GPC (18:0/18:1)
1-stearoyl-2-meadoyl-GPC (18:0/20:3n9)*
N-acetyltyrosine
N-acetylphenylalanine
N1-methylinosine
X - 11429
C-glycosyltryptophan
cotinine
o-cresol sulfate
bilirubin (E,E)*
bilirubin (Z,Z)
isoursodeoxycholate
3-phenylpropionate (hydrocinnamate)
X - 12216
indolepropionate
2-aminophenol sulfate
cysteine sulfinic acid
oxalate (ethanedioate)
threonate
carotene diol (2)
X - 17145
beta-cryptoxanthin
beta-cryptoxanthin
X - 17145
carotene diol (2)
threonate
oxalate (ethanedioate)
cysteine sulfinic acid
2-aminophenol sulfate
indolepropionate
X - 12216
3-phenylpropionate (hydrocinnamate)
isoursodeoxycholate
bilirubin (Z,Z)
bilirubin (E,E)*
o-cresol sulfate
cotinine
C-glycosyltryptophan
X - 11429
N1-methylinosine
N-acetylphenylalanine
N-acetyltyrosine
1-stearoyl-2-meadoyl-GPC (18:0/20:3n9)*
1-stearoyl-2-oleoyl-GPC (18:0/18:1)
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)
1-stearoyl-2-oleoyl-GPE (18:0/18:1)
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)
1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)
N-palmitoyl-sphingosine (d18:1/16:0)
1-stearoyl-2-docosapentaenoyl-GPC (18:0/22:5n6)*
1-stearoyl-2-dihomo-linolenoyl-GPC (18:0/20:3n3 or 6)*