nature portfolio

https://doi.org/10.1038/s41586-025-10054-6

Supplementary information

A cross-population compendium of gene–environment interactions

In the format provided by the authors and unedited

Supplementary information

A Cross-population Compendium of Gene–Environment Interactions

Namba S *et al.*

Corresponding to Shinichi Namba (snamba@m.u-tokyo.ac.jp)

and Yukinori Okada (yuki-okada@m.u-tokyo.ac.jp)

Table of contents

Supplementary Notes

Supplementary Note 1. The authors and their affiliations of the BioBank Japan Project

Koichi Matsuda[30,32], Yuji Yamanashi[34], Yoichi Furukawa[35], Takayuki Morisaki[29,30], Yukinori Okada[1–3,13,33], Yoshinori Murakami[36], Kaori Muto[37], Akiko Nagai[30], Yusuke Nakamura[38], Wataru Obara[39], Ken Yamaji[40], Kazuhisa Takahashi[41], Satoshi Asai[42,43], Yasuo Takahashi[43], Shinichi Higashiue[17], Shuzo Kobayashi[17], Hiroki Yamaguchi[18], Yasunobu Nagata[18], Satoshi Wakita[18], Yasushi Okazaki[19], Naoyuki Matsumoto[19], Chikako Nito[44], Yu-ki Iwasaki[45], Shigeo Murayama[46], Kozo Yoshimori[47], Yoshio Miki[48], Daisuke Obata[49], Masahiko Higashiyama[50], Kenta Motomura[20], Hidenobu Koga[20], and Yukihiro Koretsune[51]

[1]Department of Genome Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

[2]Department of Statistical Genetics, Graduate School of Medicine, The University of Osaka, Suita, Japan.

[2]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

[13]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), The University of Osaka, Suita, Japan.

[17]Tokushukai Group, Tokyo, Japan.

[18]Departmentof Hematology, Nippon Medical School, Tokyo, Japan.

[19]Diagnostics and Therapeutics of Intractable Diseases, Intractable Disease Research Center, Graduate School of Medicine, Juntendo University, Tokyo, Japan.

[20]Iizuka Hospital, Fukuoka, Japan.

[29]Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[30]Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.
[32]Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
[33]Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), The University of Osaka, Osaka, Japan.
[34]Division of Genetics, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
[35]Division of Clinical Genome Research, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
[36]Department of Cancer Biology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
[37]Department of Public Policy, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
[38]The Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
[39]Department of Urology, Iwate Medical University, Iwate, Japan.
[40]Department of Internal Medicine and Rheumatology, Juntendo University Graduate School of Medicine, Tokyo, Japan.
[41]Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan.
[42]Division of Pharmacology, Department of Biomedical Science, Nihon University School of Medicine, Tokyo, Japan.
[43]Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University. School of Medicine, Tokyo, Japan.
[44]Laboratory for Clinical Research, Collaborative Research Center, Nippon Medical School, Tokyo, Japan.

[45]Department of Cardiovascular Medicine, Nippon Medical School, Tokyo, Japan.

[46]Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan.

[47]Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan.

[48]The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan.

[49]Center for Clinical Research and Advanced Medicine, Shiga University of Medical Science, Shiga, Japan.

[50]Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan.

[51]National Hospital Organization Osaka National Hospital, Osaka, Japan.

Supplementary Note 2. Curation of previously reported G×E interactions from the GWAS Catalog

The GWAS catalog[27] includes a substantial number of gene–environment (G×E) studies. Although it may not encompass all such studies, it serves as a valuable resource for examining previously reported G×E interactions. To assess the overlap between the G×E interactions identified in our manuscript and those previously reported, we curated and compiled a list of G×E interactions from the GWAS catalog.

The GWAS catalog provides a specific flag “GXE” to denote G×E studies, which were marked for 268 studies (795 study–trait pairs with $P < 5.0\times10^{-8}$) as of July 8, 2024. However, a closer inspection revealed that not all flagged entries represented true G×E studies; some were gene–gene (G×G) interaction studies, while others lacked the “GXE” flag despite focusing on G×E interactions.

To define a reliable set of previously reported G×E interactions, we filtered the studies reporting genome-wide significant variants ($P < 5.0\times10^{-8}$), using partial matches in trait names followed by manual review of all retained studies, according to the criteria below:

1. Putative G×E interactions:
    (i) Included the term “interaction” or “Interaction”
    (ii) Did not match the regular expression "adjusted for .+ interaction", which meant the phenotype of standard GWAS was adjusted for some interaction term
    (iii) Did not contain the phrases indicating G×G interactions or joint tests of G×E interactions and a marginal effect: "SNP x SNP interaction", "4df test", "2df test", "(2df)", "APOE e4 interaction", "x APOL1 genotype interaction", "GSK3-beta interaction protein levels", "SNP x mitochondrial", "joint analysis main effects and ", "(main effect)", "Stromal interaction molecule 1 level", "(3df)"
2. Putative joint tests of G×E interactions and a marginal effect:
    (i) Included the term “interaction” or “Interaction”

(ii) Did not match the regular expression "adjusted for [^¥(]+ interaction"

(iii) Contained at least one of the following phrases: "4df test", "2df test", "(2df)", "(3df)", "joint analysis main effects and", "joint analysis for main effect and"

(iv) Did not include "rs671", which indicated G×G interactions

Using these criteria, we initially identified 103 studies (178 study–trait pairs) as putative G×E interaction studies, and 32 studies (89 study–trait pairs) as putative joint test studies. We then manually reviewed each study, and the following studies were excluded:

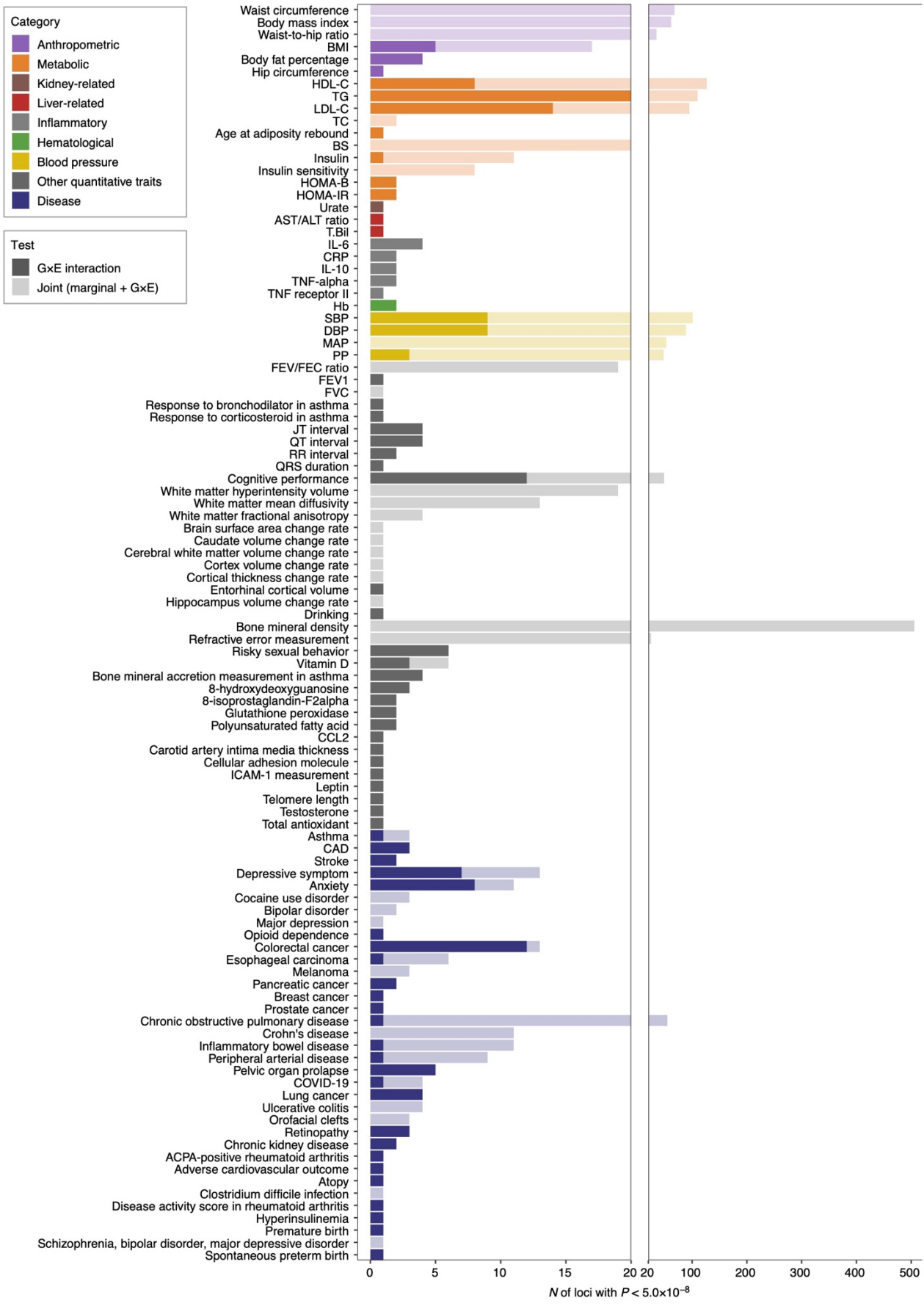
(i) Studies that did not evaluate multiplicative G×E interactions

(ii) Studies suspected of inflated results, specifically, those reporting genome-wide significant G×E loci from cohorts with only several hundred participants, or studies with no more than 2,000–3,000 participants reporting over 10 genome-wide significant G×E loci, >80% of which included only one significant variant per locus.

We also manually curated phenotype names from the GWAS catalog, as the phenotype names for G×E interactions were not standardized, often including phenotype and environment names and degrees of freedom for statistical tests.

After manual review, some entries initially categorized as G×E interaction tests were reclassified as joint tests. Ultimately, we identified 76 studies reporting G×E interactions for 203 loci (217 trait–locus pairs) and 46 studies reporting joint tests for 998 loci (1571 trait–locus pairs) **(Supplementary Figure 1)**. We found that pleiotropic G×E effects across multiple traits were rarely reported in this curated list, possibly due to the lack of systematic evaluation across phenotypes and environments.

In summary, we systematically curated and compiled G×E interactions from the GWAS catalog, providing a rigorously reviewed resource to support future validation and discovery efforts in G×E research. The full curated list is available as **Supplementary Table 7**.


Category
Anthropometric
Metabolic
Kidney-related
Liver-related
Inflammatory
Hematological
Blood pressure
Other quantitative traits
Disease
Test
GxE interaction
Joint (marginal + GxE)
Waist circumference
Body mass index
Waist-to-hip ratio
BMI
Body fat percentage
Hip circumference
HDL-C
TG
LDL-C
TC
Age at adiposity rebound
BS
Insulin
Insulin sensitivity
HOMA-B
HOMA-IR
Urate
AST/ALT ratio
T.Bil
IL-6
CRP
IL-10
TNF-alpha
TNF receptor II
Hb
SBP
DBP
MAP
PP
FEV/FEC ratio
FEV1
FVC
Response to bronchodilator in asthma
Response to corticosteroid in asthma
JT interval
QT interval
RR interval
QRS duration
Cognitive performance
White matter hyperintensity volume
White matter mean diffusivity
White matter fractional anisotropy
Brain surface area change rate
Caudate volume change rate
Cerebral white matter volume change rate
Cortex volume change rate
Cortical thickness change rate
Entorhinal cortical volume
Hippocampus volume change rate
Drinking
Bone mineral density
Refractive error measurement
Risky sexual behavior
Vitamin D
Bone mineral accretion measurement in asthma
8-hydroxydeoxyguanosine
8-isoprostaglandin-F2alpha
Glutathione peroxidase
Polyunsaturated fatty acid
CCL2
Carotid artery intima media thickness
Cellular adhesion molecule
ICAM-1 measurement
Leptin
Telomere length
Testosterone
Total antioxidant
Asthma
CAD
Stroke
Depressive symptom
Anxiety
Cocaine use disorder
Bipolar disorder
Major depression
Opioid dependence
Colorectal cancer
Esophageal carcinoma
Melanoma
Pancreatic cancer
Breast cancer
Prostate cancer
Chronic obstructive pulmonary disease
Crohn's disease
Inflammatory bowel disease
Peripheral arterial disease
Pelvic organ prolapse
COVID-19
Lung cancer
Ulcerative colitis
Orofacial clefts
Retinopathy
Chronic kidney disease
ACPA-positive rheumatoid arthritis
Adverse cardiovascular outcome
Atopy
Clostridium difficile infection
Disease activity score in rheumatoid arthritis
Hyperinsulinemia
Premature birth
Schizophrenia, bipolar disorder, major depressive disorder
Spontaneous preterm birth
0
5
10
15
20
20
100
200
300
400
500
N of loci with P < 5.0×10−8

170

Supplementary Figure 1. G×E interactions curated from the GWAS catalog.

Curated G×E interactions and joint tests of G×E interactions and a marginal effect with *P*-values less than 5.0×10^{-8}.

Supplementary Note 3. Overlaps of G×E interactions with recent UKB studies

We examined the overlap of the G×E interactions detected in this manuscript with two recent studies using UKB: Bernabeu E *et al.*[28] and Westerman KE *et al*[13]. The former focused on G×Sex interactions, with a different statistical model from our study. The latter investigated G×E interactions for cardiometabolic serum biomarkers after filtering the variants to the variance quantitative trait loci (vQTL). We note that these two studies reported per-environment *P*-values, whereas we defined significance based on locus-level, environment-aggregated *P*-values, which were inherently more stringent at the same threshold. In addition, Westerman *et al*. applied a more lenient significance threshold ($P = 2.35 \times 10^{-7}$) compared to ours, as they adopted a two-step approach using the vQTL filtering.

We regarded the G×E loci as overlapping when the lead variants of the previous studies were located within 500 kb of the genome-wide significant variants detected in this manuscript. To account for differences in the trait definition, we considered trait–locus pairs as overlapping if the traits belonged to the same category (e.g., uric acid [UA] and gout). Moreover, overlaps were counted regardless of differences in the environmental factors tested.

For the G×E interactions we detected in UKB1, we found 4 overlapping loci (6 trait–locus pairs) for Bernabeu E *et al.* and 12 overlapping loci (14 trait–locus pairs) for Westerman KE *et al*. (**Supplementary Table 8**). For those detected in BBJ1, five loci (five trait–locus pairs) were overlapped with Bernabeu E *et al.*, and one locus (one trait–locus pair) was overlapped with Westerman KE *et al*. These loci included the *PITX2* locus for arrhythmia, the *CYP19A1* locus for height, and the *SLC22A11* locus for UA, all of which were overlapped with Bernabeu E *et al.* These loci would be additional candidates of cross-population shared G×E interactions, although further validation is warranted as they were identified by the different statistical protocols.

Since Bernabeu E *et al.* focused exclusively on G×Sex interactions, it is notable that

most of the overlapping signals in our analyses were primarily driven by G×Sex interactions (7/7 loci; 9/11 trait–locus pairs). On the contrary, only a subset of loci where G×Sex was identified as the top contributor in our study overlapped with Bernabeu E *et al.* (5 out of 37 loci for UKB1 and 4 out of 11 loci for BBJ1). This observation suggests that our inclusion of a broader set of environmental factors improved the power to detect G×E interactions.

Supplementary Note 4. Comparable replication rates of G×E interactions compared to those for GWAS

To empirically compare the replication rates of G×E interactions to those of GWAS, we performed GWAS for the traits with genome-wide significant G×E interactions in UKB1 and BBJ1. We detected 5,477 and 1,802 trait–locus pairs with genome-wide significant marginal effects in UKB1 and BBJ1, respectively (7,279 in total). In UKB2, 10.1% (552 / 5,477) were significantly replicated at the Bonferroni-corrected threshold of 6.9×10^{-6} (= 0.05 / 7,279), and 34.0% (613 / 1,802) were significantly replicated in BBJ2, reflecting its larger sample size than UKB2. The replication rates of G×E interactions were comparable in UKB1 and higher in BBJ2: 9.4% (6 / 64) in UKB2 and 52.8% (19 / 36) in BBJ2 at the Bonferroni-corrected threshold of 5.0×10^{-4} (= 0.05 / 100).

In the above analysis, the Bonferroni-corrected thresholds were different between G×E and GWAS. Therefore, we next filtered the GWAS trait–locus pairs to match the Bonferroni-corrected thresholds. Specifically, for each G×E locus, we selected the marginal effect locus for the same trait whose *P*-value was closest to that of the corresponding G×E interaction. Using this matched set, the replication rates of GWAS were 14.1% (9 / 64) in UKB2 and 66.7% (24 / 36) in BBJ2, only slightly higher than those of G×E interactions.

In summary, we observed that the replication rates of G×E interactions were largely comparable to those of GWAS. Considering that differences in environmental distributions across cohorts can reduce the replicability of G×E interactions, and that past G×E studies have suffered from low replication rates[15,16], these findings suggest the high reliability of the G×E interactions identified in this manuscript.

Supplementary Note 5. Meta-analyses of G×E interactions across BBJ and UKB

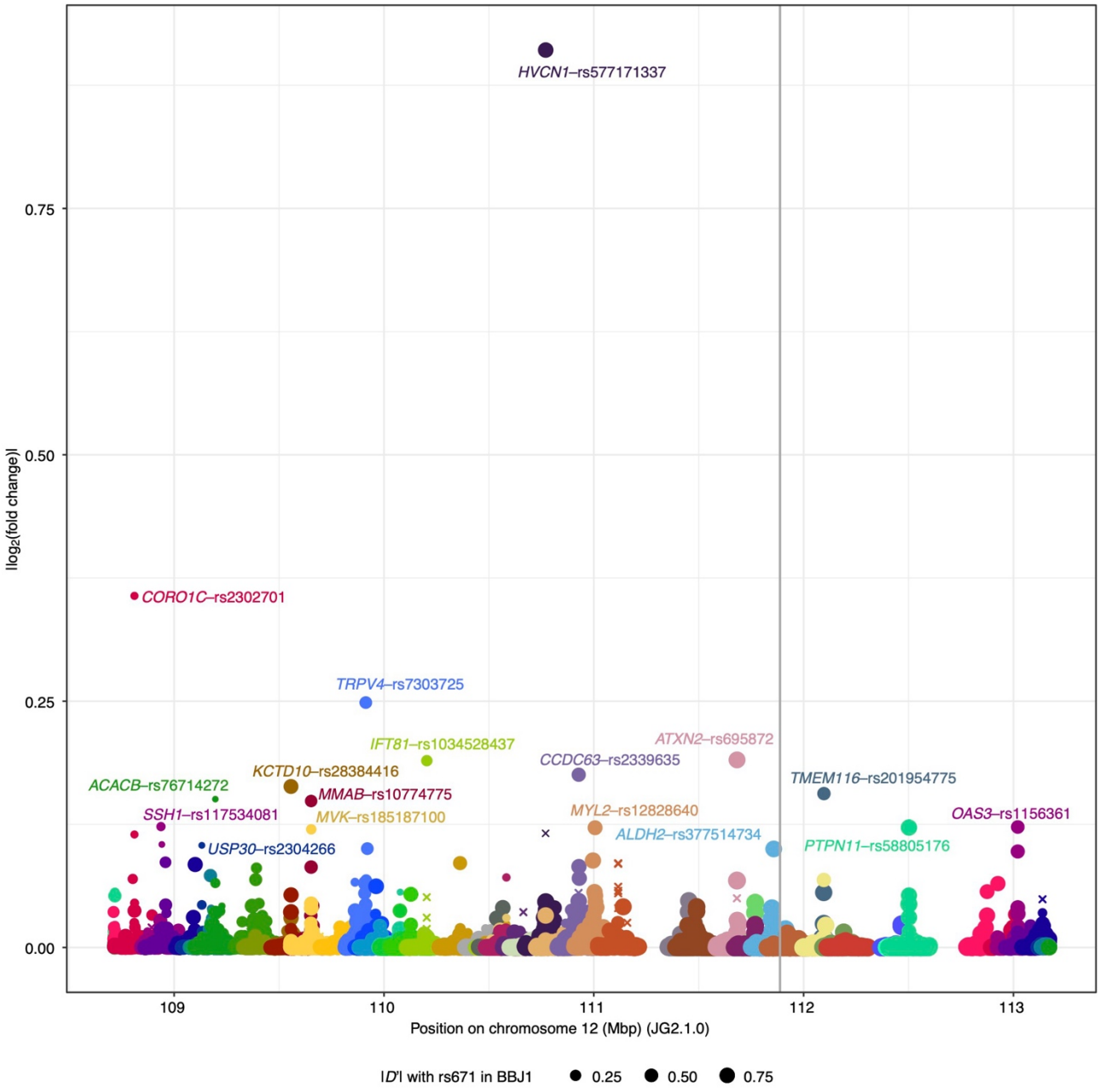
To expand the catalog of G×E interactions using larger sample sizes, we selected the variants with suggestive significance ($P_{G\times E} < 1.0\times10^{-7}$) in one of the discovery cohorts and performed meta-analyses across BBJ and UKB. Inverse-variance weighted fixed-effect meta-analyses in individual populations detected four additional G×E interactions ($P_{G\times E} < 5.0\times10^{-8}$, **Supplementary Table 11**), three of which were for disease statuses. These included a G×E interaction for T2D status in the *ALDH3B2* locus ($P_{G\times E} = 4.8\times10^{-8}$ in the UKB1+2 meta-analysis, driven by G×Ever-smoking, G×Age, and G×Ever-drinking). As the G×E interaction for T2D status was also detected in the same alcohol metabolism pathway (the *ALDH2* locus in BBJ1), these interactions might indicate the cross-population consistency of G×E interaction at the pathway level. A cross-population meta-analysis added one more G×E interaction for aspartate aminotransferase (AST) in the *HK1* locus ($P_{G\times E} = 3.5\times10^{-8}$, driven by G×Sex), which still passed the genome-wide significant threshold when using a random-effect meta-analysis ($P_{G\times E} = 3.5\times10^{-8}$).

Supplementary Note 6. Multiple environments contributed to the G×E interaction in the *ALDH2* locus in BBJ

In BBJ1, at most seven out of the nine tested environments contributed to the G×E interactions in this locus (mean 3.38 [range 2–7]) (**Supplementary Table 6**). This number was relatively larger than that for the other loci in BBJ1 (mean 2.00 [range 1–4]) and UKB1 (mean 2.17 [range 1–5]). Particularly, sex and current smoking were detected in many traits (16/21 and 13/21 traits, respectively). Out of the 21 traits that showed genome-wide significant G×E interactions, 13 and 8 traits still showed $P_{G\times E}$ less than 5.0×10^{-8} after conditioning on G×E interactions for the ever-drinking status and the amount of drink per week, respectively (**Supplementary Table 13**). We note that we defined drink per week as the amount of alcohol consumption [g/week] converted by log(1+x). These results indicated that the environments other than drinking status contributed to this locus.

Supplementary Note 7. Deep-learning based *in silico* prioritization of regulatory variants in the *ALDH2* locus

As the never-drinker-specific effects of the *ALDH2* locus on the hematopoietic traits implied the existence of causal variants other than rs671, we investigated the variants with potential regulatory effects on gene expression in this locus. Considering that the causal variants and genes were unknown, we employed a deep-learning model[73] to predict the regulatory effects for all testable gene–variant pairs found in the Japanese population in this locus comprehensively (**Supplementary Methods**). Out of 29,356 testable gene–variant pairs, which covered 62 genes from *ISCU* to *RITA1* and 10,783 variants, relatively strong regulatory effects in CD34+ hematopoietic stem cells were predicted for 21 pairs (absolute $\log_2$ fold change > 0.1, **Supplementary Figure 2**). Seven pairs (*HVCN1*-rs577171337, *ATXN2*-rs695872, *CCDC63*-rs2339635, *KCTD10*-rs28384416, *MMAB*-rs10774775, *PTPN11*-rs58805176, and *MYL2*-rs12828640) were further associated with hematopoietic traits (red blood cells, Hb, hematocrit, or white blood cells) in the never-drinkers in BBJ1 ($P < 0.05$) and were in the LD with rs671 ($|D'| > 0.1$ in BBJ1). These pairs could be the candidates of causal genes and variants, which warrant further research. To facilitate future research, we provided the list of potential regulatory effects with absolute $\log_2$ fold change larger than 0.05 in **Supplementary Table 15**.



277

278 **Supplementary Figure 2. Prioritization of variants with potential regulatory effects in**
279 **the *ALDH2* locus.**

280 Predicted regulatory effects on gene expression in CD34+ hematopoietic stem cells for
281 29,356 gene–variant pairs in the *ALDH2* locus. The *x*-axis represents the genomic positions
282 of variants in the JG2.1.0 Japanese genome assembly, and the vertical gray line shows the
283 genomic position of rs671. Colors represent genes, and the point size represents LD (|*D'*|)
284 with rs671 in BBJ1. Variants not included in the BBJ1 imputed genetic data are shown with

crosses. Variant identifiers are shown if the predicted absolute log2 fold change was the largest for individual genes and larger than 0.1.

Supplementary Note 8. Potential causal directions between environments and clinical traits

The discovery of the reverse causal relationship of the G×Natto interaction at the *PITX2* locus prompted us to a broader investigation into the causal directions of other G×E interactions identified in this manuscript. Specifically, we examined whether environmental factors influenced clinical traits by leveraging longitudinal data from multiple cohorts.

For the biomarkers, we leveraged the most recent repeat measurement data and restricted the data to those measured at least six months after the baseline survey (**Supplementary Table 16**). In BBJ1, the median time interval from the baseline to the phenotype assessment was 4.0 years (range, 0.5–9.5; N_{max} = 105,339). Among 31 trait–locus pairs, 15 were significantly verified using the repeated measurement data ($P_{G\times E} < 0.05/100 = 5.0\times10^{-4}$; **Supplementary Table 17**). In J-MICC/HERPACC, the second measurement data were available (N_{max} = 37,398), which were assessed five years after the baseline assessment. In this cohort, 11 out of 27 testable G×E interactions were significantly verified. Notably, the highly pleiotropic G×E interactions at the *ALDH2* locus were consistently verified in both BBJ1 and J-MICC/HERPACC.

For the European population, we primarily analyzed All of Us rather than UKB1, as described below. In UKB1, the second to fourth follow-up visits were available (median time interval = 5.0 years [range, 2.0–17.0]; N_{max} = 50,958). Nevertheless, the G×E interactions were not significantly verified. This result is likely attributable to the biased characteristics reported for the participants of the repeat visits in UKB, who tended to be older, less socio-economically depleted, and never smokers, with higher education and excellent overall health rating (https://biobank.ndph.ox.ac.uk/~bbdatan/repeat_assessment_characteristics.pdf). We repeated the analysis using baseline measurements data from the same subset of participants. The G×E interactions were again not significantly verified, suggesting that the

selection bias, rather than measurement timing, was a key limiting factor of longitudinal validation in UKB1. Therefore, we performed the assessment using the European population in All of Us (median time interval = 2.2 years [range, 0.5–6.3]; N_{max} = 47,292). In this cohort, 4 out of 62 testable G×E interactions were significantly verified (**Supplementary Table 17**). These results support the causal direction from environments to clinical traits for the verified G×E interactions.

As a limitation of this approach, we note that the measurement values and environments can be temporally stable and may change only gradually over time, which complicates the strict inference of causal direction based solely on measurement timing. Moreover, sex and age were inherently unaffected by clinical traits, and therefore, the repeat assessment analysis may not serve as additional evidence for the causal direction of these factors. In some cases, clinical traits and environmental exposures may arise simultaneously, or G×E signals may emerge statistically without a direct biological mechanism[14].

For the disease traits, we focused on incidental cases, whose disease onsets were at least six months after the baseline assessment, and performed Cox regression with G×E interaction terms. In BBJ1, the G×E interaction for T2D at the *ALDH2* locus, primarily driven by G×Ever-drinking, was significantly verified ($P_{G\times E} = 5.3\times10^{-5}$). On the other hand, the G×E interaction for arrhythmia at the *PITX2* locus was not significant ($P_{G\times E} = 0.077$). In the section "A reverse-causal G×E interaction," we discussed that this locus was driven by a reverse causal relationship from a disease to an environment (specifically, from warfarin medication for atrial fibrillation/flutter to the avoidance of natto consumption). Considering that the Cox regression model tested the temporal direction from environments to diseases, the failure of its verification in the Cox regression was expected. In UKB1, no G×E interaction was significantly verified, while the G×Ever-drinking interaction for arrhythmia at the *STARD3NL* locus was nominally verified ($P_{G\times E} = 0.026$).

To assess the causal direction more confidently, we performed Cox regression for overall survival, where the temporal order from environmental factors to the endpoint of death was apparent. Testing 54 loci in their discovered cohorts, we observed a significant G×E interaction for the *ALDH2* locus in BBJ1 ($P_{G\times E} = 1.7\times10^{-11}$; driven by G×Sex, G×Ever-drinking, and G×Age), indicating that the G×E interaction may affect human lifespan. In UKB1, we observed a significant G×Age interaction at the *APOE* locus ($P_{G\times E} = 5.5\times10^{-8}$), consistent with its G×Age interactions for total cholesterol (TC) and triglycerides (TG). As its effect size was in the same direction as the additive genetic effect (0.098 [95% confidence interval, 0.065–0.130] and 0.054 [0.020–0.087], respectively), this interaction suggested a larger mortality risk for carriers of the risk variant than non-carriers at older ages.

Collectively, we systematically investigated the causal directions between environments and phenotypes, which may help future studies to elucidate the mechanisms underlying these G×E interactions.

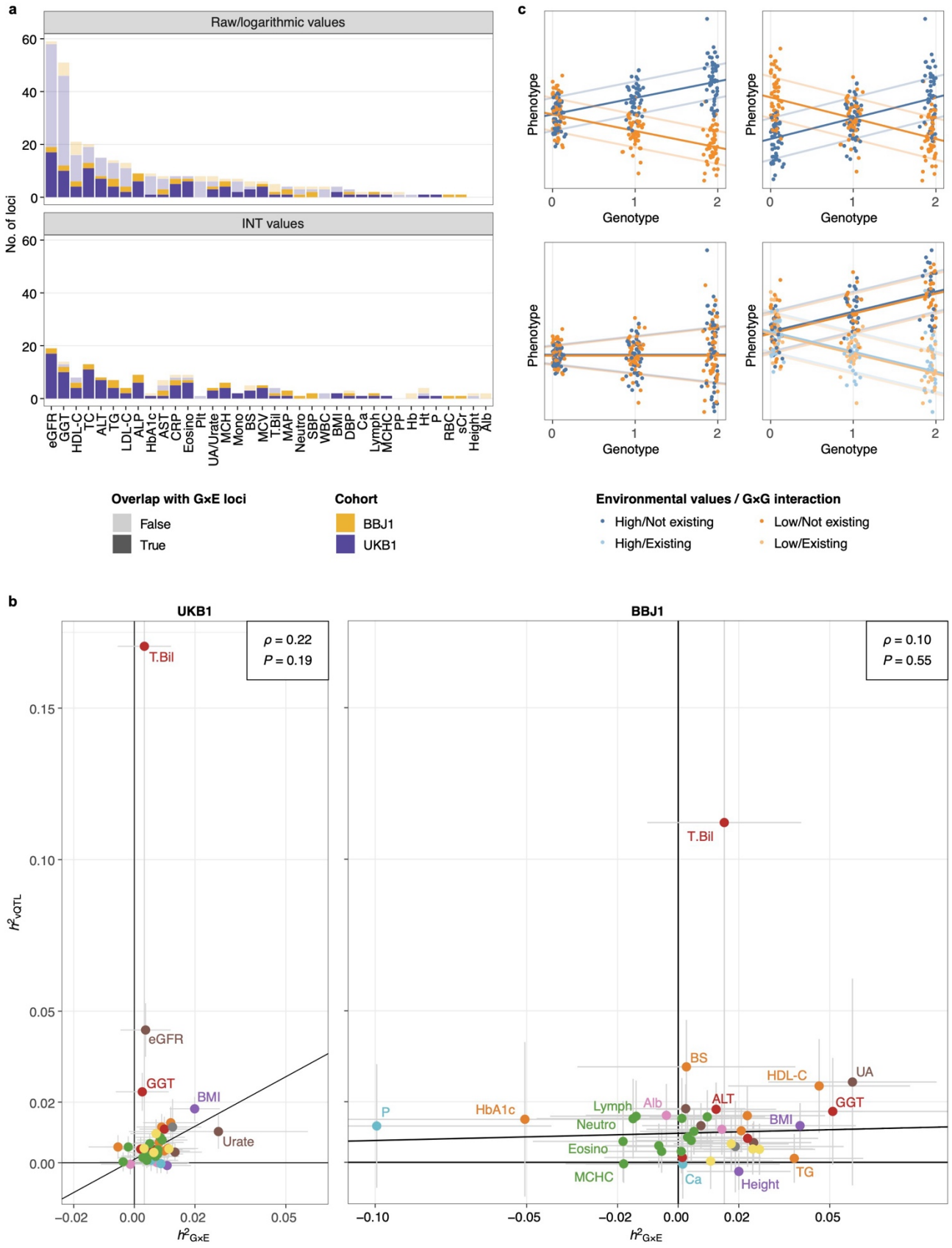
Supplementary Note 9. Variance quantitative trait loci failed to capture the entire G×E interactions

Variance quantitative trait loci (vQTL) analyses test the associations between genotypes and phenotypic variance. They have been used to prioritize or prefilter candidate G×E interaction loci, as measuring environmental factors is not necessary, and the tests are free from the heavy multiple testing burden due to the number of environments[13,74]. We detected 246 and 71 significant vQTL trait-locus pairs in UKB1 and BBJ1, respectively (**Supplementary Table 22**), using the state-of-the-art Brown–Forsythe test (**Supplementary Methods**). The number of vQTL trait-locus pairs was much larger than that of G×E interactions, but many of them were detected for a few traits in UKB1, particularly for eGFR and GGT (100 trait-locus pairs [40.7%]) (**Supplementary Figure 3a**). Most of these vQTL (73/100 pairs) were not significant when applying the rank-based inverse normal transformation to the phenotypic values, suggesting that these vQTL might be false positives due to the skew of the phenotypic distributions. The detected vQTL included 26/169 trait–locus pairs (15.4%) for which G×E interactions were also significant, after excluding eGFR and GGT. This was 14.6 times enrichment compared to GWAS loci (66 out 6,251 trait–locus pairs [1.1%]), indicating that the genome-wide scanning of vQTL was effective as a prefiltering step, as previously reported[13]. However, they amounted to only 32.5% (26/80) of the G×E interactions, indicating that vQTL could not capture the majority of G×E interactions. We observed closely similar results when using another testing method, the Deviation Regression Model (DRM)[75] (**Supplementary Table 22**).

Next, we estimated the heritability of vQTL by using LDSC. We could use LDSC without explicitly extending this method to vQTL, as the DRM test was a form of linear regression. Notably, the heritability of vQTL and G×E interactions were not correlated across traits even after removing the outlier traits (i.e., T.Bil, eGFR, and GGT) ($P > 0.05$ for Spearman's correlations in both UKB1 and BBJ1, **Supplementary Figure 3b**). These discordance

between vQTL and G×E interactions might be because G×E interactions could arise independently from phenotypic variation[14]. In addition, vQTL could arise from any non-additive genetic effects other than G×E interactions, such as heteroscedasticity and gene–gene (G×G) interactions. Furthermore, vQTL might also depend on the “baseline” phenotypic value from which the variance was calculated (**Supplementary Figure 3c**). On the other hand, it is also possible that vQTL might be related to G×E interactions from environments not evaluated in this study.

Collectively, we observed the discordance between vQTL and G×E interactions, indicating the necessity of direct evaluation of environmental values to detect G×E interactions.



Supplementary Figure 3. Variance quantitative trait loci failed to capture the entire G×E

interactions.

a, Numbers of genome-wide significant vQTL. **b**, Correlation of heritability (h^2) between G×E interactions and vQTL in UKB1 and BBJ1. Data are presented as estimated values with 95% CI. The regression line was estimated with the Deming regression to account for the estimation errors in both *x*- and *y*-axes, after excluding GGT, eGFR, and T.Bil. *P*-values were estimated with a two-sided Spearman's rank correlation test. **c**, Schematic illustrations of the relationship between G×E/G×G interactions and vQTL. Some G×E interactions can be detected as vQTL (upper left), but other might not be due to the baseline shift (upper right). In addition, heteroscedasticity and G×G interactions can cause vQTL without G×E interactions (lower panels). For **a and c**, see **Supplementary Table 1** for the sample sizes.

Supplementary Note 10. Fine-mapping and colocalization analyses of the *TNFAIP8* locus

To pinpoint causal variants, we tried fine-mapping the *TNFAIP8* locus (±500 kbps from the lead variant) using two popular methods, SuSiE[76] and FINEMAP[77]. We used default settings for all methods. First, we applied SuSiE to marginal associations and G×E interactions for clinical HDL-C using the in-sample LD matrix. SuSiE estimated that the number of causal variants was one for both marginal and G×E effects, but could not pinpoint the causal variants as the credible sets included 9 and 14 variants, respectively. Next, we applied FINEMAP and found that the inferred number of causal variants was five for both effects, the maximum possible number for FINEMAP by default. These inconsistent results between methods suggested that statistical fine-mapping was difficult for the *TNFAIP8* locus.

We then applied coloc[78] to examine whether the causal variant was identical between marginal and G×E effects. The posterior probability of H3 (the causal variants were different) and H4 (the causal variants were identical) was 33.0% and 67.0%, respectively. As these values were not high enough to determine the colocalization, we concluded that this locus had a complex LD pattern, which hampered statistical methods to infer causal variants.

Supplementary Note 11. Implications of trans-resolution G×E interaction analyses

At the locus level, the warfarin-causing G×E interaction for arrhythmia suggested that careful evaluation by specialists was necessary to interpret G×E interactions. In this regard, prospective follow-up of biobank participants would be important for enabling longitudinal studies to resolve causal directions. We also note that environment-specific bias, such as survival bias for G×Age interactions, requires careful consideration.

At the genome-wide level, we systematically demonstrated that G×E interactions significantly contribute to heritability, and affect PGS prediction accuracy and cross-population portability. Incorporating environmental context into PGS may enhance personalization and promote more equitable clinical application of genetic risk prediction across populations.

Metabolome G×E QTL analyses identified multiple sex-dimorphic and environment-specific genetic effects on the metabolome, suggesting that omics data may reveal G×E interactions with large effect sizes. The sex-discordant effects in lipid metabolites provided implications for drug development, underscoring the importance of studying G×E interactions at the molecular level.

GWAS has entered the era of global biobank collaboration and has identified thousands of loci by boosting detection power and including underrepresented populations[21]. Our locus- and genome-wide analyses consistently indicated that G×E interactions were moderately shared across populations, encouraging future international collaborative efforts. At the same time, population-specific loci emphasized the importance of studying diverse populations. In addition to differences in allele frequencies, environmental factors and their distributions could differ across populations or their correlated factors (e.g., regions), influencing the detection power of G×E interactions. In this regard, our study has limitations in discovering G×E interactions in underrepresented populations other than European and East Asian populations, and for geo-related exposures like pollution and ultraviolet radiation exposure[19].

Establishing large, diverse cohorts with detailed environmental data would be essential for further detecting unobserved G×E interactions. In addition, omics-level G×E QTL analyses have rarely been conducted due to the paucity of cohorts collecting both omics and environmental data[19]. As demonstrated using UKB and BBJ, multi-omics integration in data-rich biobanks is key to understanding molecular and clinical G×E interactions.

Supplementary Methods

Japan Multi-Institutional Collaborative Cohort (J-MICC) Study and Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC)

J-MICC is a multi-institutional collaboration of prospective and population-based cohorts comprising approximately 100,000 participants in Japan, launched in 2005 with the primary purpose of detecting G×E interactions of lifestyle-related diseases[60]. Individuals aged 35 to 69 years were enrolled from 2005 to 2010 from respondents to study announcements in specified areas, inhabitants attending health checkup examinations by local governments, visitors at health checkup centers, and patients at a cancer hospital. The second assessment was conducted five years after their enrollment. HERPACC is a hospital-based cohort including participants aged 18 or older recruited from 2001 to 2013 from first-visit outpatients at Aichi Cancer Center Hospital[61,62]. Aichi Cancer Center Hospital is one of the constitutive cohorts of J-MICC, and the same questionnaire was used for both J-MICC and HERPACC. After removing duplicated participants between J-MICC and HERPACC, we analyzed the participants of J-MICC and HERPACC together, which we refer to as J-MICC/HERPACC. For HERPACC, we analyzed the participants with no cancer or history of neoplasm to match the cohort to the population-based characteristics of J-MICC.

Quality control of genotype data

The participants were genotyped using one of the following arrays: Illumina HumanCoreExome-12, HumanCoreExome-24, Asian Screening Array, OmniExpressExome-8, and Affymetrix Axiom Japonica NEO. We genotyped Illumina arrays using GenomeStudio and the Japonica NEO array using the Applied Biosystems Analysis Power Tools and the SNPolisher R package. For the latter, we followed the best practices described in the Axiom Genotyping Solution Data Analysis User Guide, and we set the sample call rate threshold to 95%.

For quality control, we merged the genotype data from HumanCoreExome-12 and HumanCoreExome-24, considering the high overlaps of the genotyped variants between these arrays. We did not merge other genotype data, and note that we conducted statistical tests in individual datasets and meta-analyzed the results using the fixed-effect inverse-variance weighted method. For the HumanCoreExome-12/24 merged data, we excluded participants with low call rates (<0.95), high heterozygosity (>6 standard deviations), or genetic sex inconsistent with clinical sex, and outliers from the Japanese cluster estimated based on PCA. We excluded the variants meeting the following criteria: (i) with low call rates (<0.95 in the merged data or <0.90 in any individual arrays); (ii) with low minor allele frequencies (<0.01 in the merged data); (iii) array-specific variants with minor allele counts of 0 for any array; (iv) with Hardy–Weinberg equilibrium test *P*-values < 1.0×10^{-7} (in the merged data or any individual arrays); and (v) with difference in the allele frequencies between arrays (chi-squared test *P*-values < 1.0×10^{-25}). For the Asian Screening Array, we applied the same quality control as used in BBJ2. For the other array, we excluded participants with low call rates (<0.96 for OmniExpressExome-8 and <0.95 for Japonica NEO), high heterozygosity (>6 standard deviations for OmniExpressExome-8 and >10 for Japonica NEO), or genetic sex inconsistent with clinical sex, and outliers from the Japanese cluster estimated based on PCA. We excluded the variants meeting the following criteria: (i) with low call rates (<0.95); (ii) with low minor allele frequencies (<0.005); and (iii) with Hardy–Weinberg equilibrium test *P*-values < 1.0×10^{-30}.

For individual datasets, we performed statistical phasing of the genotype data using Shapeit4 and imputation using minimac4 with the in-house reference panel constructed from 11,942 Japanese whole genome sequences. After imputation, we excluded variants with an imputation quality less than 0.7 or a MAF less than 0.01. We used King to exclude relatives within second degrees in individual datasets and duplicates across datasets.

Japan Public Health Center-based Prospective (JPHC) Study

The JPHC study is a population-based cohort comprising participants enrolled from residents aged 40-69 years within 11 public health center (PHC) areas nationwide between 1990 and 1994[63]. Before conducting genetic research in the JPHC study, we obtained approval from the institutional review board of the National Cancer Centre (Approval No.: 2011-044), Tokyo, Japan, and provided eligible participants with the option to decline participation. For the present study, 13,024 participants were randomly selected as a subcohort from the JPHC study participants in 9 PHC areas, who had received a health check-up and answered a questionnaire during the baseline survey.

Quality control of genotype data

The subcohort participants were genotyped using GenomeStudio with one of the Illumina OmniExpressExome-8, HumanOmni2.5, and HumanOmniExpress-12 arrays. We merged the genotype data of individual arrays, retaining only the common variants across all arrays. We excluded participants with low call rates (<0.96), high heterozygosity (>6 standard deviations), or genetic sex inconsistent with clinical sex, and outliers from the Japanese cluster estimated based on PCA. We excluded the variants meeting the following criteria: (i) with low call rates (<0.95 in the merged data or <0.90 in any individual arrays); (ii) with low minor allele frequencies (<0.005 in the merged data); (iii) with Hardy–Weinberg equilibrium test *P*-values < 1.0×10^{-30} (in the merged data or any individual arrays); and (iv) with difference in the allele frequencies between any combinations of two arrays (chi-squared test *P*-values < 1.0×10^{-7}). We performed statistical phasing of the genotype data using Shapeit4 and imputation using minimac4 with the in-house reference panel constructed from 11,942 Japanese whole genome sequences. After imputation, we excluded variants with an imputation quality less than 0.7 or a MAF less than 0.01. We used King to exclude relatives

within second degrees.

All of Us

All of Us is a nation-wide cohort in the United States, aiming to recruit persons in demographic categories that have been and continue to be underrepresented in biomedical research[64]. National enrolment began in 2018, and adults 18 years and older who have the capacity to consent and reside in the United States or the United States territory at present are eligible. We analyzed the version 8 dataset of the All of Us.

Quality control of genotype data

We used whole-genome sequencing (WGS) data, which were called and quality-controlled as described elsewhere[79]. We utilized the ACAF-threshold genotype data available in the All of Us workbench, specifically the WGS data for variants with a population-specific allele frequency greater than 1% or a population-specific allele count (AC) greater than 100, in any computed ancestry subpopulations. We used the estimated population labels available in the All of Us workbench and retained the participants with the probability of population assignment greater than 0.5. We excluded relatives with a King kinship coefficient larger than 0.1. As the genotype data were called using the GRCh38 reference genome, we used BCFtools/liftover[80] to convert the genetic positions of our summary statistics from GRCh37.

The Human Phenotype Project (HPP)

HPP is a prospective cohort in Israel, and the participants between the ages of 40 and 70 years were recruited from 2018 to 2021 from the self-assignment of volunteers registered to the trial website[65]. The cohort is largely composed of Ashkenazi Jewish people. Individuals without recent antibiotic use and gastrointestinal morbidity were eligible if they had not yet

been diagnosed with the clinical outcomes of interest.

Quality control of genotype data

In HPP, the entire sequencing and genotyping workflow was performed by Gencove technologies, as described elsewhere[81]. Briefly, DNA was collected from a buccal swab and underwent low-pass WGS with targeting coverage levels of 0.5× and 1×. Paired-end sequencing at a length of 150 bp with Illumina Novaseq was employed for WGS. The genome data were mapped to the hs38-1kg reference genome and imputed using the 1000 Genome Project phase 3 haplotypes as the reference panel.

Quality control of phenotypes and environments in the replication cohorts

Clinical traits

We employed the same definition and quality control for clinical traits as those used in UKB1 and BBJ1. In All of Us, biomarker phenotypes were obtained from the electronic medical records, and we used the most recent records for each biomarker. When both medical records from the emergency department and those from other departments were available, we prioritized the latter. We observed frequent mislabeling of the measurement units in this cohort. For example, albumin measurements with the units of mg/mL and mg/L were of the same order of magnitude as those reported in g/dL. Therefore, we corrected the units by examining the median values for individual unit labels. The biomarker phenotypes for the other cohorts were obtained from the baseline assessment data for the main analyses, while we used the second assessment data of J-MICC for the temporal order analyses. We applied the same quality control criteria as used for BBJ and UKB. For disease statuses, we defined participants as cases if they had relevant SNOMED CT codes or self-reported disease histories in All of Us. We used self-reported disease histories for the other cohorts. For JPHC

and J-MICC/HERPACC, self-reported histories were recorded for diabetes mellitus instead of T2D, and we used diabetes mellitus as a proxy of T2D for replication purposes.

Environmental factors

The environmental factors from questionnaires, along with the conversion of categorical responses into a continuous scale, were summarized in **Supplementary Table 4**. As only one environmental factor was related to physical activity in JPHC, clustering it with dietary environmental factors would lead to a cluster consisting of both physical activity and dietary factors, opposite to our observation that the "physical activities" cluster was consistently detected as a separate cluster from other dietary clusters in UKB1 and BBJ1. Therefore, we pre-defined the "physical activities" cluster from the environmental factors related to physical activities in the replication cohorts and performed clustering only using dietary factors. In JPHC, two sets of questionnaires with slight differences in questionnaire items were used, and we used the items available in both sets. Questionnaires related to physical activities and dietary consumption were unavailable in All of Us, and we did not perform clustering in this cohort. In All of Us, smoking status was separately asked for cigar smoking, electronic smoking, hookah smoking, and smokeless tobacco smoking. We defined ever- and current-smoking statuses based on those of all types of tobacco.

Meta-analyses

We extracted variants with *P*-values less than 1.0×10^{-7} in BBJ1 or UKB1 as the target of meta-analyses. We used six environmental factors shared between biobanks: age, sex, ever-drinking, ever-smoking, current-smoking, and physical activity, as shown in **Ext. Data Fig. 1**. We used the mixmeta R package[82] to perform the inverse-variance weighted fixed-effect meta-analyses and random-effect meta-analyses. As the GEM software estimated G×E

interaction effects conditional on the additive genetic effect[22], we performed meta-analyses for interaction terms and those for the marginal effect separately. We then aggregated *P*-values across environmental factors on the Cauchy distribution[23].

Survival analyses

We used the Cox proportional hazard model to estimate hazard ratios of clinical and metabolite lipid biomarkers on mortality. We used the same covariates as those used for the metabolome G×E analyses, including the binary variable of statin medication. Survival time was defined as the difference between the date of death and the date of the blood sample collection. The death registry's follow-up ended on December 19, 2022. CAD-caused death was defined as cases in which the primary cause of death was I20–I25 (ICD-10).

Association studies stratified by environmental factors

We estimated the genetic associations using linear regression and Firth logistic regression, including age, sex, and the covariates used for the G×E interaction tests as covariates. For the sex-specific analyses, we removed sex, age×sex, and age^2×sex from the covariates. We defined the non-additive effects as deviation from the additive effect, following Palmer D *et al.*[83]. To test non-additive effects, we applied the Gram-Schmidt process weighted by the allele frequencies to individual variants to obtain the orthonormal basis of the genotype variation. We followed Palmer D *et al.*[83] for the accurate Gram-Schmidt process that did not assume the Hardy-Weinberg equilibrium. As the original genotypes could take three possible values (0, 1, and 2 for the number of the alternative allele), the orthonormalized basis also consisted of three bases: the intercept, the additive effect, and the non-additive effect. We tested the non-additive effect without including the additive effect in the regression model by

recoding the genotypes into the orthonormalized non-additive base.

Deep-learning based *in silico* prioritization of regulatory variants

We used a deep-learning model, enformer[73] (https://github.com/google-deepmind/deepmind-research/tree/master/enformer), to predict regulatory effects on gene expression in CD34-positive stem cells for the variants in the *ALDH2* locus. Enformer is a transformer-based model that takes 393,216 bp of DNA sequence as an input, and its center half (196,608 bp, the receptive field) is used for prediction. It simultaneously predicts 5,313 genomic tracks of DNase I hypersensitive sites sequencing (DNase-seq), cap-analysis gene expression sequencing (CAGE-seq), chromatin immunoprecipitation sequencing (ChIP-seq), and assay for transposase-accessible chromatin with sequencing (ATAC-seq) of various human cells and tissues. Enformer returns predicted values for the center 114,688 bp aggregated into 128-bp bins (896 bins in total). We followed the procedure described in the original article[73] to perform prediction. Specifically, given a gene–variant pair, we prepared two 393,216 bp sequences extracted from a human reference genome assembly (described below) centered on the gene's transcription start site. One sequence had the reference allele at the variant position, and the other had the alternative one. After performing prediction, we averaged the predicted values of the four center bins and then computed the log2 fold change (i.e., $\log_2$(alt_expression/ref_expression)) as the predicted regulatory effect. To obtain potential regulatory effects in hematopoietic stem cells, we focused on the predicted values for the CAGE-seq track of CD34+ stem cells derived from adult bone marrow (track #5,141 of the enformer model).

For the target variants, we thought to investigate all single nucleotide variants found in the Japanese population in the *ALDH2* locus. We started with the variants with at least one minor allele observed in the 104 Japanese samples from the high-coverage 1000 genomes

project data[84]. We restricted them to the variants in the LD block containing the *ALDH2* locus, which we defined as 108,631,585–113,088,440 bp of chromosome 12 (GRCh38) based on published East Asian LD block data[85] (https://github.com/jmacdon/LDblocks_GRCh38/blob/master/data/pyrho_EAS_LD_blocks.bed). As the causal gene was unknown, we selected all testable genes (i.e., the genes within half of the receptive field [98,304 bp]) for individual variants. To evaluate potential regulatory effects in the Japanese genome, we used a Japanese human reference genome assembly, JG2.1.0, which was constructed from the *de novo* assembly of three Japanese male genomes[86] (https://jmorp.megabank.tohoku.ac.jp/downloads/tommo-jg2.1.0-20211208). We lifted over gene annotations and variant positions from GRCh38 to JG2.1.0 using transanno (https://github.com/informationsea/transanno).

Heritability estimation

We used linkage disequilibrium score regression (LDSC)[87,88] to estimate the heritability of marginal effects, G×E interactions, and vQTL, as it was shown to be applicable to G×E interactions[37]. When estimating the total heritability of G×E interactions, simply summing up the heritability of individual G×E interaction terms might overestimate the heritability due to the correlations between G×E interaction terms. Therefore, we orthogonalized environmental factors by the Gram-Schmidt process and repeated the G×E interaction tests. We then summed up the heritability of orthogonalized G×E interaction terms to obtain the total heritability of G×E interactions. We estimated the ratio of G×E interaction heritability over marginal heritability by the delta method. When estimating heritability in a liability scale, we used the prevalence in the UK Biobank as the populational prevalence in both biobanks to ensure that the conversion from the observed scale was consistent between biobanks. We estimated standard errors by using a block jackknife over 200 blocks following the default of

LDSC. We calculated *P*-values for heritability using one-sided tests whereas we used two-sided tests for genetic correlation analyses.

Variance quantitative trait loci

We used two statistical methods for vQTL analyses, Brown–Forsythe test and the Deviation Regression Model (DRM). The former is a one-way analysis of variance (ANOVA) of the absolute difference from the median value and was implemented in OmicS-data-based Complex trait Analysis (OSCA v0.46, https://yanglab.westlake.edu.cn/software/osca/)[74]. The latter is a linear regression between genotypes and the absolute difference from the median value[75], and we performed the DRM test using the bgenie tool (v1.3, https://jmarchini.org/bgenie/). As these tests did not directly include covariates, we regressed out age, sex, and the covariates used for the G×E interaction tests from the phenotype values before performing the vQTL analyses. We estimated the heritability of vQTL by applying LDSC to the DRM statistics. We could use LDSC without explicitly extending this method to vQTL, as the DRM test was a form of linear regression.

PGS stratified by environments

We divided the discovery and replication cohorts in UKB and BBJ into two groups based on environmental factors. If the environmental factor was binary (i.e., sex, ever-smoking, current-smoking, and ever-drinking), we divided the cohorts based on the binary status. The discovery sample sizes were balanced across strata to ensure comparable GWAS power. Specifically, in the discovery cohort, we randomly selected the participants from the group with the larger sample size so that the two groups contained the same number of participants. For the other environmental factors, we divided the cohorts into two groups with the same

number of individuals. We performed linear regression using plink2[89] in the groups of the discovery cohorts to obtain GWAS summary statistics. We then constructed PGS from the GWAS summary statistics using one of the state-of-the-art methods, PRScs-auto[90]. PRScs is a Bayesian approach to estimate the variant weights with a continuous shrinkage prior, and PRScs-auto is a version of PRScs that automatically determines hyperparameters. We evaluated the prediction accuracy of PGS in the replication cohort by calculating correlations between PGS and the phenotypic residuals from which age, sex, and the covariates used in the G×E interaction tests were regressed out. We then estimated *P*-values for the difference in these correlations using two-sided Hotelling's *t* test.

Phenotype prediction using PGS constructed from G×E interactions (G×E-PGS)

The G×Eprs framework

We employed the G×Eprs framework[40] to test whether G×E-PGS significantly explains the phenotypic variation. This framework fits the following linear regression model:

$$Y \sim PGS_{additive} + PGS_{G\times E} \times E + PGS_{G\times E} + E$$

where $PGS_{additive}$ is the PGS constructed from additive genetic effects, $PGS_{G\times E}$ is PGS constructed from G×E interaction effects, and E denotes the environmental factor. We also refer to the PGS constructed from marginal effects as $PGS_{marginal}$ below. As G×E-aware PGS construction methods have not been established, we constructed all types of PGS using PRScs-auto[90] with the 1000 Genome Project reference panel for the corresponding population. For the response variable Y, we used phenotype residuals obtained by regressing the raw phenotype values on the same covariates used in the G×E interaction tests. To preserve the variance attributable to environmental factors, we excluded age×sex, age^2, and age^2×sex from the covariates when testing G×Age PGS and excluded age×sex and age^2×sex when testing G×Sex PGS. Following the original study, we evaluated the significance of the

$PGS_{G\times E} \times E$ term using both Wald *P*-values and permutation-based *P*-values, and considered it significant when FDR was less than 0.05 for both *P*-values. Permutation tests were performed with 10,000 iterations by randomly shuffling the values of $PGS_{G\times E} \times E$, as implemented in the R package "GxEprs".

Phenotype prediction using G×E-PGS

We compared two elastic net models for phenotype prediction, one with G×E-PGS and one without G×E-PGS. The baseline model included marginal PGS and the environmental factor (E), assuming a standard PGS constructed based on GWAS summary statistics:

$$Y \sim PGS_{marginal} + E$$

The full model additionally included additive PGS, G×E-PGS, and the interaction with the environment:

$$Y \sim PGS_{marginal} + PGS_{additive} + PGS_{G\times E} + PGS_{G\times E} \times E + E$$

We also included Age^2 into both models when testing G×Age-PGS to account for non-linear variation in phenotype values across age. Models were implemented using the R package "glmnet" (version 4.1.8), which performs regularized regression by linearly combining L1 (Lasso) and L2 (Ridge) penalties. The hyperparameters (Lasso and Ridge penalties) were tuned via a grid search over a range from 0 to 1 in increments of 0.1, and were selected based on 10-fold cross-validation by minimizing the mean squared error on the validation folds. We reported the R^2 from the 10-fold cross-validation in J-MICC/HERPACC as the prediction accuracy.

Pathway enrichment analyses

We ran MAGMA[41] to obtain gene-level *P*-values and perform gene-set enrichment analyses. MAGMA is a simple approach to summarize variant-level *P*-values in a gene level accounting

for LD. We used the gene bodies as the regions of individual genes. We meta-analyzed the gene-level *P*-values from individual biobanks using Stouffer's method weighted with the sample sizes implemented in MAGMA. We excluded the genes in the major histocompatibility complex region (25–34 Mbp in chromosome 6) from the gene-set enrichment analyses and the scDRS analyses described below, considering the complex LD structure of this region.

Single-cell projection of G×E interactions

The Tabula Sapiens data contained the cells sequenced by fluorescence-activated cell sorting (FACS)-sorted cells with smart-seq2 amplification and the ones sequenced by 10x microfluidic droplet capture and amplification[42]. As more cell types were included in the latter subset, we used the latter to reduce batch effects. We subsampled 100 cells per combination of tissues, anatomical regions, and cell types to treat each cell type fairly and to reduce computational costs, resulting in 50,075 cells for 118 cell types.

For the monkey artery scRNA-seq data[43], we downloaded the data from Gene Expression Omnibus (GSE117715). We retained the cells and genes that passed quality control in the original article and applied normalization and scaling using Seurat NormalizeData and ScaleData functions, resulting in 7,989 cells. We then performed UMAP using 10 PCs.

The scDRS method links scRNA-seq with polygenic effects at single-cell resolution by identifying the cells with high gene expression in the trait-associated genes[7]. This method calculates cell-level *P*-values by Monte Carlo sampling of control gene sets containing the same number of genes with matched mean expression and expression variance. This method also calculates the cell-type level *P*-values using the top 5% quantile of the cell-level scores from the given cell type as the test statistic. Considering the high polygenicity of blood pressure traits[91], we used the top 2,000 genes with the highest MAGMA gene-level

associations for the input of scDRS. These genes were weighted by the *Z*-scores of MAGMA gene-level associations as implemented in scDRS. We estimated *P*-values using the one-sided Monte Carlo test with 10,000 sampling and included the number of expressed genes as a covariate. The cell type annotations were defined in the original articles of the data. For the Tabula Sapiens data, we excluded the cell types with less than 75 cells and hepatocytes from the cell-type level analyses, as “hepatocytes” contained the cells from the heart, and they were clustered separately from the hepatocytes from the liver.

Supplementary References

73. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **18**, 1196–1203 (2021).
74. Wang, H. *et al.* Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci. Adv.* **5**, 1–13 (2019).
75. Marderstein, A. R. *et al.* Leveraging phenotypic variability to identify genetic interactions in human phenotypes. *Am. J. Hum. Genet.* **108**, 49–67 (2021).
76. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).
77. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
78. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genet.* **10**, e1004383 (2014).
79. Bick, A. G. *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
80. Genovese, G. *et al.* BCFtools/liftover: an accurate and comprehensive tool to convert genetic variants across genome assemblies. *Bioinformatics* **40**, btae038 (2024).
81. Wasik, K. *et al.* Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genomics* **22**, 197 (2021).
82. Sera, F., Armstrong, B., Blangiardo, M. & Gasparrini, A. An extended mixed-effects framework for meta-analysis. *Stat. Med.* **38**, 5429–5444 (2019).
83. Palmer, D. S. *et al.* Analysis of genetic dominance in the UK Biobank. *Science* **379**, 1341–1348 (2023).
84. Byrska-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440.e19 (2022).
85. Spence, J. P. & Song, Y. S. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci. Adv.* **5**, eaaw9206 (2019).
86. Takayama, J. *et al.* Construction and integration of three de novo Japanese human genome assemblies toward a population-specific reference. *Nat. Commun.* **12**, 226 (2021).
87. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
88. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
89. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

90. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1–10 (2019).
91. Keaton, J. M. *et al.* Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat. Genet.* **56**, 778–791 (2024).