Genetic Architecture of Circulating Metabolic Biomarkers Across Ancestral Populations

Author information

Authors and Affiliation

Yohei Takeuchi[1], Masaru Koido[1], Yunye He[1], Shota Nakamura[1], Xin Liu[1], Naoki Itokawa[1], Hiromi Tsuru[2], Yoji Sagiya[3], Akiko Nagai[3], Takayuki Morisaki[3,4], Koichi Matsuda[3,4], Yoichiro Kamatani[1]

1. Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
2. Laboratory of Genome Technology, Institute of Medical Science, The University of Tokyo
3. Laboratory of Clinical Genome Sequence, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
4. BioBank Japan, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Contributions

Y.K. and M.K. conceived and supervised this study. K.M., T.M., and A.N. managed the data of BioBank Japan data set. T.M. led the metabolomics project of BBJ. Y.T., Y.H., S.N., N.I., and M.K. performed the bioinformatics analyses. H.T. and Y.S. handled sample management and serum dispensing. Y.T. drafted the manuscript; it was critically revised by M.K., Y.H. and Y.K. All authors approved the final revision of the manuscript.

Corresponding authors

Correspondence to Yoichiro Kamatani, Masaru Koido, and Takayuki Morisaki

Abstract

Circulating metabolite levels hold promise for disease risk prediction, yet the genetic architecture and cross-ancestry transferability remain incompletely characterized. Here, we conducted the largest genome-wide association study in East Asians to date, quantifying 248 blood metabolic biomarkers using nuclear magnetic resonance spectroscopy in 37,969 Japanese individuals. We identified eight novel associations and pinpointed rs75326924, causing CD36 deficiency, as a putative causal variant influencing average diameter for low-density lipoprotein particles. A cross-ancestry meta-analysis of 657,341 individuals uncovered 28 additional novel associations and 11,167 credible sets, indicating 52.6% of ancestry-specific causal variants despite highly correlated marginal effect sizes (Pearson r = 0.78–0.98). Integration with liver single-nucleus transcriptomics highlighted perivascular hepatocytes as a key cell type of the genetic regulation of circulating biomarkers. Finally, weighted sums of metabolic biomarkers derived from the UK Biobank predicted myocardial infarction and cirrhosis risks in Japanese individuals. Our findings refine the genetic architecture of metabolic biomarkers and underscore their clinical relevance across ancestries.

Main

Metabolites are low molecular weight compounds. In organisms, they are synthesized and degraded in cellular and tissue-level metabolic pathways, and perturbations in these processes can influence disease risk. Metabolite levels in tissues and biofluids capture these dynamics[1], and blood metabolic biomarker profiling, in particular, provides a snapshot of an individual's metabolic state. Genetic factors account for approximately 20–80% of the variation in metabolic levels[2]. Thus, elucidating the genetic architecture of metabolites is essential for uncovering the biological underpinnings of human metabolism.

Large-scale metabolomic genome-wide association studies (GWASs) (N > 10,000) have been conducted in several European cohorts[3–5], which have investigated the genetic factors and elucidated new metabolic processes. On the other hand, because genetic association signals strongly depend on linkage disequilibrium (LD) structure and allele frequencies, GWASs in non-European populations have the potential to identify novel loci undetected in European studies. Despite this advantage, large-scale metabolomic GWASs in the East Asian population have been limited to a study in Japan[6] and two studies in China[7,8], each involving 10,333, 10,792, and 34,394 individuals, respectively. Therefore, studies with larger East Asian sample sizes are necessary to comprehensively elucidate the metabolomic genetic architecture.

Previous studies have shown that measurements of metabolic biomarkers by nuclear magnetic resonance (NMR) spectroscopy and their weighted sums were associated with current and future health status[9–12]. Among the metabolomics platforms, NMR spectroscopy generally exhibits fewer batch effects and lower costs than other methods, such as mass spectrometry (MS)[13], making NMR-derived biomarker panels attractive for clinical applications. However, environmental and genetic determinants of disease risk, as well as disease prevalence and primary etiologies, differ across ancestral populations. Consequently, caution is required when interpreting the predictive ability for disease risk across different ancestries. For example, a model trained in a European population demonstrated poor lung cancer predictive accuracy among Asian ever-smokers[14]. Accordingly, the relationship between metabolic biomarkers and disease risk may also vary between European and

East-Asian populations, and the findings derived from European studies cannot be directly generalized to East-Asian populations[11].

In this study, we performed a GWAS of 248 metabolic biomarkers, including amino acids, lipids, and lipoproteins, analyzed in blood samples from 37,969 Japanese participants in Biobank Japan (BBJ) using NMR spectroscopy. This study is the largest metabolomic GWAS in the East Asian population to date. Using the metabolomic GWAS results from BBJ and UK Biobank (UKB) and Estonian Biobank (EBB), encompassing a combined N = 657,341 individuals, we characterized the ancestry-specific and shared genetic effects on metabolic biomarkers. To map these genetic signals to biology, we linked the metabolomic GWAS results to transcriptome profiles from a range of tissues and cell populations most significantly associated with each biomarker. Finally, utilizing metabolomic measurements and clinical data from BBJ, we evaluated the prognostic value of metabolic biomarkers for disease risk in the Japanese population.

Results

A total of 250 metabolic biomarkers (Supplementary Table 1) were initially analyzed using the NMR platform in 45,270 serum samples from the BBJ first cohort (BBJ1), which consisted of two independently genotyped sub-cohorts: BBJ1-180K, with 182,536 individuals, and BBJ1-12K, with 11,716 individuals. Glycerol and 3-hydroxybutyrate were excluded from this study because available measurements were less than 90% of the total samples. 248 analyzed metabolic biomarkers were classified into 18 groups based on their biological functions. For the present study, we used samples that passed both metabolomic and genotype quality control (QC) (see Methods). After QC, 37,969 samples from BBJ1-180K were used for metabolomic GWAS, 3,522 samples from BBJ1-12K were used for additional genetic association tests, and 15,137 to 34,765 samples from BBJ1-180K were available for survival analysis evaluating the associations between metabolic biomarker(s) and disease risk.

Investigation of genetic effects on metabolic biomarkers in the Japanese population

Using 37,969 Japanese participants from BBJ1-180K, we investigated the associations between 11 million imputed variants (minor allele frequency (MAF) > 0.1 % and an imputation quality score R^2 > 0.7) and 248 metabolic biomarkers using a generalized linear mixed model implemented in BOLT-LMM[15]. To account for multiple testing across 248 metabolic biomarkers, we applied a study-wide significance threshold (P value $< 5 \times 10^{-8}/20$, Bonferroni correction; see Methods) in the metabolomic GWAS. We identified 2,140 associations involving 297 unique lead variants and 244 metabolic biomarkers and detected 76 regions (one to 34 unique lead variants per region) without overlap across metabolic biomarkers (Fig. 1a and Supplementary Table 2). Of these, eight pairs (involving six unique lead variants and eight metabolic biomarkers) were novel, corresponding to six loci across metabolic biomarkers (Table 1).

To replicate these eight novel associations, we analyzed BBJ1-12K (N = 3,528). Six associations showed concordant effect directions with the discovery GWAS results, although the sign test did not

102 reach nominal significance (*P* value = 0.14). Furthermore, three associations (the *CD36* region with
103 the average diameter for low-density lipoprotein particles (LDL size) and the *ALDH2* region with
104 omega-3 fatty acids and citrate) showed nominal significance (*P* value < 0.05) in the same direction,
105 with three of these remaining significant after Bonferroni correction (*P* value < 0.05/8)
106 (Supplementary Table 3).

107 To find putative causal variants for metabolic biomarkers, we performed statistical fine-mapping of
108 2 Mb regions centered on each lead variant or index variant from secondary signals (see Methods).
109 We merged partially overlapping regions corresponding to the same metabolic biomarker, resulting
110 in 2,111 regions comprising 285 lead or index variants and 244 metabolic biomarkers. Using SuSiE-
111 RSS[16], we identified 3,652 independent causal signals (95% credible sets; CSs) across 244 metabolic
112 biomarkers (Fig. 1b). Each CS contained one to 203 variants (median: 4). 883 CSs (239 metabolic
113 biomarkers and 61 variants) comprised a single variant with a posterior inclusion probability (PIP) >
114 0.95, and 1,240 CSs (242 metabolic biomarkers and 89 variants) harbored at least one variant with
115 PIP > 80%. Among the 89 variants, 24 directly affect the protein function (functional variants; see
116 Methods), indicating that the proteins encoded by the corresponding 20 genes alter metabolite levels
117 (Supplementary Table 4). Notably, these coding variants included rs75326924, a missense variant in
118 *CD36* (Fig. 1c). This variant is known to cause CD36 deficiency and is also associated with
119 atherosclerotic cardiovascular disease[17]. rs75326924 was fine-mapped as a causal variant for 15
120 metabolic biomarkers, including a novel association with LDL size (PIP = 1.0). For the remaining 65
121 fine-mapped variants, we evaluated whether they were located within the enhancer elements linking
122 to target genes in 131 human tissue- or cell-types predicted by the Activity-by-Contact (ABC)
123 model[18,19] (ABC score > 0.015). This analysis identified 23 variants in 155 elements targeting 65
124 genes in 106 tissue or cell-types (Supplementary Table 5). Among non-coding causal variants, we
125 identified rs77869005, 5' untranslated region in *HAL*, as a potential causal variant for histidine levels
126 (PIP = 0.99). The MAF of this variant was markedly higher in the East Asian population, particularly
127 in the Japanese population, than in the European population (MAF in 1000 genomes project phase3
128 (1KGP3)[20] Europeans (EUR) = 0.0010, 1KGP3 East Asians (EAS) = 0.051; 1KGP3 Japanese in
129 Tokyo = 0.082, and BBJ1-180K = 0.084), suggesting that its higher frequency in the Japanese

population enabled the finding in our study. The ABC model predicts that elements containing rs77869005 influence expression levels of *HAL* in 12 cell and tissue types, with the highest ABC score observed in the liver. Histidine ammonia-lyase coded by *HAL* is the enzyme that converts histidine to urocanic acid in the liver[21]. These results suggested that rs77869005 influences blood histidine levels by regulating the *HAL* enhancer activity in the liver.

Meta-analysis and comparisons of genetic effect sizes across ancestral populations

To investigate genetic effects on metabolic biomarkers shared across ancestral populations, we performed a meta-analysis of 8 million imputed variants and 248 metabolic biomarkers using a fixed effects model with the inverse variance weighted (IVW) method implemented in METAL[22]. We integrated the metabolomic GWAS results in BBJ1-180K with the cross-ancestry meta-analysis results from UKB and EBB (UKB-EBB)[23] (N_{total} = 657,272). After applying the study-wide Bonferroni correction (P value < $5 \times 10^{-8}/20$), we identified 44,976 associations (7,983 unique lead variants and 248 metabolic biomarkers) and 665 regions with no overlap across metabolic biomarkers (Supplementary Table 6). Of these, 28 associations (16 unique lead variants and 26 metabolic biomarkers) and 14 regions across metabolic biomarkers were identified as novel associations in addition to those detected by GWAS in BBJ1-180K (Table 2).

To replicate these 28 novel associations, we evaluated the corresponding associations in BBJ1-12K. 19 associations showed concordant effect directions with the discovery meta-analysis results, and a sign test revealed a statistically significant enrichment of consistent effects (P value = 0.044) (Supplementary Table 7). Since statistical powers in BBJ1-12K were low (Power = 0.064-0.27) (Supplementary Table 7), we further evaluated these novel associations using almost independent large-scale cross-ancestry meta-analysis results (N_{total} = 136,016), with partial overlap with our meta-analysis (N = 3,701 in EBB)[3]. Of the 28 novel associations, 27 showed consistent effect directions in the large-scale meta-analysis, and a sign test again confirmed significant enrichment (P value = 1.1×10^{-7}) (Supplementary Table 7). In addition, 11 associations showed nominal significance (P value <

156 0.05) in the same direction; one of these remained significant after Bonferroni correction (*P* value <
157 0.05/28, Supplementary Table 7).

158 We assessed the cross-population similarity of genetic effects on metabolic biomarkers by evaluating
159 their correlations between the East Asian population from BBJ1-180K and six ancestral populations
160 from UKB-EBB or UKB: the European population from UKB-EBB, and the Central/South Asian,
161 African, East Asian, Middle Eastern, and Admixed American populations from UKB. First, we
162 extracted the lead variants identified in BBJ1-180K. We compared their effect sizes with those in each
163 UKB-EBB or UKB population, using variants with MAF > 1% in both populations being compared.
164 The Pearson correlation coefficients of effect sizes ranged from 0.78 in Admixed Americans to 0.95
165 in Europeans, with a relatively high correlation of 0.93 observed in UKB East Asians (Fig. 2a and
166 Supplementary Table 8). Next, we performed a reverse comparison: for each UKB-EBB or UKB
167 population, we extracted lead variants associated with metabolic biomarkers. We compared their
168 effect sizes with those in BBJ1-180K, again using variants with MAF > 1% in both populations.
169 Pearson correlation coefficients ranged from 0.85 in Middle Easterns to 0.98 in UKB East Asians (Fig.
170 2b and Supplementary Table 9). These results indicate that the genetic effects on metabolic
171 biomarkers are largely consistent across populations, particularly between East Asians in BBJ1-180K
172 and those in UKB.

173 Next, we performed a multi-ancestry fine-mapping using MESuSiE[24], integrating the GWAS results
174 from BBJ1-180K and the European-specific meta-analysis from UKB-EBB. We focused on 2,111
175 regions associated with traits in BBJ1-180K and identified 14,266 95% CSs across 244 metabolic
176 biomarkers. Each of the 11,167 CSs contained at least one variant with a PIP > 0.5 for being ancestry-
177 specific or shared (Supplementary Table 10). Despite the largely consistent effect sizes of lead
178 variants between the East Asian and European populations, we estimated 5,881 variants (52.6%) to
179 be population-specific causal variants, comprising 5,802 European-specific and 79 East Asian-
180 specific causal variants. The 3,099 remaining CSs did not contain a variant with PIP > 0.5 for being
181 ancestry-specific or shared. To investigate the potential functional mechanisms, we examined variants
182 with the highest PIP (> 0.5) in 95% CSs and found that population-shared variants were more

183 prevalent in categories including missense variants among the 14 types of annotation, while the most
184 frequently observed variants were intronic (Fig. 2c and Supplementary Table 11).

185

186 **Genetic links to blood metabolic biomarkers and tissue- and cell- types**

187 To identify tissue-type specificity of genetic associations for metabolic biomarkers, we investigated
188 the gene-level associations between circulating 248 metabolic biomarkers and 30 types of tissues
189 using data from the Genotype-Tissue Expression (GTEx)[25] using MAMGA[26] implemented in FUMA
190 (v.1.6.2)[27]. Using the GWAS results from BBJ1-180K, we identified five significant biomarker-tissue
191 pairs, all of which were for the liver (P value < 0.05/(20×30) after Bonferroni correction; see
192 Methods). Using the European-specific meta-analysis results from UKB-EBB, we identified 252
193 significant pairs (227 metabolic biomarkers and seven tissues), including the same five pairs
194 identified in BBJ1-180K (P value < 0.05/(20×30), after Bonferroni correction; see Methods) (Fig. 3a
195 and Supplementary Table 12). Across all analyzed tissues, the liver showed the largest number of
196 associations (222 metabolic biomarkers). Z-scores for these liver-biomarker associations were
197 significantly higher than those for the other 29 tissues (t-test P value < 0.05/(29 × 2), after Bonferroni
198 correction; Supplementary Fig. 1) in both BBJ1-180K and UKB-EBB, suggesting the central role of
199 the liver in the genetic regulation of these metabolic biomarkers in both populations. Beyond the liver,
200 the small intestine was associated with 13 metabolic biomarkers, whole blood with 10, the spleen
201 with three, the skeletal muscle with two, and the kidney and thyroid with creatinine.

202 To pinpoint specific liver cell-types linked to the 222 liver-associated metabolic biomarkers, we
203 integrated the GWAS results from BBJ1-180K and the European-specific meta-analysis results from
204 UKB-EBB with publicly available single-nucleus RNA sequencing (snRNA-seq) data from healthy
205 human liver[28]. This snRNA-seq analysis identified 10 cell clusters based on gene expression patterns.
206 Applying MAGMA to perform gene-level association tests between the 222 metabolic biomarkers
207 and 10 clusters, we identified 350 pairs (197 metabolic biomarkers and 10 clusters) (FDR < 5%) (Fig.
208 3b and Supplementary Table 13). Of the 10 clusters, hepatocytes near the periportal vein showed the

209 largest number of significant associations (151 pairs). This was followed by hepatic stellate cells
210 (HSC) near the periportal vein clusters (HSC 1, 58 pairs) and hepatocytes near the central vein (52
211 pairs). In contrast, interzonal hepatocytes showed only seven pairs, suggesting a preferential
212 involvement of perivascular hepatocytes in the genetic regulation of the circulating analyzed
213 metabolic biomarkers.

214

215 **Associations between blood metabolic biomarker levels and disease risk**

216 Leveraging metabolomic and clinical data from BBJ1-180K, we investigated the relationships
217 between 248 metabolic biomarker levels measured at the blood sampling date (baseline) and the
218 incidence of 19 diseases with at least 20 events after blood sampling. Using Cox proportional hazards
219 models adjusted for sex (excluding prostate cancer) and age at blood sampling, we estimated hazard
220 ratios (HRs) for 4,701 pairs (248 metabolic biomarkers and 19 disease onsets). As a result, 566 pairs
221 (216 metabolic biomarkers and 11 diseases) showed significant associations after Bonferroni
222 correction (P value $< 0.05/(20 \times 19)$; see Methods) (Supplementary Table 14).

223 To replicate these 566 pairs, we examined the estimated HRs of 122 matched biomarker-disease pairs
224 in UKB[10]. Of these, 64 pairs were also significant in the same direction (P value $< 0.05/122$ in UKB,
225 after Bonferroni correction) (Supplementary Table 15), with a high Pearson correlation coefficient (r
226 $= 0.95$). By disease, we replicated 39 pairs for myocardial infarction, 17 pairs for ischemic stroke,
227 three pairs for rheumatoid arthritis, one pair for lung cancer, and one pair for cataract. Of the 17
228 replicated metabolic biomarkers for ischemic stroke, 16 (high density (HDL)-related biomarkers,
229 albumin, and histidine) were also replicated for myocardial infarction (Supplementary Fig. 2),
230 suggesting that these metabolic biomarkers play a broad role in cardiovascular diseases in both
231 populations.

232 Finally, to examine cross-ancestry disease prediction based on baseline metabolic biomarker levels,
233 we performed an IVW meta-analysis of 2,971 matched pairs (248 metabolic biomarkers and 12
234 diseases) across BBJ1-180K and UKB. This analysis identified 891 significant pairs (237 metabolic

235 biomarkers and 11 diseases) (P value < 0.05/1000, as defined in a previous study[10]), including 47
236 additional associations (42 metabolic biomarkers and 10 diseases) that were not significant in either
237 BBJ1-180K or UKB alone (Supplementary Table 15).

238

239 **Prediction of disease onsets using metabolomic scores**

240 To evaluate the predictive ability of multiple metabolic biomarkers for disease risk in BBJ1-180K,
241 we constructed disease-specific metabolomic scores for five diseases (ischemic stroke, myocardial
242 infarction, lung cancer, cirrhosis, and colorectal cancer). These scores were derived by integrating the
243 baseline metabolomic data from BBJ1-180K with the pre-determined weights of 36 metabolite levels
244 for each disease derived from UKB11. We then assessed the utility of these scores over a 10-year
245 follow-up period in BBJ1-180K. A log-rank test revealed significant differences in disease risk
246 between individuals in the top 10% ranked by metabolomic score (the high-risk group) and the
247 remaining 90% for myocardial infarction and cirrhosis (P value < 0.05/5, after Bonferroni correction)
248 (Fig. 4a and Supplementary Table 16). Using Cox proportional hazard models adjusted for sex and
249 age at blood sampling as covariates, we estimated HRs for the high-risk group, showing that cirrhosis
250 and myocardial infarction had significant HRs (HR for myocardial infarction = 2.5, HR for cirrhosis
251 = 8.2, P value < 0.05/5, Supplementary Fig. 3 and Supplementary Table 17).

252 To compare our findings with those from the predominantly European populations using four-year
253 follow-up data, we repeated the analysis with the same follow-up period. We observed consistent
254 results: the high-risk group had significantly higher risks of myocardial infarction and cirrhosis (P
255 value < 0.05/5, after Bonferroni correction) (Supplementary Table 18). When comparing HRs
256 between our study and predominantly European cohorts, we revealed significant heterogeneity for
257 ischemic stroke and lung cancer (Cochran's Q test P value < 0.05/5, after Bonferroni correction) but
258 no heterogeneity for myocardial infarction and cirrhosis (Fig. 4b and Supplementary Table 18).

259 Next, we compared the predictive performance of metabolomic scores and polygenic risk scores
260 (PRS) for disease risk. PRS were calculated based on GWAS results in BBJ1-180K (samples lacking

metabolic biomarker measurement) and evaluated using Cox proportional hazard models adjusted for sex, age, and the top 10 principal components of genetic ancestry. For myocardial infarction, the PRS was significantly associated with disease risk (HR of PRS for myocardial infarction = 3.3, P value < 0.05/5, after Bonferroni correction) (Supplementary Table 19) and showed a higher HR than the metabolomic score. We further compared two models: model 1, which incorporated covariates and either the metabolomic score or PRS, and model 2, which incorporated covariates plus both the metabolomic score and the PRS simultaneously. The HRs estimated from model 2 were similar to those from model 1 (Supplementary Fig. 4 and Supplementary Table 20). Metabolomic scores showed significant HRs for cirrhosis and myocardial infarction, while the PRS showed a significant HR only for myocardial infarction. To evaluate the predictive ability, we calculated the concordance index (C-index) for four models: (i) base (age and sex only), (ii) base + metabolomic score, (iii) base + PRS, and (iv) base + metabolomic score + PRS. For cirrhosis and lung cancer, the metabolomic score alone demonstrated the highest predictive value. For myocardial infarction, ischemic stroke and colorectal cancer, the model that combined both metabolomic score and PRS had the highest C-index (Fig. 4c and Supplementary Table 21).

Discussion

In this large-scale GWAS of 248 blood metabolic biomarkers, we identified eight novel associations in the Japanese population and 28 additional associations in the largest cross-ancestry meta-analysis conducted to date. Statistical fine-mapping prioritized 47 putative causal variants with potential functional consequences, predicted to either alter protein function (n = 24) or regulate gene expression via enhancer elements (n = 23). Integrative gene-level analyses further underscored perivascular hepatocytes as the most genetically relevant cell type contributing to inter-individual variation in circulating metabolic biomarker levels. Moreover, integrating metabolomic and clinical data from BBJ1-180K and comparing our findings with those from primarily European populations showed that metabolomic scores have predictive potential for myocardial infarction and cirrhosis in the Japanese population. Collectively, our study uncovers key genetic variants and cell types that shape inter-individual variation in circulating metabolite levels and provides an atlas linking metabolic biomarkers to disease risk in the Japanese population.

We identified 32 metabolic biomarkers, such as docosahexaenoic acid, ApoA1, phospholipids in small VLDL, cholesteryl esters in small HDL, and LDL size, that were associated with the *CD36* region. Of these, 31 were reported in previous studies, and the association with LDL size represented one of the novel associations identified by this study and was replicated in an independent cohort. Fine-mapping and functional annotation pinpointed rs75326924 in *CD36*, known to cause CD36 deficiency, as the putative causal variant for 15 metabolic biomarkers, including LDL size. CD36 expressed on macrophages recognizes oxidized LDL, a modified form of LDL that tends to form smaller particles[29]. A previous study[30] reported that macrophages derived from CD36-deficient patients exhibited approximately a 50% reduced oxidized LDL uptake compared to those from healthy individuals. Our results demonstrated that rs75326924 was significantly associated with reduced LDL particle size, potentially attributable to impaired uptake of oxidized LDL due to CD36 deficiency.

In line with previous studies[31,32], our gene-level analysis showed that the liver is the tissue most genetically associated with metabolic biomarkers, with associations observed for 222 out of 248

304 biomarkers assessed in this study. To further dissect this relationship, we focused on liver cell types
305 and found that hepatocytes near the periportal and central veins showed significantly greater
306 enrichment than those in the interzonal region. The lack of enrichment in interzonal hepatocytes
307 represents a novel observation. Supporting our findings, a previous study[33] showed that single-cell
308 disease relevance scores for nine laboratory values, including total LDL and HDL cholesterol, were
309 positively associated with the scores for hepatocytes in both periportal and pericentral regions,
310 whereas our analysis targeted a broader range of metabolic biomarkers and provided a more
311 detailed characterization of lipoproteins in terms of their particle size and components. In adult
312 patients, metabolic dysfunction-associated steatotic liver disease (MASLD), formerly known as
313 non-alcoholic fatty liver disease, has been reported frequently as a liver disease that involves
314 zonation[34]. When comparing metabolic biomarkers associated with the hepatocytes in the periportal
315 and pericentral regions, 24 metabolic biomarkers, including three fatty acids (ratio of omega-3 fatty
316 acids to total fatty acids, ratio of omega-6 fatty acids to omega-3 fatty acids, and ratio of saturated
317 fatty acids to total fatty acids), and two amino acids (glycine and glutamine), showed significant
318 genetic associations exclusively in the pericentral region. Fatty acids are well-known risk factors of
319 MASLD[35], and glycine and glutamine, key substrates for glutathione synthesis, were also reported
320 to be associated with MASLD development[36]. Together, our findings reveal that the genetic
321 architecture of blood metabolic biomarkers is more tightly linked to perivascular hepatocytes than
322 interzonal ones and provide evidence that some metabolic biomarkers, known to be related to the
323 development of MASLD, are genetically associated with the hepatocytes in the pericentral region.

324 By integrating metabolomic profiles with longitudinal clinical information from BBJ1-180K, we
325 identified 566 biomarker-disease onset pairs with predictive potential. Of the 122 testable pairs, 64
326 were successfully replicated in UKB. Inverse associations between cardiovascular disease onset and
327 HDL cholesterol or albumin have been well documented in previous studies[37–39] and were
328 replicated in the present study. While the protective role of histidine in cardiovascular disease onset
329 has previously been reported only in the European or African populations[10,40,41] to our knowledge,
330 its association with reduced risk of myocardial infarction was consistently observed in both the
331 BBJ1-180K and UKB studies. Histidine is an essential amino acid with antioxidant and anti-

332 inflammatory properties[42], and its pivotal physiological functions may contribute to a
333 cardioprotective effect across diverse ancestral populations.

334 We evaluated the predictive ability of metabolomic scores for five disease risks in BBJ1-180K
335 using pre-determined weights derived from UKB metabolomic data. Our results showed significant
336 HRs for cirrhosis and myocardial infarction, suggesting that the prediction model developed in the
337 European population remains effective in the Japanese population. Notably, the metabolomic score
338 outperformed the PRS in predicting the risk of cirrhosis. In contrast, we observed substantial
339 heterogeneity in HRs for ischemic stroke and lung cancer compared with the corresponding
340 European results, whereas a previous study12 reported no such heterogeneity for these two diseases
341 across three European cohorts. The heterogeneity identified here may reflect differences across
342 ancestral populations, such as the subtype distribution for ischemic stroke and lung cancer[43,44].

343 A primary limitation of our study is the relatively small number of metabolic biomarkers analyzed.
344 Although NMR spectroscopy is relatively cost-effective and provides consistent quantification, it
345 captures fewer traits than MS, potentially underrepresenting the broader human metabolome,
346 particularly lipid species. Furthermore, among 14,266 95% CSs identified by cross-ancestry fine-
347 mapping, we found that 3099 CSs (21.7%) did not contain a variant with PIP > 0.5 of being
348 ancestry-specific or shared. This likely reflects the smaller sample size in BBJ-180K (N = 37,969)
349 compared to that in the UKB-EBB Europeans (N = 599,249). Therefore, although this study, the
350 largest metabolomic GWAS in East Asian populations to date, contributes to findings across
351 ancestral populations, further expansion of sample sizes in non-European populations is required to
352 more finely understand the genetic architecture of metabolic biomarkers.

353 In summary, by integrating metabolomic, genomic, and clinical data from BBJ1-180K and
354 comparing our findings with those from primarily the European population, we have elucidated the
355 genetic architecture of metabolic biomarkers and demonstrated their predictive potential for disease
356 onset across ancestral populations.

357

358 ## Methods

359 ### Study participants

360 All samples used in this study were collected from BBJ, a multicenter, hospital-based cohort in
361 Japan. From fiscal year 2003 to 2007, BBJ collected DNA, serum, and clinical information from
362 approximately 200,000 patients in 66 hospitals across 12 medical institutions as the first cohort
363 (BBJ1). We obtained informed consent from all participants by following protocols approved by
364 their institutional ethical committees. We obtained approval from the ethics committee of the
365 Institute of Medical Sciences, The University of Tokyo (approval number 2023-77-0118). We have
366 complied with all of the relevant ethical regulations. Serum metabolic biomarkers were quantified in
367 a total of 45,270 samples from BBJ1. BBJ1 comprised two studies using different genotyping
368 arrays, BBJ1-180K (approximately 180,000 individuals) and BBJ1-12K (approximately 12,000
369 individuals), as described in our previous paper[45]. No individuals overlapped between the two sub-
370 cohorts (see Genotyping and genomic imputation). All study participants were diagnosed by
371 physicians at the cooperating hospitals with at least one of the 47 target diseases[46]. Among these
372 diseases was amyotrophic lateral sclerosis (ALS); however, none of the subjects included in the
373 current study were diagnosed with ALS (Supplementary Table 22). Clinical information was
374 collected annually through interviews and medical records up to fiscal year 2013.[46]

375

376 ### Metabolic biomarkers quantification, QC, and preprocessing

377 Blood samples were collected at the collaborating hospitals of BBJ and stored at -80 °C. Samples
378 were transferred to BBJ in a cooling container with dry ice and stored at -150 °C in BBJ. Samples
379 were thawed at 4 °C for aliquoting into 100 μL and then frozen at -80 °C in BBJ before analysis.
380 The serum samples were shipped to Nightingale Health laboratories in Tokyo in 96-well plates on
381 dry ice in batches of 10,000–20,000 samples. Samples were stored in a freezer at -80 °C at
382 Nightingale Health laboratories after arrival from BBJ laboratory. Before preparation, frozen
383 samples were slowly thawed at +4 °C overnight, and then mixed gently and centrifuged (3 min,

384 3400 × g, +4 °C) to remove possible precipitate. Aliquots of each sample were transferred into 3-
385 mm outer-diameter NMR tubes and mixed in 1:1 ratio with a phosphate buffer (75 mM Na2HPO4
386 in 80%/20% H2O/D2O, pH 7.4, including also 0.08% sodium 3-(trimethylsilyl) propionate-2,2,3,3-
387 d4 and 0.04% sodium azide) automatically with an automated liquid handler (PerkinElmer Janus
388 Automated Workstation)[10]. The samples were measured using a 500 MHz NMR spectrometer
389 (Bruker AVANCE IIIHD) by Nightingale Health Japan. Nightingale Health's proprietary software
390 (quantification library 2020) was used to quantify 250 serum metabolic biomarker levels of 45,270
391 samples from BBJ1 (Supplementary Table 1). These metabolic biomarkers were classified into 108
392 directly measured metabolites (non-derived biomarkers) and 142 biomarkers, sum and/or ratio of
393 the non-derived biomarkers (derived biomarkers), following the protocol described in a previous
394 study[47] (Supplementary Table 1).

395 We performed QC on 45,147 samples from BBJ1-180K (N = 40,659) and BBJ1-12K (N = 4,488).
396 First, we excluded glycerol and 3-hydroxybutyrate, which showed missing rates exceeding 90%
397 across BBJ1-180K and BBJ1-12K, leaving 248 biomarkers for the present study.[47] We excluded
398 samples if they lacked age information, were under 18 years old, lacked imputation data (due to
399 failure in genotype QC or overlap with the imputation reference panel), or had a diagnosis date for
400 any of the initial 47 target diseases more than six months after the blood sampling date. Samples
401 with over 90% missing biomarker values were also excluded. If individuals had multiple serum
402 measurements, only the sample corresponding to the earliest blood sampling date was retained.
403 Following these basic QC procedures applied across all analyses, we additionally applied analysis-
404 specific QC steps tailored for metabolomic GWAS and for the survival analysis as described below.

405 Metabolomic GWAS was conducted after accounting for potential confounding by disease status or
406 medication use. We reviewed clinical information recorded within six months before or after blood
407 sampling for conditions or medications listed in the exclusion criteria (Supplementary Table 23).
408 Following a previously proposed QC protocol[47], we first estimated and adjusted the impact of
409 technical variability (metabolomic measurement batches and time from loading to measurement)
410 during the quantification of non-derived biomarkers in each study ($N_{BBJ1\text{-}180K}$ = 40,764; $N_{BBJ1\text{-}12K}$ =
411 4,495). We then recalculated derived biomarkers using the R package ukbnmr (v.2.2.1). Metabolic

412 biomarker values beyond ± 4 standard deviations from the mean were set to missing. These steps
413 resulted in 37,969 samples for BBJ1-180K and 3,528 samples for BBJ1-12K.

414 For survival analysis between metabolic biomarker(s) and disease risk, we excluded samples whose
415 initial date of interviews or medical records was more than six months after the blood sampling
416 date, or who were diagnosed with the target or related diseases (Supplementary Table 24) before or
417 within six months after blood sampling. For the analysis of prostate cancer, we used a subset
418 comprising only men. When building metabolomic scores for disease prediction, we excluded
419 samples with any missing values among the 36 metabolic biomarkers and used the remaining
420 samples to calculate metabolomic scores. These biomarkers have been clinically validated and
421 certified for diagnostic use in the NMR metabolomics assay to facilitate rapid clinical translation[10].
422 Following a previous study[11], we estimated a physiological glucose concentration based on glucose
423 and lactate (adjusted glucose concentration = glucose concentration + 0.38 × lactate concentration)
424 to account for potential glucose degradation during sample preparation. Afterward, metabolic
425 biomarker values exceeding ± 4 SDs from the mean were treated as missing. Metabolic biomarker
426 levels were subsequently log1P-transformed and standardized to Z-scores (mean = 0 and variance =
427 1).

428

429 **Genotyping and genomic imputation**

430 All BBJ1-180K samples were genotyped using Illumina HumanOmniExpressExome BeadChip or a
431 combination of Illumina HumanOmniExpress and HumanExome BeadChips. BBJ1-12K samples
432 were genotyped using Infinium Asian Screening Array BeadChip. QC for the genome data of samples
433 in each cohort followed the protocol described in a previous study[48]. For sample-level QC, individuals
434 with a call rate of less than 98% were removed. After removing duplicates and twins within each
435 cohort based on estimated identical-by-descent (IBD) sharing and retained only the samples with the
436 highest call rate, we re-estimated IBD sharing to confirm no sample overlapping between the two sub-
437 cohorts (BBJ1-180K and BBJ1-12K). For genotype QC, we excluded variants with a genotype call

rate < 99%, heterozygous counts < 5, Hardy-Weinberg equilibrium P value < 1.0×10^{-6}, or for BBJ1-180K alone, a concordance rate < 99.5% or a non-reference discordance rate ≥ 0.5% between array genotypes and whole-genome sequence data in the same individuals (N = 939). Pre-phasing was performed using Eagle (v2.4.1)[49]. Imputation was conducted using a reference panel constructed from whole-genome sequence data of Japanese ancestry (n = 3,256) combined with the 1KGP3 version 5 (n = 2,504)[50]. All genomic coordinates are in hg19 (GRCh37), except for the VEP downstream analysis. The imputation was carried out using Minimac4 (v1.0.2; https://github.com/statgen/Minimac4). Variants with an imputation quality R^2 < 0.7 or MAF < 0.1% were excluded.

Metabolomic GWAS using samples from BBJ1-180K

After applying an inverse rank normal transformation to the post-QC metabolic biomarker levels, we performed a GWAS using the linear mixed model implemented in BOLT-LMM (v2.4.1)[15]. The model included age at blood sampling, sex, the top 10 principal components of genetic ancestry, and the presence or absence of 46 registered diseases developed before or within six months after blood sampling as covariates. For the discovery GWAS, we used samples from BBJ1-180K (N=37,969). We calculated minus $\log_{10}$ (P value) based on χ^2.

Principal component analysis of metabolic biomarker levels

Using the Python package scikit-learn (v.0.24.2), we performed principal component analysis (PCA) on the inverse rank-normalized levels of 248 metabolic biomarkers in 9,444 samples from the QC-processed BBJ1-180K that had no missing values. Default parameters were used for the analysis. The top 20 principal components cumulatively explained 95% of the variance in the 248 metabolic biomarkers. Considering that this study was conducted on multiple correlated metabolic biomarkers, we adjusted the conventional genome-wide significance level (5×10^{-8}) by these 20 principal

463 components, yielding a study-wide corrected significance level[51].

464

465 **Functional annotation of variants by VEP**

466 To take advantage of more up-to-date annotations, we lifted over the coordinates of the GWAS
467 variants from hg19 to hg38 using the Python package gwaslab (v.3.5.1;
468 https://github.com/Cloufield/gwaslab)[52] and then annotated the variants with VEP
469 (https://useast.ensembl.org/info/docs/tools/vep/index.html)[53] version 99, using GENCODE v49 on
470 hg38. We adopted the Ensembl severity ranking
471 (https://asia.ensembl.org/info/genome/variation/prediction/predicted_data.html)[31] and defined
472 functional variants as those containing any of the following terms: splice_acceptor_variant,
473 splice_donor_variant, stop_gained, frameshift_variant, stop_lost, start_lost, transcript_amplification,
474 inframe_insertion, inframe_deletion, missense_variant, protein_altering_variant,
475 splice_region_variant.

476

477 **Definition of associated regions and novel associations**

478 We identified variants with *P* value below the study-wide corrected significance level for each
479 metabolic biomarker and defined 1 Mb windows centered on these variants as associated loci. To
480 capture all associated regions across metabolic biomarkers, we merged partially overlapping windows
481 to eliminate redundancy. Within each merged region, the variant showing the strongest association
482 with a given metabolic biomarker was designated the lead variant. Among these lead variants, the one
483 with the smallest *P* value across all associated biomarkers was selected as the cross-metabolic
484 biomarkers' lead variant for the region (Supplementary Fig. 4).

485 To name each region, we first examined whether the cross-metabolic biomarker lead variant and
486 variants in LD with it ($r^2 > 0.6$) were annotated as functional variants. We assessed LD in East Asians
487 from 1KGP3 for BBJ1-180K GWAS results and LD in both East Asians and Europeans from 1KGP3

488 for cross-ancestry meta-analysis results. If functional variants were identified, we adopted the gene(s)
489 affected by these variants as the region name. When multiple functional variants were available, we
490 selected the gene associa the variant that had the highest Combined Annotation-Dependent Depletion
491 score (v.1.6)[54]. If no functional variant was identified, we used the name(s) of the closest gene(s) of
492 the lead variant. A lead variant for a metabolic biomarker was considered a novel association when it
493 was located more than 1 Mb away from any previously reported associated variants (*P* value < $5\times10^{-}$
494 ^{8} or stronger study-specific thresholds). We consulted the GWAS catalog (v.2024-12-19)[55] and
495 additional studies not listed in the GWAS catalog (six to 200 studies for each GWAS; Supplementary
496 Table 25), including metabolomic GWAS results from Europeans[7,56–58] and East Asians[6,7,59–61]. The
497 experimental factor ontology (EFO) IDs were used to comprehensively map the same or similar
498 phenotypes in the GWAS catalog.

499

500 **Replication for novel associations**

501 In BBJ-12K, we tested variant-metabolic biomarker associations that were identified as novel in either
502 the GWAS in BBJ1-180K or the cross-ancestry meta-analysis, using a linear model implemented in
503 PLINK2 (v2.00a3LM; https://www.cog-genomics.org/plink/2.0/)[62] after excluding related individuals
504 (N = 3,522). To evaluate these novel associations, we performed a sign test using replication results
505 from BBJ-12K for associations identified in BBJ1-180K, and using both BBJ-12K and an additional
506 previous large-scale cross-ancestry meta-analysis[3] for those identified in the present cross-ancestry
507 meta-analysis. This test evaluated whether the proportion of variants with matching effect directions
508 exceeded what would be expected by chance. We performed a binomial test to assess whether the
509 observed number of concordant effect directions in both studies was significantly greater than 50%.
510 A one-sided test *P* value < 0.05 was considered evidence of replication.

511

512 **Stepwise conditional analysis**

513 To identify independent signals within and around the associated region (± 500 kb), we performed a
514 stepwise conditional analysis. In each iteration, we regressed the metabolic biomarker on every variant
515 while adjusting for the most significant variant in the region (the index variant) and for the covariates
516 used in the original GWAS. The procedure was repeated until no variant reached the genome-wide
517 significance threshold (P value < 5×10^{-8}). This analysis identified 1,443 additional associations (213
518 index variants with P value < 5×10^{-8} and 236 metabolic biomarkers) (Supplementary Table 26).

519

520 **Fine-mapping in BBJ1-180K**

521 For each of the 3,583 independent signals identified by GWAS and stepwise conditional analysis,
522 excluding the MHC region (chromosome 6: 25-36 Mb), we defined a 2-Mb genomic region centered
523 on the lead variants or index variants. Overlapping regions for the same metabolic biomarker were
524 merged, resulting in 2,111 distinct regions ranging from 2.0 Mb to 3.9 Mb. To pinpoint causal variants,
525 we performed statistical fine-mapping using SuSiE-RSS (v.0.12.35)[16], allowing up to 10 CSs per
526 region. For genetic relationship inference, we used the KING framework implemented in PLINK2 to
527 compute the KING kinship coefficient. Individuals with a KING kinship coefficient >0.0884 were
528 excluded, resulting in a set of 37,288 unrelated individuals from the 37,969 used in this study within
529 BBJ1-180K. Using this dataset, we constructed the LD matrix using LD store2 (v.2.0;
530 http://www.christianbenner.com) for the fine-mapping with SuSiE-RSS.

531

532 **Nomination of genes for non-coding variants by the ABC model**

533 We used the genome-wide mapping of enhancer-gene connections in 131 human cell or tissue types
534 predicted by the ABC model. The dataset included the element-gene links with ABC scores >= 0.015
535 (https://mitra.stanford.edu/engreitz/oak/public/Nasser2021/AllPredictions.AvgHiC.ABC0.015.minus
536 150.ForABCPaperV3.txt.gz)[18]. When a putative non-coding causal variants falls within a regulatory

element, we assume that it influences enhancer activity and, consequently modulates the expression of the associated target genes.

Cross-ancestry meta-analysis

Using the metabolomic GWAS results from BBJ1-180K and the meta-analysis results of six ancestral populations from UKB-EBB, we conducted fixed-effects meta-analyses with the IVW method implemented in METAL (v.2011-03-25, https://csg.sph.umich.edu/abecasis/Metal/download/)[22]. Variants were retained only if they were present in both BBJ1-180K and UKB-EBB.

Comparison of genetic effect sizes across ancestral populations

Using GWAS results from BBJ1-180K, the European-specific meta-analysis from UKB-EBB, as well as the GWAS results from five ancestral populations (the Central/South Asian, African, East Asian, Middle Eastern, and Admixed American population) within the UKB, we investigated the correlations of the genetic effect sizes of extracted lead variants.

We first selected lead variants from BBJ1-180K with MAF > 1% in both BBJ1-180K and each of the six comparison studies from UKB or UKB-EBB, resulting in 1,638 associations for the European population, 1,680 for the Central/South Asian population, 1,660 for the African population, 1,846 for the East Asian population, 1,659 for the Middle Eastern population, and 1,642 for the Admixed American population (Supplementary Table 8).

Next, we defined associated regions across metabolic biomarkers based on the European-specific meta-analysis or GWAS results from each population. We again extracted lead variants with MAF > 1% in both BBJ1-180K and each comparison study, resulting in 32,187 associations for the European population, 749 for the Central/South Asian population, 300 for the African population, 241 for the

East Asian population, 18 for the Middle Eastern population, 6 for the Admixed American population (Supplementary Table 9).

We then calculated Pearson correlation coefficients to assess the concordance of genetic effect sizes between BBJ1-180K and each of the six ancestral populations from UKB-EBB or UKB.

Multi-ancestry fine-mapping

For the 2,111 regions constructed for fine-mapping with SuSiE-RSS, we performed multi-ancestry fine-mapping to identify causal variants across ancestries with MESuSiE (v.1.0)[24], using GWAS results from the East-Asian specific BBJ1-180K and the European-specific meta-analysis from UKB-EBB, and LD matrices from BBJ1-180K and the unrelated British-ancestry individuals from UKB (N = 337,491)[63]. MESuSiE models genetic effects under both ancestry-shared (non-zero effect in both ancestries) and ancestry-specific conditions and calculates a PIP for each variant in each condition. Following the previous study[24], we excluded variants with MAF < 0.1% in European-specific meta-analysis and strand-ambiguous variants in both studies. Variants were classified as shared causal variants if PIP > 0.5 in both East Asian and European populations, East Asian-specific causal variants if PIP > 0.5 in the East Asian population only, and European-specific causal variants if PIP >0.5 in the European population only. We allowed up to 10 signals per region.

Gene-level association analysis between metabolic biomarkers and tissue or cell types

Using the GWAS results from BBJ1-180K and the European-specific meta-analysis from UKB-EBB, we performed gene-based analysis using MAGMA (v.1.08) implemented in FUMA (v.1.6.2) with default parameters. We used the LD reference panel constructed from genomic data of the 1KGP3 EAS or EUR population that matched the target population. We performed MAGMA[26] gene-property analysis implemented in the SNP2GENE method of FUMA (v.1.6.2)[27] integrating the gene-based analysis results for 248 metabolic biomarkers in BBJ1-180K and UKB-EBB with the transcriptome

586 data from the GTEx consortium v.8 across 30 tissues[25]. Tissue-metabolic biomarker pairs with a one-
587 sided test P value < 0.05/(20×30) in each study were considered significant after Bonferroni correction,
588 where 20 represents the number of metabolic principal components explaining 95% of the variance
589 in the 248 metabolic biomarkers, and 30 is the number of the analyzed tissue types.

590 To identify cell types in the liver associated with the metabolic biomarkers, we obtained a publicly
591 available snRNA-seq dataset of the human liver[26]. In the snRNA-seq analysis, approximately 90,000
592 cells from three healthy human livers were analyzed and 10 cell populations (clusters) were identified
593 based on gene expression patterns. The top 10% highly associated genes compared with the other
594 clusters based on fold-change were considered cluster-type-specific gene sets. Subsequently, for the
595 222 metabolic biomarkers associated with the liver in the tissue enrichment analysis in UKB-EBB,
596 we performed gene-set enrichment analysis using MAGMA with the gene-based analysis results from
597 BBJ1-180K and UKB-EBB together with cluster-specific gene sets. We used 1KGP3 EAS and EUR
598 population datasets as reference panels. To increase statistical power, we performed an IVW meta-
599 analysis by combining statistics from BBJ1-180K and UKB-EBB. Cluster-metabolic biomarker pairs
600 with FDR < 5% were considered significant.

601

602 **Disease risk prediction by each metabolic biomarker**

603 We examined the associations between 248 metabolic biomarkers and 19 disease incidences, each
604 with at least 20 events, based on 34,933 samples from BBJ1-180K. Disease onset was defined as the
605 first recorded diagnosis date after blood sampling. We estimated HRs for the association between Z-
606 normalized metabolic biomarkers and disease incidence using Cox proportional hazards model,
607 adjusting for age at blood sampling and sex (except for prostate cancer, where only male samples
608 were analyzed). Disease-metabolic biomarker pairs with a P value < 0.05/(20×19) were considered
609 significant, where 20 is the number of metabolic principal components explaining 95% of the variance
610 in the 248 metabolic biomarkers and 19 is the number of analyzed diseases.

611 For comparisons and meta-analysis with results from the UKB results[10], we used the publicly available

summary statistics (https://nightingalehealth.com/atlas), which estimated HRs using Cox proportional hazards models adjusted for age at blood sampling, sex and UKB assessment centre. Diseases in BBJ and UKB were matched based on 10th version of International Classification of Diseases (ICD-10) codes (Supplementary Table 27). When a disease-biomarker pair in BBJ corresponded to multiple pairs in UKB, we retained the one with the smallest *P* value in UKB. We performed an IVW meta-analysis to combine results across cohorts.

Disease risk prediction by metabolomic score

Among the 12 previously analyzed diseases, we evaluated the predictive ability of metabolomic scores for the onset of five diseases (ischemic stroke, myocardial infarction, lung cancer, cirrhosis, and colorectal cancer) for which diagnosis dates were available in BBJ1-180K. Metabolomic scores were calculated by multiplying the Z-normalized metabolic biomarker levels in BBJ1-180K by the corresponding coefficients derived from UKB[11].

To obtain GWAS results for constructing the PRS models, we performed GWAS for five diseases using individuals without metabolic biomarker measurements in BBJ1-180K (N_{total} = 139,952) with SAIGE (v.1.1.9)[64] (Supplementary Table 28). For lung cancer and colorectal cancer, we used individuals without any cancers among BBJ target diseases as controls. We included age, age^2, sex, age × sex, age^2 × sex, and the top 10 genetic principal components as covariates. The genomic inflation factors λ_{GC} indicated no substantial inflation (λ_{GC} = 1.02-1.04). We employed PRS-CS to compute PRS models using HapMap3 SNPs with an EAS LD reference panel from the 1KG. Global shrinkage parameters were inferred from the data by PRS-CS using a fully Bayesian approach (PRS-CS-auto).

Individuals were stratified into the top 10% (high-risk group) and the remaining 90% ranked by their metabolomic score or PRS. Kaplan-Meier curves stratified by the score were estimated with the R package survival (v.3.5.7). Using Schoenfeld's residuals, we did not reject the proportional hazards assumption for any of the five diseases (*P* value > 0.05/5). We also estimated HRs for the

high-risk group within four or 10 years after blood sampling using the Cox proportional hazards model, adjusting for sex and age as covariates. Age at blood sampling was also Z-normalized. We compared the HRs estimated in this study with those from meta-analyses of UKB, EBB, and Finnish THL Biobank. In each study, HRs were estimated using Cox proportional hazards models, adjusting for age and sex as covariates using four years of follow-up data. Furthermore, we computed the C-index to evaluate the predictive ability of the following four models: (1) base (age + sex), (2) base + metabolomic score, (3) base + PRS, and (4) base + metabolomic score + PRS.

Data availability

Summary statistics for GWAS in BBJ1-180K and the cross-ancestry meta-analysis of 248 metabolic biomarkers are available in the GWAS catalog (accessions GCST90709281 - GCST90709776, Table S29). Summary statistics for ancestry-specific GWAS in UKB or UKB-EBB and the cross-ancestry meta-analysis across these studies are available in the GWAS catalog. Genotype and metabolic biomarker data in BBJ1-180K and BBJ1-12K were deposited at the NBDC Human Database (Genotype data in BBJ1-180K, research ID: hum0014 JGAS000114; Genotype data in BBJ1-12K and metabolic biomarker data, research ID: hum0311 JGAS000412 and JGAS000561). The imputation reference panel constructed using whole genome data of 3,256 Japanese individuals from BBJ and 2,504 individuals from 1KG was also deposited at the NBDC Human Database (research ID: hum0014 JGAS000746).

Code availability

In the present study, we used publicly available software and packages for bioinformatics analysis. The details were described in Methods.

661 **References**

662 1. Idle, J. R. & Gonzalez, F. J. Metabolomics. *Cell Metab* **6**, 348–351 (2007).

663 2. Shah, S. H. *et al.* High heritability of metabolomic profiles in families burdened
664 with premature cardiovascular disease. *Mol Syst Biol* **5**, 258 (2009).

665 3. Karjalainen, M. K. *et al.* Genome-wide characterization of circulating metabolic
666 biomarkers. *Nature* **628**, 130–138 (2024).

667 4. Yin, X. *et al.* Genome-wide association studies of metabolites in Finnish men
668 identify disease-relevant loci. *Nat Commun* **13**, 1644 (2022).

669 5. Richardson, T. G. *et al.* Characterising metabolomic signatures of lipid-
670 modifying therapies through drug target mendelian randomisation. *PLoS Biol* **20**,
671 e3001547 (2022).

672 6. Toki, R. *et al.* Genome-wide association study of plasma amino acids and
673 Mendelian randomization for cardiometabolic traits. *Sci Rep* **15**, 14569 (2025).

674 7. Lin, C. *et al.* Cross-ancestry analyses of Chinese and European populations
675 reveal insights into the genetic architecture and disease implication of
676 metabolites. *Cell Genomics* **5**, 100810 (2025).

677 8. Liu, S. *et al.* Genome-wide association study of maternal plasma metabolites
678 during pregnancy. *Cell Genomics* **4**, 100657 (2024).

679 9. Buergel, T. *et al.* Metabolomic profiles predict individual multidisease outcomes.
680 *Nat Med* **28**, 2309–2320 (2022).

681 10. Julkunen, H. *et al.* Atlas of plasma NMR biomarkers for health and disease in
682 118,461 individuals from the UK Biobank. *Nat Commun* **14**, 604 (2023).

683 11. Barrett, J. C. *et al.* Metabolomic and genomic prediction of common diseases in
684 700,217 participants in three national biobanks. *Nat Commun* **15**, 10092 (2024).

685 12. Yamamoto, Y. *et al.* Dissecting cross-population polygenic heterogeneity across
686 respiratory and cardiometabolic diseases. *Nature Communications* **16**, (2025).

687 13. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Current Opinion in*
688 *Biotechnology* vol. 43 34–40 Preprint at
689 https://doi.org/10.1016/j.copbio.2016.08.001 (2017).

690 14. Yang, J. J. *et al.* Lung Cancer Risk Prediction Models for Asian Ever-Smokers.
691 *Journal of Thoracic Oncology* **19**, 451–464 (2024).

692 15. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association
693 power in large cohorts. *Nat Genet* **47**, 284–290 (2015).

694 16. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary
695 data with the 'Sum of Single Effects' model. *PLoS Genet* **18**, (2022).

696 17. Hanawa, H. *et al.* Identification of cryptic splice site, exon skipping, and novel
697 point mutations in type I CD36 deficiency. *J Med Genet* **39**, 286 (2002).

698 18. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes.
699 *Nature* **593**, 238–243 (2021).

700 19. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation
701 from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664–1669 (2019).

702 20. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–
703 74 (2015).

704 21. Brosnan, M. E. & Brosnan, J. T. Histidine metabolism and function. *Journal of*
705 *Nutrition* **150**, 2570S-2575S (2020).

706 22. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis
707 of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

708 23. Tambets, R. *et al.* Genome-wide association study for circulating metabolites in
709 619,372 individuals. Preprint at https://doi.org/10.1101/2024.10.15.24315557
710 (2024).

711 24. Gao, B. & Zhou, X. MESuSiE enables scalable and powerful multi-ancestry fine-
712 mapping of causal variants in genome-wide association studies. *Nat Genet* **56**,
713 170–179 (2024).

714 25. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects
715 across human tissues. *Science (1979)* **369**, 1318–1330 (2020).

716 26. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA:
717 Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* **11**, e1004219
718 (2015).

719 27. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional
720 mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**,
721 1826 (2017).

722 28. Watson, B. R. *et al.* Spatial transcriptomics of healthy and fibrotic human liver at
723 single-cell resolution. *Nat Commun* **16**, 319 (2025).

724 29. Chait, A., Brazg, R. L., Tribble, D. L. & Krauss, R. M. Susceptibility of small,
725 dense, low-density lipoproteins to oxidative modification in subjects with the
726 atherogenic lipoprotein phenotype, pattern B. *Am J Med* **94**, 350–356 (1993).

727 30. Janabi, M. *et al.* Oxidized LDL-Induced NF-B Activation and Subsequent
728 Expression of Proinflammatory Genes Are Defective in Monocyte-Derived

Macrophages From CD36-Deficient Patients. *Arterioscler Thromb Vasc Biol* **20**, 1953–1961 (2000).

31. Ottensmann, L. *et al.* Genome-wide association analysis of plasma lipidome identifies 495 genetic associations. *Nat Commun* **14**, 6934 (2023).

32. van der Meer, D. *et al.* Pleiotropic and sex-specific genetic mechanisms of circulating metabolic markers. *Nature Communications* **16**, (2025).

33. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat Genet* **54**, 1572–1580 (2022).

34. Brunt, E. M. Pathology of nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* **7**, 195–203 (2010).

35. Allard, J. P. *et al.* Nutritional assessment and hepatic fatty acid composition in non-alcoholic fatty liver disease (NAFLD): A cross-sectional study. *J Hepatol* **48**, 300–307 (2008).

36. Rom, O. *et al.* Glycine-based treatment ameliorates NAFLD by modulating fatty acid oxidation, glutathione synthesis, and the gut microbiome. *Sci. Transl. Med* **12**, 2841 (2020).

37. Linsel-Nitschke, P. & Tall, A. R. HDL as a target in the treatment of atherosclerotic cardiovascular disease. *Nat Rev Drug Discov* **4**, 193–205 (2005).

38. Bucci, T. *et al.* Albumin Levels and Risk of Early Cardiovascular Complications after Ischemic Stroke: A Propensity-Matched Analysis of a Global Federated Health Network. *Stroke* **55**, 604–612 (2024).

39. Djoussé, L., Rothman, K. J., Cupples, L. A., Levy, D. & Ellison, R. C. Serum albumin and risk of myocardial infarction and all-cause mortality in the framingham offspring study. *Circulation* **106**, 2919–2924 (2002).

40. Jauhiainen, R. *et al.* The Association of 9 Amino Acids with Cardiovascular Events in Finnish Men in a 12-Year Follow-up Study. *Journal of Clinical Endocrinology and Metabolism* **106**, 3448–3454 (2021).

41. Yu, B. *et al.* Association of Rare Loss-Of-Function Alleles in HAL, Serum Histidine: Levels and Incident Coronary Heart Disease. *Circ Cardiovasc Genet* **8**, 351–355 (2015).

42. Wade, A. M. & Tucker, H. N. Antioxidant characteristics of L-histidine. *J Nutr Biochem* **9**, 308–315 (1998).

43. Zhang, Y. *et al.* Global variations in lung cancer incidence by histological subtype in 2020: a population-based study. *Lancet Oncol* **24**, 1206–1218 (2023).

763 44. Bang, O. Y. Considerations when subtyping ischemic stroke in Asian patients.
764 *Journal of Clinical Neurology (Korea)* **12**, 129–136 (2016).

765 45. He, Y. *et al.* East Asian-specific and cross-ancestry genome-wide meta-analyses
766 provide mechanistic insights into peptic ulcer disease. *Nat Genet* **55**, 2129–2138
767 (2023).

768 46. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and
769 profile. *J Epidemiol* **27**, S2–S8 (2017).

770 47. Ritchie, S. C. *et al.* Quality control and removal of technical variation of NMR
771 metabolic biomarker data in ~120,000 UK Biobank participants. *Sci Data* **10**, 64
772 (2023).

773 48. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated
774 variants in the Japanese population. *Nat Commun* **10**, 4393 (2019).

775 49. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference
776 Consortium panel. *Nat Genet* **48**, 1443–1448 (2016).

777 50. Koyama, S. *et al.* Population-specific putative causal variants shape quantitative
778 traits. *Nat Genet* **56**, 2027–2035 (2024).

779 51. Lotta, L. A. *et al.* A cross-platform approach identifies genetic regulators of
780 human metabolism and health. *Nat Genet* **53**, 54–64 (2021).

781 52. He, Y., Koido, M., Shimmori, Y. & Kamatani, Y. GWASLab: a Python package
782 for processing and visualizing GWAS summary statistics. Preprint at
783 https://doi.org/https://doi.org/10.51094/jxiv.370 (2023).

784 53. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122
785 (2016).

786 54. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD:
787 Predicting the deleteriousness of variants throughout the human genome. *Nucleic*
788 *Acids Res* **47**, D886–D894 (2019).

789 55. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition
790 resource. *Nucleic Acids Res* **51**, D977–D985 (2023).

791 56. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci
792 influencing human serum metabolite levels. *Nat Genet* **44**, 269–276 (2012).

793 57. Smith, C. J. *et al.* Integrative analysis of metabolite GWAS illuminates the
794 molecular basis of pleiotropy and genetic correlation. *Elife* **11**, e79348 (2022).

795 58. Gallois, A. *et al.* A comprehensive study of metabolite genetics reveals strong
796 pleiotropy and heterogeneity across time and context. *Nat Commun* **10**, 4788
797 (2019).

59. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415–1424 (2021).

60. Kodate, S. *et al.* Simulating metabolic pathways to enhance interpretations of metabolome genome-wide association studies. *Sci Rep* **15**, (2025).

61. Iwasaki, T. *et al.* Genetic influences on human blood metabolites in the Japanese population. *iScience* **26**, 105738 (2023).

62. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

63. Weissbrod, O. *et al.* Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat Genet* **52**, 1355–1363 (2020).

64. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335–1341 (2018).

Acknowledgments

We thank Peter Würtz and Machiko Usui (Nightingale Health) for their valuable support, and we acknowledge Nightingale Health Japan Plc for providing the metabolic biomarker data through a collaboration with BBJ. We also thank Chikashi Terao, Kohei Tomizuka, and Satoshi Koyama for their generous support. We are also gratetful to all the participants and staff of the BBJ project. This study was supported by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of the Japanese government and the Japan Agency for Medical Research and Development (AMED) under grant Numbers JP18km0605001 / JP23tm0624002 (the BioBank Japan project), JP19km0405215, and JP223fa627011, and JST SPRING GX, Grant Number JPMJSP2108. The super-computing resource was provided by Human Genome Center, the Institute of Medical Science, the University of Tokyo (http://sc.hgc.jp/shirokane.html).

Ethics declarations

Competing interests

M.K. is a consultant for Takeda Pharmaceutical Co., Ltd. S.N. is an employee of Daiichi Sankyo Co., Ltd. This work was neither part of a collaboration with nor funded by Daiichi Sankyo. Y.K. holds stock in StaGen Co., Ltd. The other authors declare no competing interests.

Figures

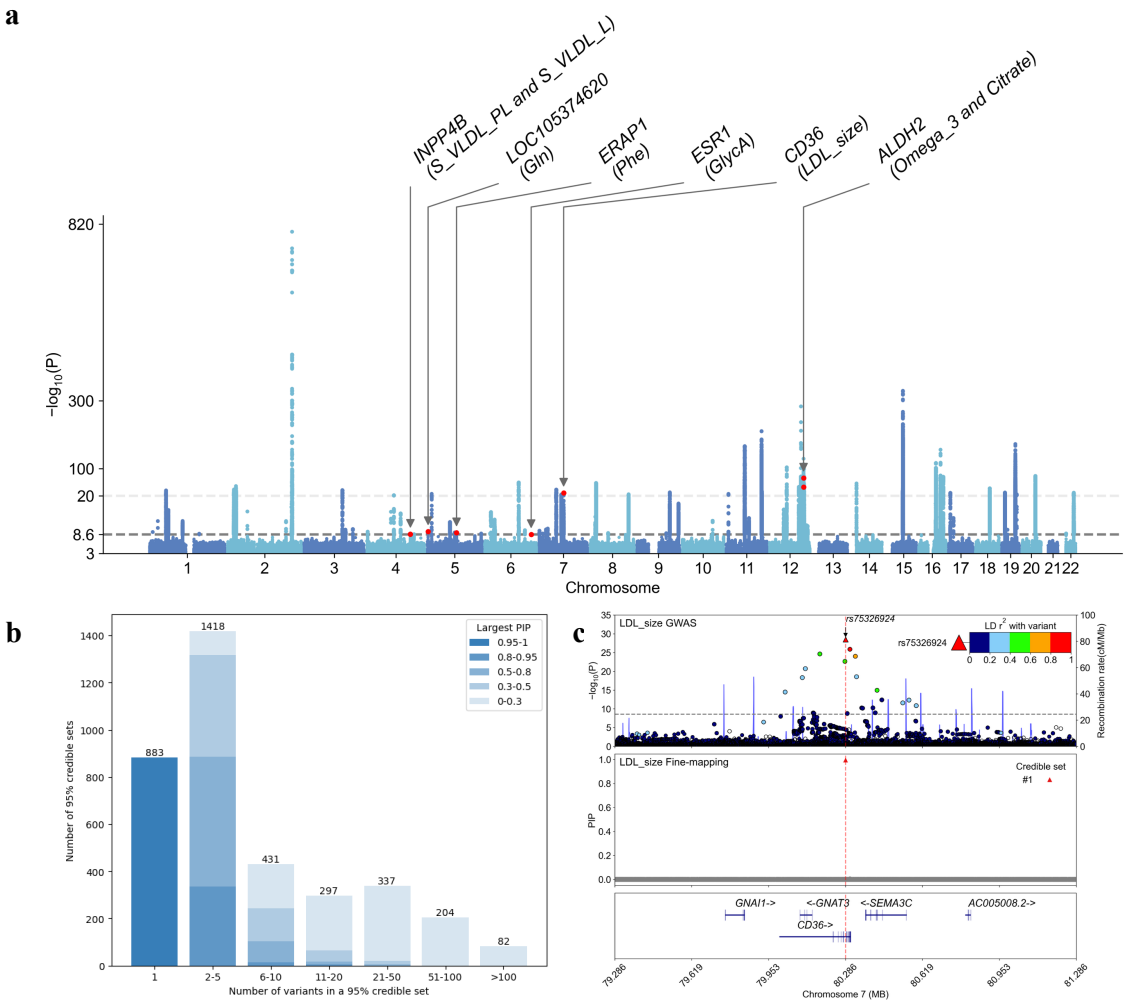

Fig. 1 GWAS results in BBJ1-180K

a, Manhattan plot of the metabolomic GWAS in BBJ1-180K. Association results for 248 metabolic biomarkers are over-laid into a single panel. For variants above the blue dashed line, values are rescaled. The gray dashed line denotes the study-wide significance threshold (*P* value < $5 \times 10^{-8}/20$). Red dots mark variants that represent novel associations; each dot is annotated with the region name and corresponding metabolic biomarker. Abbreviations of metabolic biomarkers are listed in Supplementary Table 1. **b**, Number of variants and the maximum PIP within 3,652 95% credible sets identified by fine-mapping. Each bin is colored according to the number the largest PIP in the credible set (The larger PIP, blue is darker.). **c**, Regional plot of the *CD36* locus for LDL size from GWAS (top) and fine-mapping (bottom) results in BBJ1-180K. In the top panel, the dotted line shows the study-wide significance threshold (*P* value < $5 \times 10^{-8}/20$). LD with rs75326924, calculated based on 1KGP3 EAS, is shown by color scale.

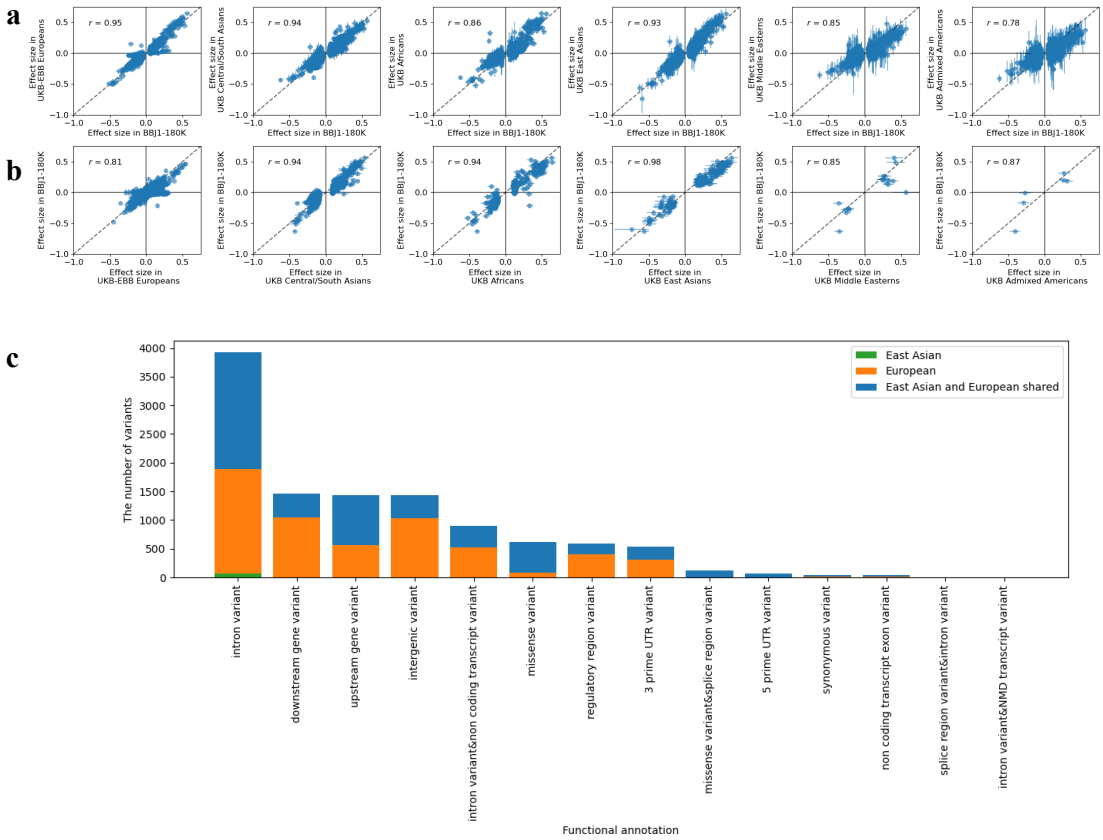

Fig. 2 Comparison of genetic effects on metabolic biomarkers across ancestral populations

a, **b** Effect size correlations for the lead variants identified in (**a**) BBJ1-180K or (**b**) UKB-EBB or UKB. For each panel the effect sizes in BBJ1-180K are plotted against those in the six ancestral populations of UKB. *r*: Pearson's correlation. **c,** Stacked bar plot of 11,167 putative causal variants in either population or both populations across 14 types of functional annotations.

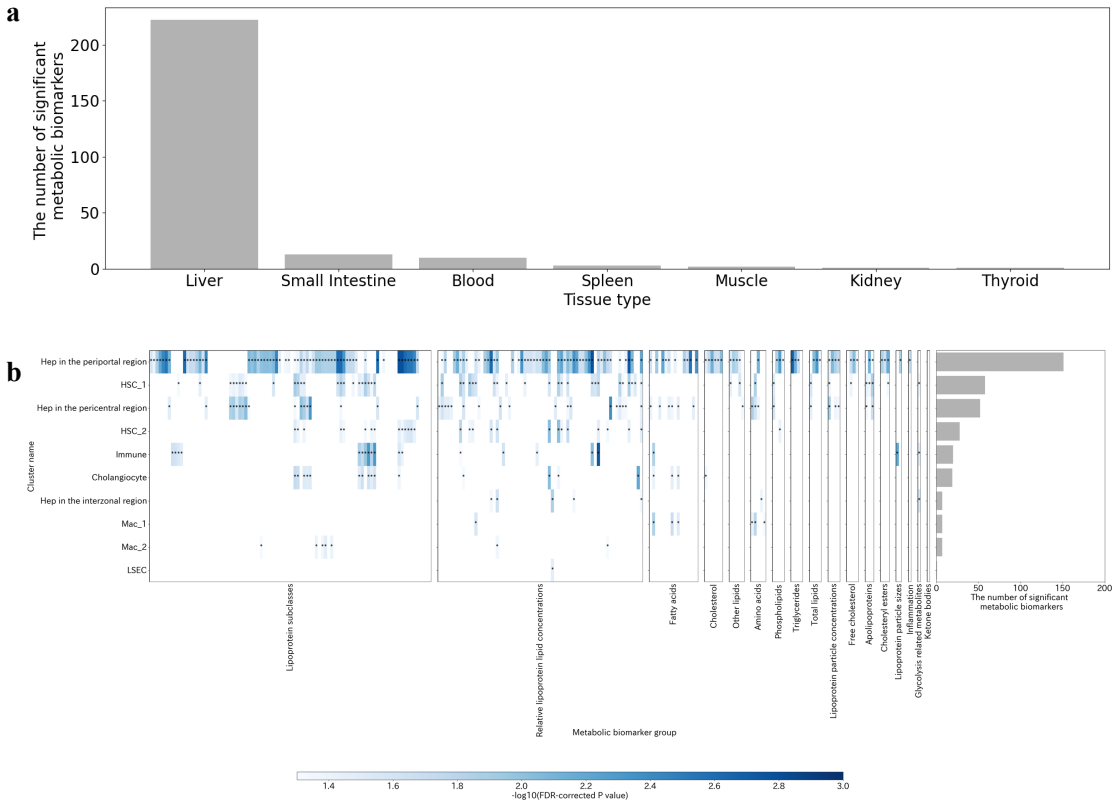

Fig. 3 Gene-level association landscape for metabolic biomarkers across tissues and cell types in the liver

a, Bar plot of the number of significant metabolic biomarkers (P value < 0.05 / (20x30)) for each tissue. Only the seven tissues out of the 30 whole-body tissues examined that yielded at least one significant biomarker are shown. **b**, Composite view of the liver cell type-biomarker analysis results. Left panel shows heat map of the -log10 corrected P-value for pairs of metabolic biomarker and each of the 10 cell types in the liver. Pairs with FDR < 5% (i.e., *) are colored blue according to their -$\log_{10}$ (FDR-corrected *P* value). Non-significant pairs are colored white. Right panel shows bar plot of the number of significant metabolic biomarkers (FDR < 5%) for each cell type in the liver. Abbreviations for metabolic biomarkers are described in Supplementary Table 1. Hep: hepatocytes, HSC: hepatic stellate cells, Mac: Macrophage, and LESC: Liver sinusoidal endothelial cells.

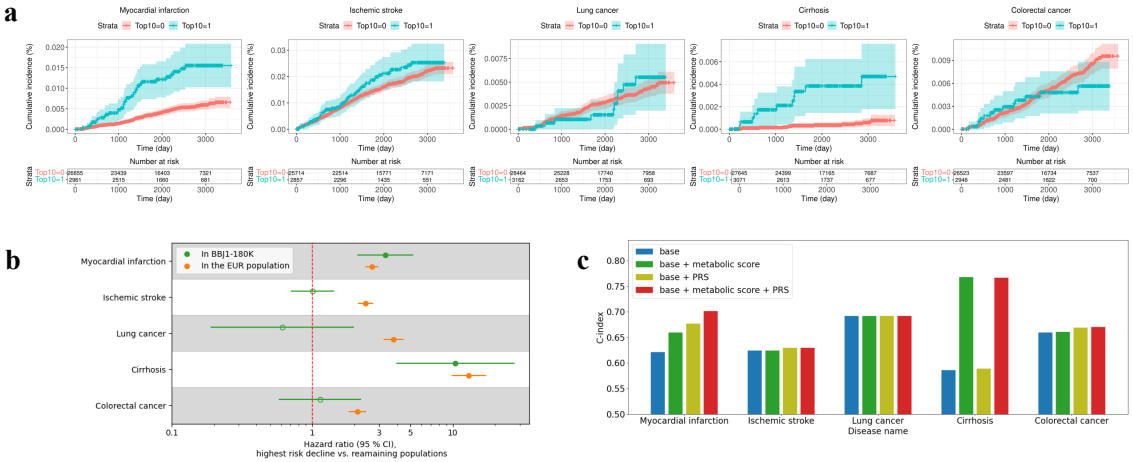

Fig. 4 Predictive performance of the metabolomic scores for five diseases

a, Kaplan-Meier curves comparing individuals in the top 10% high-risk group with the remaining 90% of the participants in BBJ1-180K. Top10: The top 10% high-risk. **b**, Four-year HRs for the metabolomic scores applied to the top 10% high-risk group for each of the five diseases. Each dot represents the point estimate of HR from Cox proportional hazards model, and the horizontal error bars denote the 95% confidence intervals. Dots are filled when the *P* value is < 0.05/5; otherwise, they are left unfilled. **c**, Comparisons of the C-index for five diseases using three models: base (= age + sex), base + metabolomic score, base + PRS model, and base + metabolomic score + PRS. 10-year follow-up data are used to calculate the C-index.

881 **Tables**

882 **Table 1 Eight novel associations identified by GWAS in BBJ1-180K**

Metabolite	rs ID	Chromosome	Position (GRCh37)	Region name	Effect allele/ noneffect allele	Beta	SE	*P* value	EAF	EAF_1KG_EUR
Total lipids in small VLDL	rs1497401	4	143321490	*INPP4B*	T/C	0.056	0.0092	2.00E-09	0.302	0.368
Phospholipids in small VLDL	rs1497401	4	143321490	*INPP4B*	T/C	0.056	0.0091	1.80E-09	0.302	0.368
Glutamine	rs56981040	5	2992414	*LOC105374620*	G/T	-0.061	0.0096	3.10E-10	0.828	0.974
Phenylalanine	rs70981840	5	96119579	*ERAP1*	A/ACT	0.044	0.0073	6.90E-10	0.378	0.147
Glycoprotein acetyls	rs560647736	6	152278333	*ESR1*	C/T	0.64	0.11	2.30E-09	0.999	1
Average diameter for LDL particles	rs75326924	7	80286003	*CD36*	C/T	0.23	0.020	3.20E-29	0.955	1
Omega-3 fatty acids	rs149212747	12	111836771	*ALDH2*	A/AC	0.16	0.0088	2.20E-73	0.754	1
Citrate	rs149212747	12	111836771	*ALDH2*	A/AC	-0.13	0.0088	1.10E-46	0.754	1

883

884 SE: Standard error. EAF: Effect allele frequency. EAF_BBJ: EAF in BBJ1-180K. EAF_EUR: EAF in
885 Europeans of 1KGP3.

886

887 **Table 2 28 novel associations identified by cross-ancestry meta-analysis**

Metabolite	rs ID	CHR	Position (GRCh37)	Locus name	Effect allele/ noneffect allele	BETA	SE	*P* value	EAF	EAF_BBJ	EAF_UKBEST
Tyrosine	rs958581	1	169128664	*NME7*	A/G	-0.010	0.0017	2.45E-09	0.587	0.363	0.600
Lactate	rs2199936	4	89045331	*PPM1K-DT*	A/G	-0.017	0.0027	1.95E-10	0.130	0.298	0.108
Valine	rs77065511	4	101093015	*LOC100507053*	A/C	-0.030	0.0049	1.82E-09	0.0743	0.192	0.0215
Cholesteryl esters in large VLDL	rs2173208	4	143315473	*FREM3*	T/G	-0.010	0.0017	8.32E-10	0.617	0.690	0.615
Cholesterol in large VLDL	rs2173208	4	143315473	*FREM3*	T/G	-0.010	0.0017	1.45E-09	0.617	0.690	0.615
Phospholipids in very large VLDL	rs6822835	4	143316581	*FREM3*	A/T	-0.010	0.0017	2.34E-09	0.641	0.696	0.639
Free cholesterol in very large VLDL	rs6822835	4	143316581	*FREM3*	A/T	-0.010	0.0017	1.70E-09	0.641	0.696	0.639
Cholesteryl esters in very large VLDL	rs6822835	4	143316581	*FREM3*	A/T	-0.010	0.0017	1.45E-09	0.641	0.696	0.639
Cholesterol in very large VLDL	rs6822835	4	143316581	*FREM3*	A/T	-0.010	0.0017	8.71E-10	0.641	0.696	0.639
Phospholipids to total lipids ratio in very small VLDL	rs10006795	4	143318673	*FREM3*	A/C	0.010	0.0017	2.45E-09	0.354	0.304	0.356
Free cholesterol to total lipids ratio in large LDL	rs10006795	4	143318673	*FREM3*	A/C	-0.010	0.0017	1.32E-09	0.354	0.303	0.356
Histidine	rs1500187	5	122824678	*HMGB3P17*	A/G	0.010	0.0017	1.45E-09	0.528	0.309	0.539
Total cholines	rs9399137	6	135419018	*HBS1L*	T/C	0.018	0.0029	6.46E-10	0.685	0.656	0.690
Valine	rs11971790	7	1055582	*C7orf50*	T/G	-0.012	0.0019	2.00E-09	0.496	0.460	0.499
Total lipids in very small VLDL	rs11763645	7	155025210	*SHH*	A/G	-0.017	0.0027	1.17E-09	0.887	0.641	0.914
Free cholesterol in medium LDL	rs117969049	9	112242562	*SVEP1*	T/C	-0.024	0.0039	2.14E-09	0.0638	0.209	0.0389
Cholesteryl esters to total lipids ratio in large LDL	rs117969049	9	112242562	*SVEP1*	T/C	-0.026	0.0041	2.57E-10	0.0660	0.209	0.0389
Citrate	rs12414179	10	111981935	*GPAM*	A/C	0.018	0.0029	7.94E-10	0.107	0.282	0.0766
Cholesteryl esters to total lipids ratio in IDL	rs1998848	14	21492229	*NDRG2*	A/G	-0.035	0.0058	1.55E-09	0.0504	0.159	0.0163
Total lipids in IDL	rs139262716	14	31725111	*NARS1P1*	A/G	-0.030	0.005	1.66E-09	0.0766	0.238	0.0192
Phospholipids in medium VLDL	rs7140599	14	39942467	*YTHDF2P1*	T/C	0.0099	0.0017	2.24E-09	0.428	0.620	0.421
Concentration of VLDL particles	rs7140599	14	39942467	*YTHDF2P1*	T/C	0.010	0.0016	1.00E-09	0.428	0.620	0.421
Free cholesterol to total lipids ratio in large LDL	rs80287865	15	70181451	*LINC00593*	T/G	0.047	0.0074	1.51E-10	0.0769	0.170	0.008
Cholesteryl esters to total lipids ratio in very small VLDL	rs80287865	15	70181451	*LINC00593*	T/G	0.048	0.0074	6.31E-11	0.0769	0.169	0.00756
Triglycerides to total lipids ratio in IDL	rs80287865	15	70181451	*LINC00593*	T/G	-0.045	0.0074	1.45E-09	0.0794	0.170	0.00756
Cholesterol to total lipids ratio in IDL	rs80287865	15	70181451	*LINC00593*	T/G	0.046	0.0075	5.13E-10	0.0798	0.170	0.00756

Free cholesterol to total lipids ratio in medium LDL	rs80287865	15	70181451	*LINC00593*	T/G	0.047	0.0074	3.24E-10	0.0772	0.170	0.00756
Cholesterol to total lipids ratio in very small VLDL	rs80287865	15	70181451	*LINC00593*	T/G	0.047	0.0074	3.47E-10	0.0772	0.169	0.00756

SE: Standard error. EAF: Effect allele frequency in cross-ancestry meta-analysis. EAF_BBJ: EAF in BBJ1-180K. EAF_UKB: EAF in UKB.