Metabolomic profiles predict individual multidisease outcomes

Buergel, T.; Steinfeldt, J.; Ruyoga, G.; Pietzner, M.; Bizzarri, D.; Vojinovic, D.; ... ;
Landmesser, U.

OPEN
Metabolomic profiles predict individual multidisease outcomes

Thore Buergel [1,24], Jakob Steinfeldt[2,24], Greg Ruyoga[1], Maik Pietzner[3,4], Daniele Bizzarri [5,6], Dina Vojinovic[7,8], Julius Upmeier zu Belzen [1], Lukas Loock[1], Paul Kittner[1], Lara Christmann[1], Noah Hollmann [1], Henrik Strangalies[1], Jana M. Braunger[1], Benjamin Wild [1], Scott T. Chiesa [9], Joachim Spranger [10,11], Fabian Klostermann[12,13], Erik B. van den Akker [5,6,14], Stella Trompet [15,16], Simon P. Mooijaart[15], Naveed Sattar [17], J. Wouter Jukema [16,18], Birgit Lavrijssen[7,19], Maryam Kavousi[7], Mohsen Ghanbari [7], Mohammad A. Ikram [7], Eline Slagboom[5,20], Mika Kivimaki [21,22], Claudia Langenberg[3,4], John Deanfield[9,25], Roland Eils [1,23,25] ✉ and Ulf Landmesser [2,25]

Risk stratification is critical for the early identification of high-risk individuals and disease prevention. Here we explored the potential of nuclear magnetic resonance (NMR) spectroscopy-derived metabolomic profiles to inform on multidisease risk beyond conventional clinical predictors for the onset of 24 common conditions, including metabolic, vascular, respiratory, musculoskeletal and neurological diseases and cancers. Specifically, we trained a neural network to learn disease-specific metabolomic states from 168 circulating metabolic markers measured in 117,981 participants with ~1.4 million person-years of follow-up from the UK Biobank and validated the model in four independent cohorts. We found metabolomic states to be associated with incident event rates in all the investigated conditions, except breast cancer. For 10-year outcome prediction for 15 endpoints, with and without established metabolic contribution, a combination of age and sex and the metabolomic state equaled or outperformed established predictors. Moreover, metabolomic state added predictive information over comprehensive clinical variables for eight common diseases, including type 2 diabetes, dementia and heart failure. Decision curve analyses showed that predictive improvements translated into clinical utility for a wide range of potential decision thresholds. Taken together, our study demonstrates both the potential and limitations of NMR-derived metabolomic profiles as a multidisease assay to inform on the risk of many common diseases simultaneously.

R isk stratification is central to disease prevention[1,2]. Over the past decade, increasingly complex information on an individual's phenotype has become available beyond conventional demographic and laboratory information. While blood metabolites such as cholesterols are established clinical predictors for cardiovascular disease risk[3], many more have been linked to common disease phenotypes[4–8]. In recent years, studies have moved beyond associations of individual markers by linking metabolomic profiles to aging[9], disease onset[10] and mortality[11], appreciating the human blood metabolome as a direct reflection of the physiological state.

Proton nuclear magnetic resonance (1H-NMR) spectroscopy enables a standardized assessment of a multitude of small circulating molecules in the blood simultaneously. NMR differs from other techniques in metabolomics, such as mass spectrometry, by its virtual absence of batch effects, minimal requirements of expensive reagents and high throughput at comparatively low cost[12]. In the current assay >150 original markers are quantified, including amino and fatty acids and metabolites related to carbohydrate metabolism and fluid balance, partly overlapping with conventional clinical predictors including glucose, albumin and creatinine[13–15]. Further, the

assay has a high resolution of lipoprotein particles, measuring their components, sizes and concentrations[13,14]. This high-throughput NMR metabolomics platform has been explored in multiple studies investigating all-cause mortality[11,16], cardiovascular disease[13,17], type 2 diabetes (T2D)[18,19], Alzheimer's disease[8] and COVID-19 (ref. [20]). Importantly, recent work has indicated a broad metabolic basis across diseases, suggesting a shared etiology[21]. This systemic information contained in metabolomic profiles has been insufficiently considered in the risk prediction of common diseases.

Here we exploited the potential of NMR-based blood profiling as a single-domain assay to simultaneously predict multidisease onset. We developed, trained and validated a deep residual multitask neural network to simultaneously learn disease-specific metabolomic states for 24 conditions, including common metabolic, vascular, respiratory, musculoskeletal and neurological disorders and cancers (Fig. 1). The scalar metabolomic states, contained in a 24-dimensional vector, were derived from 168 circulating metabolomic markers measured in ~120.000 individuals in the UK Biobank population cohort[22]. We extensively investigated the learned metabolomic states by integrating them in Cox proportional hazard (CPH) models[23], modeling the risk for individual endpoints and demonstrating that information gained through NMR metabolomic profiling is additive to known clinical predictors. Moreover, we externally validated the metabolomic states in four independent cohorts, the Whitehall II cohort[24], the Rotterdam Study[25], the Leiden Longevity Study[26] and the PROspective Study of Pravastatin in the Elderly at Risk[27] (Fig. 1c), and investigated their clinical utility.

Results

Study population and the metabolomic state model. Based on the UK Biobank cohort[22,28], we derived an integrated metabolomic state capturing information on incident disease risk in a general population sample (Fig. 1a,b). We extracted clinical predictors and disease endpoints for 117,981 individuals with serum NMR profiling at the time of cohort recruitment (Methods and Supplementary Tables 1–3). The study population had a median age of 58 years (interquartile range (IQR) 50, 63), of whom 54.2% were female, 11% current smokers and 5.2% diagnosed with T2D (Table 1). Median body mass index (BMI) was 26.8 (IQR 24.2, 29.9), systolic blood pressure was 136 mmHg (IQR 124, 149), total cholesterol was 5.65 mmol l^{-1} (IQR 4.90, 6.42) and glucose was 4.93 mmol l^{-1} (IQR 4.60, 5.32). Median follow-up was 12.2 years with ~1,435,340 overall person-years. To maximize the generalizability and transferability of our results, we partitioned the data spatially by the 22 recruitment centers. For each center, all individuals from a single center were retained for testing of models that were trained on individuals pooled from the 21 remaining recruitment centers and then randomly split into training and validation sets to develop the models. After model selection on the validation datasets and obtaining the selected models' final predictions on the individual test sets, test set predictions were aggregated for downstream analysis (Fig. 1b).

We externally validated disease-specific metabolomic states in four independent cohorts analyzed with the same 1H-NMR metabolomics assay, the Whitehall II cohort[24], and three independent cohorts of the BBMRI-NL consortium (Fig. 1c). The Whitehall II cohort[24] is an ongoing prospective cohort study, including metabolomics for 6,197 participants aged 44–69 years. The Rotterdam Study is a prospective, population-based cohort study among individuals living in the Ommoord district in the city of Rotterdam (the Netherlands)[25], offering metabolomics for 2,949 participants with a median age of 74 years (IQR 70–79). The Leiden Longevity PAROFF Study (LLS)[26] comprises offspring and spouses of long-lived individuals, with metabolomics available for 1,655 individuals with a mean age of 59 years (IQR 54–63). Finally, the PROspective Study of Pravastatin in the Elderly at Risk (PROSPER) is a clinical trial investigating pravastatin effects[27], of which 960 samples with a

median age of 76 years (IQR 73–78) are included in the BBMRI-NL platform. Detailed characteristics of the four replication cohorts are presented in Supplementary Data and Supplementary Table 4.

The metabolomic state model is a multitask residual neural network trained on the entire set of 168 original metabolomic markers to model the integrative metabolomic state for all 24 endpoints simultaneously (Fig. 1b, Extended Data Fig. 1 and Metabolomic state model). This allowed us to leverage the shared metabolite profiles while retaining flexibility in fitting endpoint-specific variations, outperforming endpoint-specific linear models and linear models on principal components (Extended Data Fig. 2).

To test whether multidisease states could be equally informative from readily accessible information from study participants at baseline, we investigated three different scenarios with increasingly comprehensive predictor sets. First, we considered age and sex only, both highly predictive for common diseases and available at no cost. Second, we investigated cardiovascular predictors from well-validated primary prevention scores, the American Heart Association (ASCVD)[3], which are easily accessible at minimal cost and are predictive beyond cardiovascular disease, including neurological and neoplastic conditions[29–31]. Third, we extended these predictors with a comprehensive set of clinical predictors beyond what is typically available in primary care. These included >30 predictors with information on lifestyle factors, physical measurements and laboratory values, as well as further validated disease-specific predictors from FINDRISC[32] (T2D) and CAIDE[33] (dementia) scores (Fig. 1d and Supplementary Table 2).

Metabolomic state stratifies the risk of disease onset. A critical component of prevention is identification of individuals at high risk of developing a disease, often at an early subclinical stage. To investigate whether the NMR-derived metabolomic state informs disease risk, we assessed the link with incident event rates in the observation period (Fig. 2a). To allow comparison between the endpoints despite the large differences in event rates (Supplementary Table 7; for example, Parkinson's disease, 0.6%; major adverse cardiac event (MACE), 8.7%), we also calculated the observed event rate ratio between individuals in the top and bottom 10% of metabolomic states (Fig. 2 and Supplementary Table 7) with 95% confidence intervals (CIs).

We observed increasing event rates over metabolomic state percentiles for all 24 investigated diseases, except breast cancer. For 15 of the 24 diseases, the top 10% of the metabolomic state corresponded to a rate more than fivefold higher compared with the bottom 10%. For conditions such as T2D (top 10%, 21.87%; bottom 10%, 0.36%; odds ratio (OR) 61.45, 95% CI 47.00, 86.12), abdominal aortic aneurysm (AAA) (top 10%, 2.46%; bottom 10%, 0.18%; OR 14.1, 95% CI 9.93, 24.45) and heart failure (top 10%, 10.80%; bottom 10%, 0.96%; OR 11.27, 95% CI 9.43, 13.50) the ratio was >10. Ratios for most other diseases were lower—for example, cerebral stroke 9.66 (95% CI 7.64, 12.14), MACE 9.25 (95% CI 8.12, 10.53), atrial fibrillation 8.13 (95% CI 6.95, 9.37), all-cause dementia 6.39 (95% CI 5.40, 8.09) or chronic obstructive pulmonary disease (COPD) 4.98 (95% CI 4.37, 5.62). In contrast, we observed much smaller ratios for some diseases—for example, glaucoma (top 10%, 3.47%; bottom 10%, 1.57%; OR 2.19, 95% CI 1.91, 2.62) or asthma (top 10%, 5.52%; bottom 10%, 2.48%; OR 2.22, 95% CI 2.01, 2.57), thus suggesting less information contained in the respective metabolomic states. In summary, the disease-specific metabolomic state stratified risk trajectories for all investigated endpoints except breast cancer (Fig. 2b), separating the rates of cumulative events most notably for T2D, renal disease and heart failure but also, to a much lesser extent, for glaucoma or asthma.

Information is shared with clinical predictors. Many clinical predictors are readily available in primary care and commonly used to

Fig. 1 | Study overview. a, To learn metabolomic states from circulating blood metabolites, the eligible UK Biobank population (with NMR blood metabolomics and valid consent) was split into training, validation and test sets with 22-fold nested cross-validation based on the assigned UK Biobank assessment center. **b**, For each of the 22 partitions, the metabolomic state model was trained on the 168 metabolomic markers to predict metabolomic risk against 24 common disease endpoints. Subsequently, for each endpoint, CPH models were developed on the metabolomic state in combination with sets of commonly available clinical predictors to model disease risk. Predictions of the CPH model on the test set were aggregated for downstream analysis. **c**, The metabolomic state model was externally validated in four independent cohorts—the Whitehall II cohort and three from the BBMRI-NL consortium: the Rotterdam Study, the Leiden Longevity Study and the PROSPER cohort. **d**, In this study we consider clinical predictors from scores commonly applied in primary prevention. We additionally integrate variables into a comprehensive predictor set (PANEL) to investigate overlapping information with the metabolomic state. FH, family history.

Table 1 | The study population

Characteristic	Male, $n = 54{,}078$[a]	Female, $n = 63{,}903$[a]	Overall, $n = 117{,}981$[a]
Age at recruitment	58 (50, 64)	57 (50, 63)	58 (50, 63)
Education years	15.00 (11.00, 15.00)	13.00 (11.00, 15.00)	13.00 (11.00, 15.00)
Current smoker	6,724 (12%)	5,747 (9.0%)	12,471 (11%)
Daily alcohol intake	13,651 (25%)	10,191 (16%)	23,842 (20%)
Daily moderate to vigorous physical activity	50 (15, 105)	45 (10, 90)	45 (10, 90)
Daily healthy food	52,974 (98%)	63,290 (99%)	116,264 (99%)
Family history of diabetes	8,827 (16%)	11,266 (18%)	20,093 (17%)
T2D	3,882 (7.2%)	2,295 (3.6%)	6,177 (5.2%)
Weight (kg)	84 (76, 94)	69 (62, 79)	76 (66, 88)
Standing height (cm)	176 (171, 180)	162 (158, 167)	168 (162, 175)
BMI	27.3 (25.0, 30.1)	26.1 (23.5, 29.7)	26.8 (24.2, 29.9)
Waist/hip ratio	0.93 (0.89, 0.98)	0.81 (0.77, 0.86)	0.87 (0.80, 0.94)
Waist circumference (cm)	96 (89, 103)	83 (76, 92)	90 (80, 99)
Systolic blood pressure (mmHg)	139 (128, 152)	133 (121, 147)	136 (124, 149)
Total cholesterol (mmol l^{-1})	5.45 (4.70, 6.21)	5.80 (5.07, 6.58)	5.65 (4.90, 6.42)
LDL cholesterol (mmol l^{-1})	3.46 (2.87, 4.05)	3.56 (3.00, 4.17)	3.52 (2.94, 4.12)
HDL cholesterol (mmol l^{-1})	1.24 (1.06, 1.45)	1.55 (1.32, 1.82)	1.40 (1.17, 1.67)
Triglycerides (mmol l^{-1})	1.69 (1.18, 2.44)	1.33 (0.96, 1.89)	1.48 (1.04, 2.14)
Glucose (mmol l^{-1})	4.96 (4.61, 5.37)	4.91 (4.59, 5.28)	4.93 (4.60, 5.32)
Glycated hemoglobin (%)	35.3 (32.8, 38.1)	35.2 (32.7, 37.7)	35.2 (32.8, 37.9)
Creatinine (umol l^{-1})	80 (72, 88)	63 (57, 70)	70 (61, 81)
Cystatin C (mg l^{-1})	0.92 (0.84, 1.01)	0.86 (0.78, 0.95)	0.88 (0.80, 0.98)
Urea (mmol l^{-1})	5.45 (4.68, 6.33)	5.10 (4.33, 5.95)	5.26 (4.49, 6.13)
Urate (umol l^{-1})	350 (305, 399)	264 (225, 309)	303 (250, 361)
Aspartate aminotransferase (U l^{-1})	26 (23, 31)	23 (20, 27)	24 (21, 29)
Alanine aminotransferase (U l^{-1})	24 (18, 32)	18 (14, 23)	20 (15, 27)
Alkaline phosphatase (U l^{-1})	79 (67, 93)	82 (67, 98)	80 (67, 96)
Albumin (g l^{-1})	45.52 (43.80, 47.24)	44.91 (43.21, 46.63)	45.20 (43.47, 46.93)
C-reactive protein (mg l^{-1})	1.29 (0.67, 2.55)	1.38 (0.65, 2.95)	1.33 (0.66, 2.76)
Erythrocytes (10^{12} cells l^{-1})	4.74 (4.51, 4.98)	4.32 (4.10, 4.54)	4.50 (4.23, 4.79)
Leukocytes (10^9 cells l^{-1})	6.68 (5.66, 7.89)	6.61 (5.60, 7.81)	6.64 (5.62, 7.85)
Platelets (10^9 cells l^{-1})	234 (202, 269)	261 (226, 301)	248 (214, 287)
Hemoglobin (g dl^{-1})	15.00 (14.37, 15.64)	13.50 (12.90, 14.10)	14.15 (13.31, 15.02)
Hematocrit (%)	43.3 (41.4, 45.2)	39.2 (37.5, 41.0)	41.0 (38.7, 43.5)
Mean corpuscular volume (fl)	91.4 (88.8, 94.1)	91.1 (88.4, 93.7)	91.2 (88.6, 93.9)
Mean corpuscular hemoglobin (pg)	31.69 (30.70, 32.70)	31.37 (30.33, 32.37)	31.50 (30.50, 32.50)
Mean corpuscular hemoglobin (g dl^{-1})	34.60 (34.00, 35.22)	34.36 (33.80, 35.00)	34.48 (33.90, 35.10)
Antihypertensives	1,090 (2.0%)	680 (1.1%)	1,770 (1.5%)

[a]Median (IQR); n (%)

stratify the risk of common diseases such as cardiovascular disease[3], kidney disease[34] or diabetes[32]. While more complex risk scores have been proposed[35], the trade-off between the added predictive information and resources in time and cost required to collect the new data has limited clinical adoption[36]. We therefore investigated the predictive information of the relatively affordable and standardized NMR metabolomics assay against common clinical variables in the UK Biobank and in four independent validation cohorts.

First, we modeled disease risk for each endpoint in the UK Biobank using CPH models for three clinical predictor sets with increasing complexity: Age+Sex, highly predictive and available

ahead of any test; ASCVD, a set of readily available cardiovascular predictors; and PANEL, a comprehensive selection of clinical predictors including in-depth blood measurements (Fig. 1d) exceeding those typically available in primary care. For all sets, the performance of CPH models was benchmarked against those based on the sets' combinations with the metabolomic state. As quantified by Harrell's C-index, the discriminative performances of all models at 10 years after baseline are shown in Fig. 3a. Subsequently, to validate metabolomic states, we applied the trained metabolomic state model to the external validation cohorts and replicated the CPH models with and without metabolomic state addition for
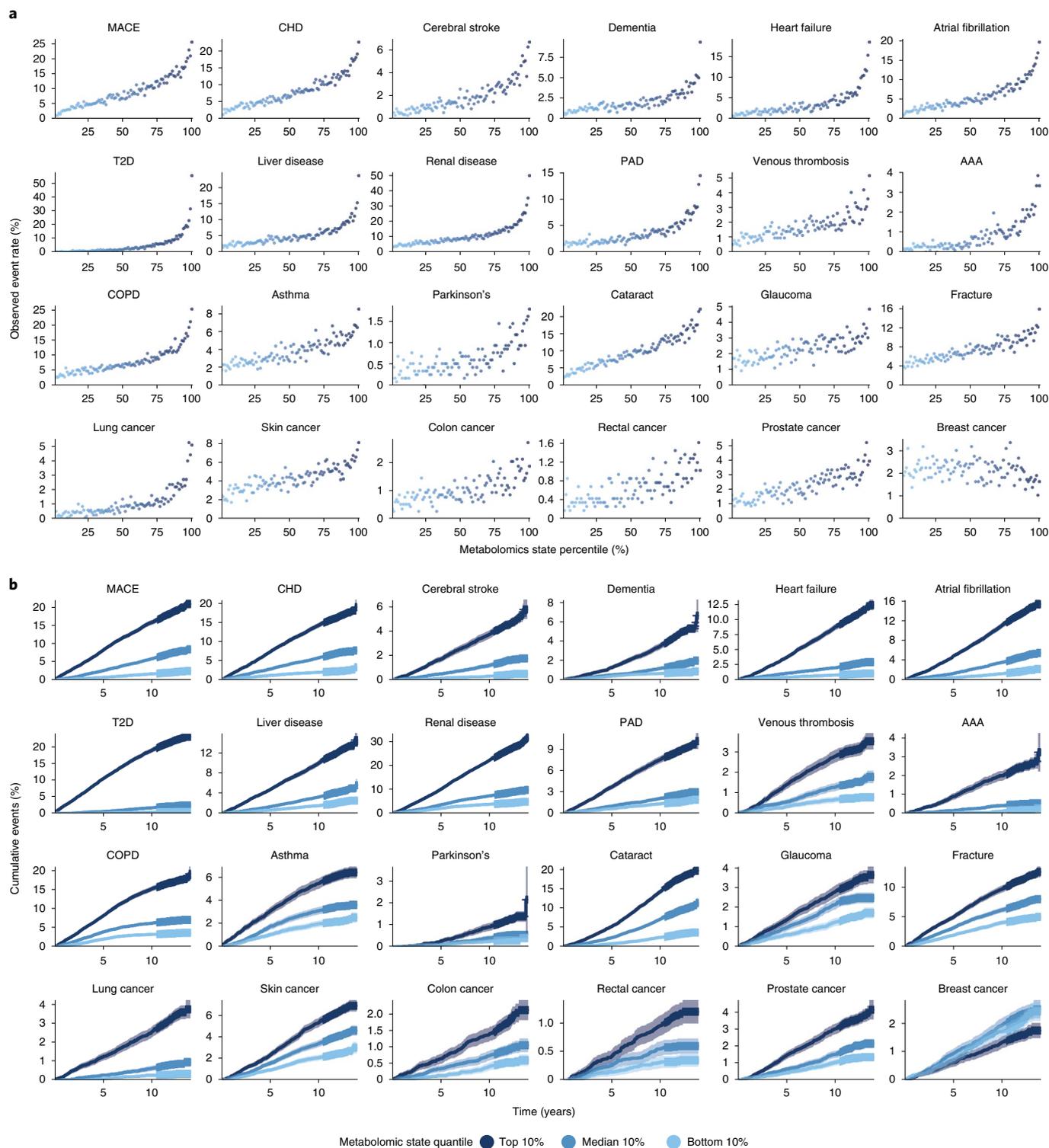
Fig. 2 | Metabolomic state is associated with ORs and stratifies survival. a, Observed event frequency for incident disease plotted against metabolomic state percentiles over the entire study population for all 24 endpoints. **b**, Cumulative event rates over the observation time for all assessed endpoints, stratified by metabolomic state quantiles (light blue, bottom 10%; blue, median 10%; dark blue, top 10%), with 95% CIs indicated. PAD, peripheral artery disease.

the Age+Sex predictor set for all endpoints available. The results of the external validation are shown in Extended Data Fig. 3. We noted the discriminative performance of the metabolomic state to be highly disease dependent: while the metabolomic state contained significantly less predictive information than clinical

predictors for cataract, glaucoma and skin, colon, rectal and prostate cancers, this was not the case for renal disease, liver disease and T2D. Here, the metabolomic state contained a greater predictive value than Age+Sex and even ASCVD. Generally, we observed an increase in discriminative performance with the addition of more
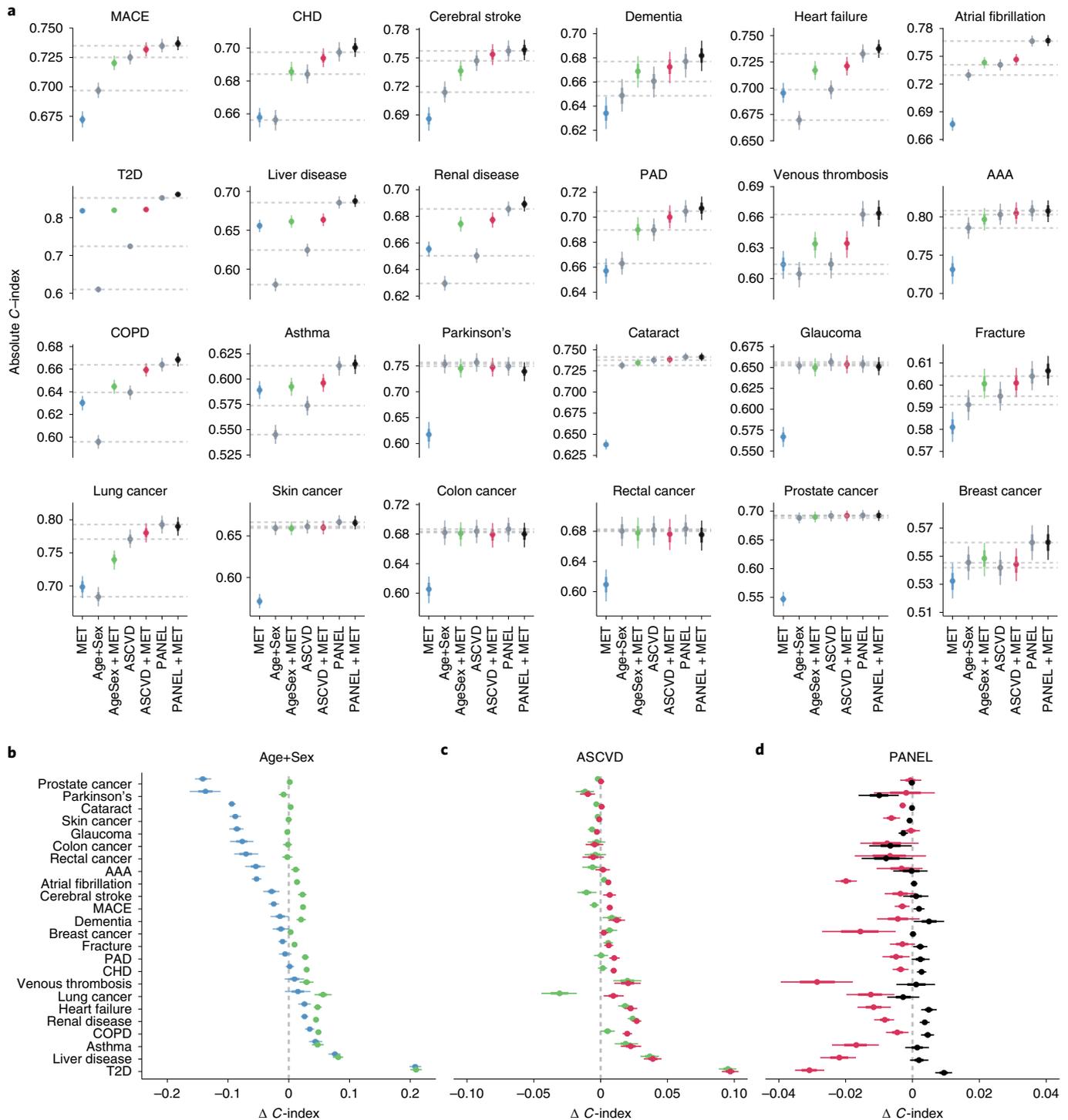
Fig. 3 | Predictive value of the metabolomic state is endpoint dependent. a, Comparison of discriminative performance of CPH models trained on the metabolomic state only (MET), the three clinical predictor sets (Age+Sex, ASCVD and PANEL) and the sets' combinations with the metabolomic state. Horizontal dashed lines indicate the median performance of the three clinical predictor sets. **b**, Differences in discriminative performance between the Age+Sex baseline (dashed line), metabolomic state only (blue) and the combination of Age+Sex and metabolomic state (green). **c**, Differences in discriminative performance between ASCVD predictors (dashed line), the combination of Age+Sex and the metabolomic state (green) and the combination of metabolomic state and ASCVD predictors (red). **d**, Difference in discriminative performance between comprehensive PANEL predictors (dashed line), ASCVD + MET (red) and PANEL + MET (black). **a–d**, Statistical measures were derived from $n = 117.981$ individuals; those with previous events were excluded (Supplementary Table 1). Data are presented as median (center of error bar) and 95% CI (line of error bar) determined by bootstrapping of with 1,000 iterations. **b–d**, The x-axis range differs across panels; vertical grid lines indicate differences of 0.02 C-index.

comprehensive clinical predictors across all endpoints, and performances were stable over different age groups, biological sexes and ethnic backgrounds (Extended Data Fig. 4).

To better assess the predictive value of the metabolomic state (MET) in comparison with clinical variables, we calculated C-index deltas (Fig. 3b–d). We noted that CPH models fit solely on the metabolomic state performed competitively or better than Age+Sex for ten of the 24 endpoints, including T2D and COPD, but also for heart failure, liver disease and renal disease (Fig. 3b). The competitive performance compared with Age+Sex was replicated in the validation cohorts for T2D, COPD, heart failure, coronary heart disease (CHD) and all-cause dementia (Extended Data Fig. 3 and Supplementary Table 10).

Interestingly, CPH models fit on the combination of the metabolomic state with Age+Sex (Age+Sex+MET) performed comparably to, or better than, the ASCVD predictors for 15 of the 24 endpoints, including T2D, liver disease, renal disease, heart failure, venous thrombosis and dementia (Fig. 3c). While the comprehensive PANEL score generally contained the most predictive information, surprisingly we observed only modest gains over the combination of ASCVD and the metabolomic state, and Age+Sex and the metabolomic state (Fig. 3d). Applying the complex metabolomic state model architecture to the predictors of the PANEL, we did not observe systematic performance improvements (Extended Data Fig. 5).

Discriminative improvements over clinical predictors. In addition to investigating the shared information, we were interested in quantifying the additive predictive value of metabolomic state over readily available clinical variables. To understand how the information is distributed over the PANEL predictors, we first assessed the aggregated coefficients of the CPH model and found that basic demographic information, medical history and physical measurements provided the most predictive information over all endpoints (Supplementary Table 5). In addition, apart from shared measures (for example, glucose, albumin or creatinine), lipids and creatinine/cystatin c, we did not observe strong correlations ($|r| > 0.5$) between the PANEL predictors and NMR metabolites (Supplementary Table 6). Therefore, we continued assessment of performance differences between the CPH models' fit on clinical predictors and those with the added metabolomic state by calculating differences in the C-index (Supplementary Table 9).

In the UK Biobank, the metabolomic state significantly added predictive information over age and sex for 18 of the 24 endpoints; in contrast, endpoints with a comparably low predictive value of the metabolomic state, such as Parkinson's disease, skin cancer, colon cancer, rectal cancer, glaucoma and cataract, did not benefit from the addition of the metabolomic state. Results from four external cohorts independently confirmed significant discriminative improvements over Age+Sex for CHD, heart failure, atrial fibrillation, T2D and COPD (for detailed results and event counts for the independent cohorts, see Extended Data Fig. 3a and Supplementary Table 10).

Beyond basic demographic predictors, addition of the metabolomic state to cardiovascular predictors further significantly improved discriminative performance for 15 of the 24 endpoints (Fig. 3c). Even when added to the comprehensive PANEL set, the metabolomic state provided significant additional discriminatory value for eight of the 24 endpoints (Fig. 3d) as quantified by C-index, including T2D (0.009, 95% CI 0.007, 0.012), dementia (0.005, 95% CI 0, 0.009), heart failure (0.005, 95% CI 0.003, 0.007), COPD (0.005, 95% CI 0.003, 0.006), renal disease (0.004, 95% CI 0.002, 0.005), CHD (0.003, 95% CI 0.001, 0.004) and MACE (0.002, 95% CI 0, 0.004).

We further sought to understand the potential of the metabolomic state in regard to individual risk under consideration of established

clinical predictors. Therefore, we examined the partial effects and hazard ratios (HRs, per s.d. metabolomic state, with 95% CI) of the CPH models trained on the combinations of the metabolomic state and predictor sets Age+Sex, ASCVD and PANEL (Extended Data Fig. 6a) for those 18 endpoints with discrimination benefits over the Age+Sex set. We observed a notable separation between the top, median and bottom 10% of the metabolomic state in 14 of the 18 endpoints when adjusted for more comprehensive clinical predictors (for HRs, see Extended Data Fig. 6b). A change of 1 s.d. in the metabolomic state for T2D resulted in substantially adjusted HRs (HR$_{Age+Sex}$ 3.83 (95% CI 3.71–4.01), HR$_{PANEL}$ 2.5 (95% CI 2.34–2.67)), which were replicated with adjustment for Age+Sex in the independent cohorts (Extended Data Fig. 3b). Other investigated endpoints, such as all-cause dementia (HR$_{Age+Sex}$ 1.56 (95% CI 1.54–1.72), HR$_{PANEL}$ 1.46 (95% CI 1.43–1.47)), heart failure (HR$_{Age+Sex}$ 1.8 (95% CI 1.74–1.86), HR$_{PANEL}$ 1.45 (95% CI 1.38–1.52)), COPD (HR$_{Age+Sex}$ 1.56 (95% CI 1.53–1.6), HR$_{PANEL}$ 1.35 (95% CI 1.31–1.39)) or MACE (HR$_{Age+Sex}$ 1.63 (95% CI 1.58–1.69), HR$_{PANEL}$ 1.4 (95% CI 1.33–1.46)), showed less pronounced, yet clear, separation of risk trajectories. In regard to T2D, the HRs of the metabolomic states were externally validated with adjustment for Age+Sex for all-cause dementia, heart failure, atrial fibrillation, CHD and COPD (Extended Data Fig. 3b). In contrast, the metabolomic state only marginally modified the risk trajectories for asthma (HR$_{Age+Sex}$ 1.37 (95% CI 1.3–1.44), HR$_{PANEL}$ 1.09 (95% CI 1.03–1.16)) and cataract (HR$_{Age+Sex}$ 1.22 (95% CI 1.18–1.25), HR$_{PANEL}$ 1.08 (95% CI 1.05–1.11)).

Discriminative performance translates to clinical utility. While discrimination is critical, the clinical utility of any risk model depends on calibration and the choice of adequate thresholds for interventions. We found all models well calibrated in the UK Biobank Cohort (see Fig. 4a–c and Supplementary Fig. 1 for details on all endpoints). UK Biobank[37], as one of the largest and most comprehensive population cohorts in the world, therefore, allowed us to estimate clinical utility with high precision over a wide range of clinically reasonable intervention thresholds. However, adequate clinical decision thresholds directly depend on the benefits and harms of interventions and disease prevalence. We therefore calculated decision curves[38] to estimate the benefit of adding metabolomic information to a prediction model (see Fig. 4d–i and Supplementary Fig. 1 for details on all endpoints). Further, we calculated clinically relevant metrics such as sensitivity, positive predictive value and positive likelihood ratio over multiple false-positive rates (Supplementary Table 11)[39].

Specifically, we investigated the application of the metabolomic state in two scenarios. First, as a potentially economical and practical option, we assessed the combination of the metabolomic state with Age+Sex and with the less resource-intensive, non-laboratory predictors of the PANEL set. Second, we combined the metabolomic state with the entire PANEL set (including all laboratory predictors) to assess whether there is a net benefit even beyond comprehensive predictors.

Generally we found that discriminative gains (Fig. 3) translated to utility gains (see Fig. 4d–i and Supplementary Fig. 1 for details on all endpoints). The metabolomic state substantially added to age and sex for most endpoints, and additional non-laboratory predictors either closed (12 of the 24 endpoints, including T2D, stroke, heart failure and lung cancer) or narrowed the gap (an additional four of the 24 endpoints, including dementia, atrial fibrillation and renal disease) with the comprehensive set of PANEL predictors. The addition of the metabolomic state to the comprehensive PANEL predictors led to further improvements in the utility for reasonable ranges of decision thresholds for 11 of the 24 endpoints (most notably T2D, heart failure and, to a lesser extent, dementia; see Supplementary Fig. 1 for details on all endpoints and Extended Data Fig. 7 for
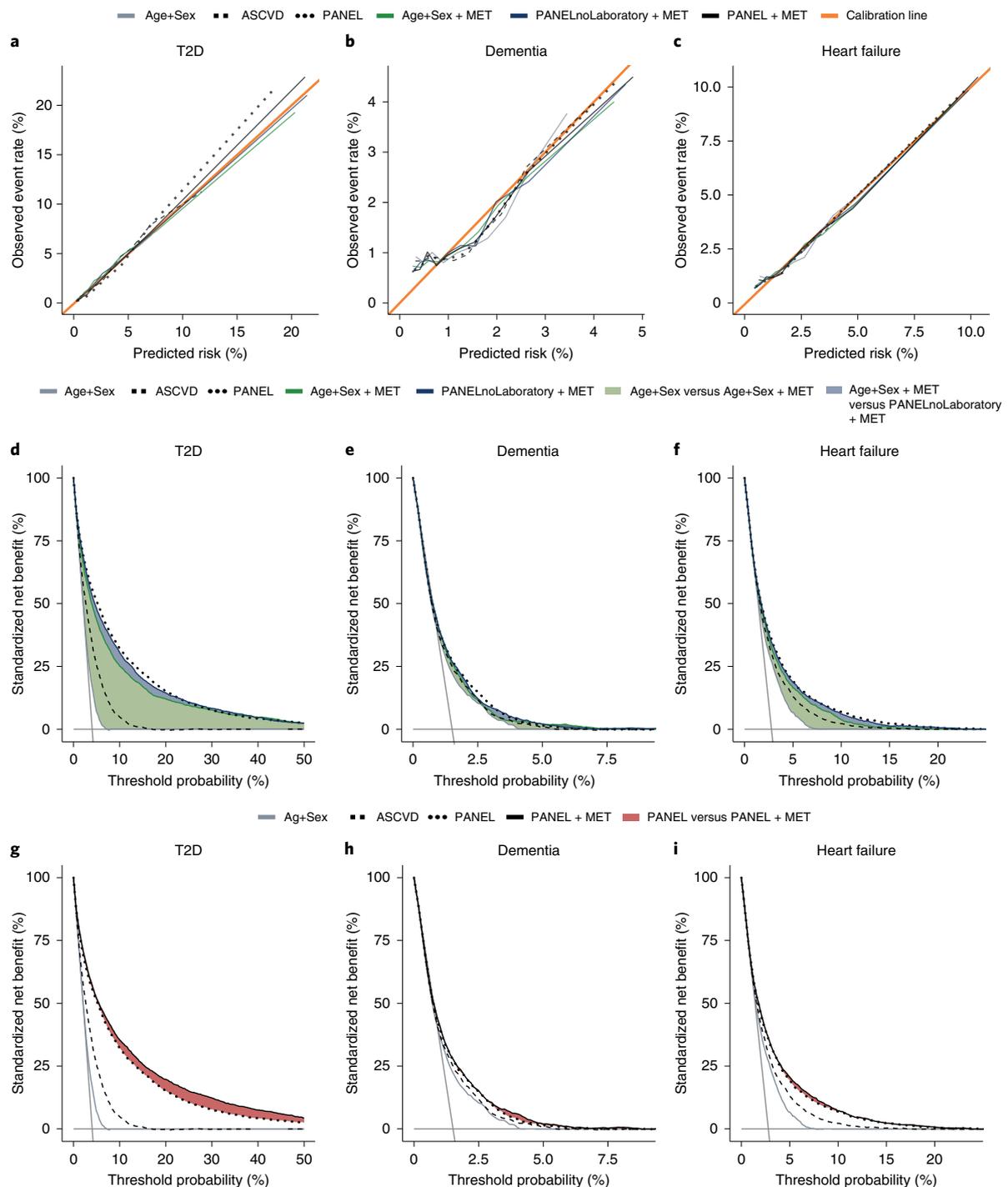
Fig. 4 | Model calibration and additive predictive value of the metabolomic state translate to potential clinical utility. a–c, Calibration curves for CPH models, including baseline parameter sets Age+Sex, ASCVD and PANEL, as well as their combinations with the metabolomic state (Age+Sex+MET) for the endpoints T2D (**a**), dementia (**b**) and heart failure (**c**). **d–f**, Endpoint-specific net benefit curves standardized by endpoint prevalence, where horizontal solid gray lines indicate 'treat none' and vertical solid gray lines indicate 'treat all'; T2D (**d**), dementia (**e**) and heart failure (**f**). The standardized net benefits of sets Age+Sex, ASCVD and PANEL are compared with Age+Sex+MET and additional non-laboratory predictors of PANEL (PANELnoLaboratory). Green and blue color-filled areas indicate the added benefit of the combination of the metabolomic state and Age+Sex and PANELnoLaboratory, respectively. **g–i**, Standardized net benefit curves comparing the performance of PANEL+MET against baselines Age+Sex, ASCVD and PANEL; T2D (**g**), dementia (**h**) and heart failure (**i**). Decision curves were derived from $n = 111,745$ (T2D), $n = 117,245$ (dementia) and $n = 113,636$ (heart failure) individuals.

additional analyses investigating apolipoprotein 4 (*APOE4*) carrier status for dementia). Conversely, where there were no improvements in the discriminatory value, no relevant improvements in clinical utility could be found. These observations were further reflected in the positive predictive values and positive likelihood ratios (Supplementary Table 9).
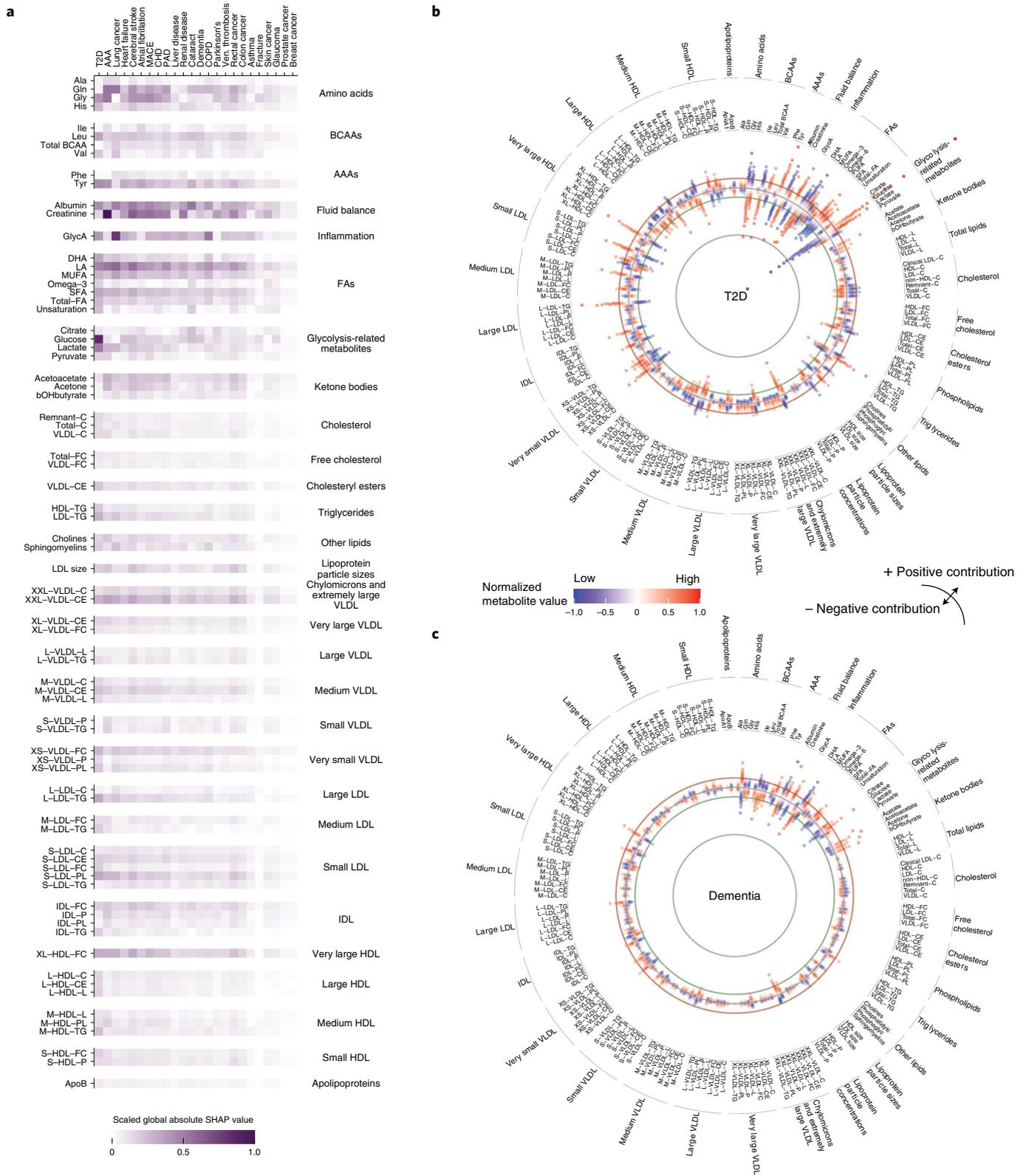
Fig. 5 | Analysis of the metabolomic state informs on metabolite profiles associated with disease risk. a, Heatmap showing the importance of metabolites in regard to the estimated metabolomic states, represented by absolute global SHAP value estimates per endpoint for the 75 globally most important metabolites. Endpoints are sorted by the discriminative performance of the metabolomic state (left to right; Fig. 3a). **b**, Global metabolite attributions for T2D; individual attributions are aggregated by percentiles and each dot indicates one percentile. The more distant a dot from the circular baseline, the stronger the absolute attribution for that percentile. Deviations toward the center and periphery represent negative and positive contributions, respectively, to the metabolomic state. Colors indicate the metabolite's mean plasma value. **c**, Global metabolite attributions for all-cause dementia. IDL, intermediate-density lipoprotein.

Identification of disease-specific metabolite profiles. A requirement for the adoption of neural networks in medicine is explainability. While neural networks are not inherently interpretable, methods have been developed to overcome this challenge[40]. To identify which metabolites most affect disease risk, we approximated Shapely additive explanation (SHAP) values[41] for all investigated diseases. Generally, the larger the absolute SHAP value the more important a metabolite for an individual prediction. Based on the direction of the effect of a metabolite's contribution, increasing or decreasing the predicted risk, SHAP can take a positive or negative value.

To understand individual metabolites in the context of the 24 investigated diseases, we investigated global metabolite attributions, the sum of absolute SHAP values per metabolite and disease (Fig. 5a and Extended Data Fig. 8). We found that most high-impact metabolites were linked to multiple diseases: plasma levels of metabolites with consistently high contribution included the amino acids glutamine, glycine and tyrosine, metabolites related to carbohydrate metabolism, albumin, the kidney function marker creatinine, glycoprotein acetylation (GlycA) and the ketone bodies acetone and acetoacetate. Further implicated were fatty acids (FA) such as linoleic acid (LA) and multiple lipoprotein components, including free cholesterol in very large high-density lipoprotein (VHDL), triglycerides in large low-density lipoprotein (LDL), phospholipids in small LDL and sphingomyelins. In addition to shared metabolite profiles, we pinpointed marked associations of creatinine with AAA, glucose with T2D and GlycA with lung cancer and COPD. For diseases with a high discriminatory value for metabolomic state, predicted metabolite contributions were considerably higher than for diseases with little discriminatory metabolomic information (Fig. 5a).

Subsequently we focused on T2D (Fig. 5b) and all-cause dementia (Fig. 5c), two diseases with strong metabolomic contributions over the comprehensive clinical predictors and indications of clinical utility (see above). Metabolites related to carbohydrate metabolism, such as glucose and lactate, dominated the predicted metabolomic state of T2D in our model (Fig. 5b). In line with earlier findings[5,19,42], we observed contributions of amino acids, ketone bodies, lipids and FAs as well as creatinine and albumin. We confirmed that higher plasma levels of FAs, docosahexaenoic acid (DHA) and LA were associated with lower risk[43,44]. Further, we observed a distinct contribution of lipid content across the whole density gradient of lipoproteins, including a high triglyceride content in LDL and IDL or free cholesterol content in very small very-low-density lipoprotein (VLDL) and HDL. For all-cause dementia, we identified creatinine, albumin and the amino acids glutamine, leucine and tyrosine as predominant contributors to predicted risk (Fig. 5c). In line with earlier findings[8,45], we observed a notable role of FAs such as LA and monounsaturated and saturated FAs, as well as a protective effect of branched-chain amino acids (BCAAs). Our results further implicate associations of glucose, ketone bodies acetate, acetoacetate and acetone, and beta-hydroxybutyrate. Finally we found several lipoproteins to be associated, most notably free cholesterol in very large HDL and cholesterylester in extremely large VLDL. Comprehensive data for all investigated endpoints, including the most important metabolites and disease-specific attribution profiles, can be found in Extended Data Fig. 8, Supplementary Table 12 and Supplementary Fig. 2).

Computation of SHAP values also allowed us to derive individual risk attribution profiles for individual participants and diseases, informing on the impact of single metabolites on a given prediction. We visualize the attribution profiles for T2D in two-dimensional uniform manifold approximation and projection (UMAP)[46] space (Extended Data Fig. 9), which is resolved by the estimated importance of glucose (that is, SHAP values assigned to glucose regarding the predicted risk for T2D; Extended Data Fig. 9a). While most high-risk individuals (top 1% metabolic state) are located at coordinates with strong glucose attribution, we found high-risk individuals scattered over the entire attribution space (Extended Data Fig. 9b). Interestingly, the attribution profiles of high-risk individuals were not consistently dominated by glucose but rather by, for instance, low levels of albumin, LA, DHA, histidine and glycine (Extended Data Fig. 9c). This observation is further reflected in NMR metabolite concentrations, because we found substantial differences in the concentrations of glucose, LA, FAs and triglycerides when comparing the metabolite distributions of individuals in the area with the strongest glucose attribution with those of individuals in two spatially distinct, high-risk UMAP areas (Extended Data Fig. 10).

Discussion

The assessment of risk is a critical component of disease prevention. However, comprehensive risk assessment often requires the careful acquisition of predictors, one disease at a time. Thus, for each disease-specific risk score, the resources (time and cost) required for the collection can severely limit adoption and utility[47]. Interestingly, many common diseases involve metabolic alterations and human blood metabolomic patterns contain rich systemic information on the underlying physiology[9–11,20,21]. While individual metabolites have long been linked to disease risk, systemic information from blood metabolomics could inform on multiple diseases simultaneously. Importantly, in recent years, assays such as 1H-NMR spectroscopy have matured and allowed the assessment of serum metabolite information robustly at comparatively low cost[13,14]. However, the potential of metabolomic profile as a single-domain, multidisease assay in primary care has not been investigated thus far.

We have assessed the potential of NMR-derived metabolomic profiles as a tool for individualized prediction of onset across 24 common diseases. With >1.4 million person-years of follow-up, we leveraged the systemic information in metabolomic profiles to derive integrative metabolomic states for many diseases simultaneously. We found the metabolomic states to be predictive for all but one of the investigated diseases and externally validated these findings in four independent cohorts for available endpoints. Further, we investigated the predictive value beyond clinical variables and identified a subset of endpoints with potential clinical utility. Finally, we examined metabolite attributions confirming a multitude of disease-associated metabolites and a shared metabolomic background of common diseases.

Importantly, we found that the predictive information of the metabolomic state matched established clinical variables for many of the investigated endpoints. In line with previous reports on NMR–metabolite associations, we confirm that metabolomic profiles are highly predictive for, for example, T2D[19], dementia[8] and cardiovascular diseases[6,11,17] such as CHD and heart failure[48]. Generally, the additional predictive information decreased over comprehensive clinical predictors, indicating that substantial parts of the metabolomic state's discriminatory information are shared with established clinical predictors. However, for multiple endpoints, including T2D, all-cause dementia and heart failure, the metabolomic state contained complementary information that added predictive value even over comprehensive laboratory measurements. These findings largely translate into potential clinical utility for NMR-based metabolomic profiling, both as a replacement for comprehensive laboratory examinations and as an additional source of discriminatory information to refine comprehensive risk assessments for multiple diseases simultaneously.

Calculation of attributions for each individual allowed us to assess how differences in the metabolomic profile affect disease risk. We confirmed the role of metabolites such as albumin and creatinine, which have previously been associated with all-cause and disease-specific mortality[11,16] and are already part of routine care[49,50]. Further, we confirmed the role of LA, tyrosine, glycine and cholesterylesters in extremely large VLDL in multiple diseases,

further supporting metabolomic multidisease-spanning information[21]. Dissecting disease-specific attribution profiles, we found that metabolite attributions reflect metabolite–disease associations previously reported in the literature. In the case of T2D, we confirmed the associations between disease risk and metabolites beyond glucose. Specifically, our model captured the positive association between high levels of glycoprotein acetyls, BCAAs, lactate and FAs (both monounsaturated and saturated) and the protective role of metabolites such as LA or glycine[5,19]. In the attribution profile of dementia we replicated associations with BCAAs, including leucine and valine, and with FAs, most notably LA[8,45]. In addition, the associations of GlycAs with cardiovascular disease, T2D, COPD and lung cancer[51,52] are reflected in the attributions. Consequently, our metabolomic state model learns systemic information in NMR-derived metabolomic profiles based on established shared and highly specific metabolite–disease associations.

In our perspective, 1H-NMR metabolomics profiling is an attractive candidate for a single-domain, multidisease assay. Because many countries already recommend regular check-ups entailing blood tests in the prevention of selected common diseases[53], our results indicate the potential of NMR metabolomic profiling in combination with simple demographic, but also with comprehensive laboratory predictors to estimate disease risk. In addition, metabolomic risk profiles could be of potential value in the guidance of pharmacological and lifestyle interventions. This is especially relevant for diseases such as T2D, where interventions on modifiable risk factors have been shown to delay disease onset[54] and prevent subsequent comorbidities[55,56]. Similarly, the Lancet 2020 commission suggested that up to 40% of worldwide dementia may be preventable by interventions on modifiable risk factors[57]. This is particularly compelling because today's pharmacological treatment options for dementia are scarce. However, the efficacy of various lifestyle interventions[58,59] is disputed, calling for further experimental investigation.

Before application in routine care, substantial challenges remain. While the 1H-NMR assay is robust and cheaper than mass-spectrometry-based alternatives, sensitivity is lower. Also, current metabolite coverage is relatively narrow and lipid focused[13,14,60]. Although a future expansion of metabolite coverage is expected, it presents a limitation for clinical utility to date. Further, downstream quantification from raw NMR spectra needs to be harmonized for the reliable application of multivariable prediction models. While our study population is more healthy and less deprived than the general UK population[37], the results of external validation in four independent cohorts indicate general transferability of metabolomic states. However, the scope of validation was limited by the available endpoint information, constraining the replication to a subset of seven endpoints. In light of these limitations, we recommend careful scrutinization before application of the metabolomic state model beyond the validated conditions or in specific populations outside the research context. Ultimately, a broad rollout of NMR metabolomics for clinical care requires multiple logistical questions to be addressed, including both sample processing and transport.

Taken together, our work demonstrates the potential and limitations of NMR-derived metabolomic profiles as a multidisease assay to inform on the risk of many common diseases simultaneously.

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/s41591-022-01980-3.

References

1. WHO CVD Risk Chart Working Group. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob. Health* **7**, e1332–e1345 (2019).
2. A and B recommendations. *U.S. Preventive Services Task Force* https://www.uspreventiveservicestaskforce.org/uspstf/recommendation-topics/uspstf-and-b-recommendations (2022).
3. Goff David, C. et al. 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *Circulation* **129**, S49–S73 (2014).
4. Würtz, P. et al. Circulating metabolite predictors of glycemia in middle-aged men and women. *Diabetes Care* **35**, 1749–1756 (2012).
5. Mahendran, Y. et al. Association of ketone body levels with hyperglycemia and type 2 diabetes in 9,398 Finnish men. *Diabetes* **62**, 3618–3626 (2013).
6. Holmes, M. V. et al. Lipids, lipoproteins, and metabolites and risk of myocardial infarction and stroke. *J. Am. Coll. Cardiol.* **71**, 620–632 (2018).
7. Lécuyer, L. et al. NMR metabolomic signatures reveal predictive plasma metabolites associated with long-term risk of developing breast cancer. *Int. J. Epidemiol.* **47**, 484–494 (2018).
8. Tynkkynen, J. et al. Association of branched-chain amino acids and other circulating metabolites with risk of incident dementia and Alzheimer's disease: a prospective study in eight cohorts. *Alzheimers Dement.* **14**, 723–733 (2018).
9. Ahadi, S. et al. Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nat. Med.* **26**, 83–90 (2020).
10. Schüssler-Fiorenza Rose, S. M. et al. A longitudinal big data approach for precision health. *Nat. Med.* **25**, 792–804 (2019).
11. 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).
12. Markley, J. L. et al. The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **43**, 34–40 (2017).
13. Soininen, P., Kangas, A. J., Würtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ. Cardiovasc. Genet.* **8**, 192–206 (2015).
14. Würtz, 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).
15. Ala-Korpela, M., Zhao, S., Järvelin, M.-R., Mäkinen, V.-P. & Ohukainen, P. Apt interpretation of comprehensive lipoprotein data in large-scale epidemiology: disclosure of fundamental structural and metabolic relationships. *Int. J. Epidemiol.* **51**, 996–1011 (2021).
16. 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).
17. Würtz, P. et al. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131**, 774–785 (2015).
18. Fizelova, M. et al. Associations of multiple lipoprotein and apolipoprotein measures with worsening of glycemia and incident type 2 diabetes in 6607 non-diabetic Finnish men. *Atherosclerosis* **240**, 272–277 (2015).
19. 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).
20. Julkunen, H., Cichońska, A., Slagboom, P. E. & Würtz, P., Nightingale Health UK Biobank Initiative. Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *eLife* **10**, e63033 (2021).
21. Pietzner, M. et al. Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat. Med.* **27**, 471–479 (2021).
22. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
23. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34**, 187–202 (1972).
24. Marmot, M. & Brunner, E. Cohort profile: the Whitehall II study. *Int. J. Epidemiol.* **34**, 251–256 (2005).
25. Ikram, M. A. et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur. J. Epidemiol.* **32**, 807–850 (2017).
26. 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).
27. Shepherd, J. et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet* **360**, 1623–1630 (2002).
28. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
29. Qiu, C. et al. Association of blood pressure and hypertension with the risk of Parkinson disease: the National FINRISK Study. *Hypertension* **57**, 1094–1100 (2011).
30. de Bruijn, R. F. A. G. & Ikram, M. A. Cardiovascular risk factors and future risk of Alzheimer's disease. *BMC Med.* **12**, 130 (2014).

31. Johnson, C. B., Davis, M. K., Law, A. & Sulpher, J. Shared risk factors for cardiovascular disease and cancer: implications for preventive health and clinical care in oncology patients. *Can. J. Cardiol.* **32**, 900–907 (2016).

32. Lindström, J. & Tuomilehto, J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* **26**, 725–731 (2003).

33. Sindi, S. et al. The CAIDE Dementia Risk Score App: the development of an evidence-based mobile application to predict the risk of dementia. *Alzheimers Dement.* **1**, 328–333 (2015).

34. van der Velde, M. et al. Screening for albuminuria identifies individuals at increased renal risk. *J. Am. Soc. Nephrol.* **20**, 852–862 (2009).

35. Mars, N. et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat. Med.* **26**, 549–557 (2020).

36. Huang, C. et al. Performance metrics for the comparative analysis of clinical risk prediction models employing machine learning. *Circ. Cardiovasc. Qual. Outcomes* **14**, e007526 (2021).

37. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).

38. Vickers, A. J. & Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making* **26**, 565–574 (2006).

39. Hingorani, A. D. et al. Polygenic scores in disease prediction: evaluation using the relevant performance metrics. Preprint at *medRxiv* https://doi.org/10.1101/2022.02.18.22271049 (2022).

40. Gilpin, L. H. et al. Explaining explanations: an overview of interpretability of machine learning. Preprint at https://arxiv.org/abs/1806.00069v3 (2018).

41. Lundberg, S. & Lee, S.-I. A unified approach to interpreting model predictions. Preprint at https://arxiv.org/abs/1705.07874v2 (2017).

42. Mahendran, Y. et al. Glycerol and fatty acids in serum predict the development of hyperglycemia and type 2 diabetes in Finnish men. *Diabetes Care* **36**, 3732–3738 (2013).

43. Wu, J. H. Y. et al. Omega-6 fatty acid biomarkers and incident type 2 diabetes: pooled analysis of individual-level data for 39 740 adults from 20 prospective cohort studies. *Lancet Diabetes Endocrinol.* **5**, 965–974 (2017).

44. Virtanen, J. K., Mursu, J., Voutilainen, S., Uusitupa, M. & Tuomainen, T.-P. Serum omega-3 polyunsaturated fatty acids and risk of incident type 2 diabetes in men: the Kuopio Ischemic Heart Disease Risk Factor study. *Diabetes Care* **37**, 189–196 (2014).

45. 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).

46. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

47. Steyerberg, E. W. et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* **10**, e1001381 (2013).

48. Delles, C. et al. Nuclear magnetic resonance-based metabolomics identifies phenylalanine as a novel predictor of incident heart failure hospitalisation: results from PROSPER and FINRISK 1997. *Eur. J. Heart Fail.* **20**, 663–673 (2018).

49. Wannamethee, S. G., Shaper, A. G. & Perry, I. J. Serum creatinine concentration and risk of cardiovascular disease: a possible marker for increased risk of stroke. *Stroke* **28**, 557–563 (1997).

50. Ronit, A. et al. Plasma albumin and incident cardiovascular disease: results From the CGPS and an updated meta-analysis. *Arterioscler. Thromb. Vasc. Biol.* **40**, 473–482 (2020).

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. Komaromy, A., Reider, B., Jarvas, G. & Guttman, A. Glycoprotein biomarkers and analysis in chronic obstructive pulmonary disease and lung cancer with special focus on serum immunoglobulin G. *Clin. Chim. Acta* **506**, 204–213 (2020).

53. NHS Health Check. *NHS* https://www.nhs.uk/conditions/nhs-health-check/ (accessed 23 August 2022).

54. Balk, E. M. et al. Combined diet and physical activity promotion programs to prevent type 2 diabetes among persons at increased risk: a systematic review for the Community Preventive Services Task Force. *Ann. Intern. Med.* **163**, 437–451 (2015).

55. Yusuf, S. et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* **364**, 937–952 (2004).

56. Patel, S. A., Winkel, M., Ali, M. K., Narayan, K. M. V. & Mehta, N. K. Cardiovascular mortality associated with 5 leading risk factors: national and state preventable fractions estimated from survey data. *Ann. Intern. Med.* **163**, 245–253 (2015).

57. Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet* **396**, 413–446 (2020).

58. Silarova, B. et al. Effect of communicating phenotypic and genetic risk of coronary heart disease alongside web-based lifestyle advice: the INFORM randomised controlled trial. *Heart* **105**, 982–989 (2019).

59. Ngandu, T. et al. The effect of adherence on cognition in a multidomain lifestyle intervention (FINGER). *Alzheimers Dement.* https://doi.org/10.1002/alz.12492 (2021).

60. Emwas, A.-H. M. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol. Biol.* **1277**, 161–193 (2015).