

Supplementary Materials for

PhenomeXcan: Mapping the genome to the phenome through the transcriptome

Milton Pividori, Padma S. Rajagopal, Alvaro Barbeira, Yanyu Liang, Owen Melia, Lisa Bastarache, YoSon Park,
GTEx Consortium, Xiaoquan Wen*, Hae K. Im*

*Corresponding author. Email: xwen@umich.edu (X.W.); haky@uchicago.edu (H.K.I.)

Published 11 September 2020, *Sci. Adv.* **6**, eaba2083 (2020)
DOI: 10.1126/sciadv.aba2083

The PDF file includes:

Legends for tables S1 and S2
Tables S3 to S5
Figs. S1 to S4
Supplementary Methods
References

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/37/eaba2083/DC1)

Tables S1 and S2
Other file

Supplementary Materials

File name: suppl_table_S1-significant_gene_trait_associations.xlsx

Supplementary Table S1: Gene-trait associations in PhenomeXcan that were significant by Bonferroni correction p-value and with locus RCP > 0.1.

This table contains all **22,219 gene-trait associations** with p-value < 5.49×10^{-10} and locus RCP > 0.1.

File name: suppl_table_S2-UKBiobank_to_OMIM-standard.xlsx

Supplementary Table S2: Standard of OMIM gene-trait associations used to validate PhenomeXcan

This table contains 7,809 high-confidence gene-trait associations from OMIM that were used to evaluate the performance of PhenomeXcan.

Strength of Evidence	Gene List
Causal	<i>ABCC8, ANGPTL4, ANKH, APOE, CDKN1B, GCK, GCKR, GIPR, GLIS3, GLP1R, HNF1A, HNF1B, HNF4A, IGF2, INS, IRS2, KCNJ11, LPL, MC4R, MNX1, MTNR1B, NEUROG3, NKX2-2, PAM, PATJ, PAX4, PDX1, PLCB3, PNPLA3, POC5, PPARG, QSER1, RREB1, SLC16A11, SLC30A8, SLC5A1, TBC1D4, TM6SF2, WFS1, WSCD2, ZNF771</i>
Strong	<i>ABCB9, BCAR1, C2CD4B, CAMK1D, CCND2, DGKB, INSR, IRS1, IRX3, IRX5, KLF14, KLHL42, LMNA, SLC2A2, STARD10, TCF7L2, ZMIZ1</i>
Moderate	<i>ADCY5, AGPAT2, AGTR2, AP3S2, BCL11A, CISD2, FAM63A, FOXA2, GPSM1, IGF2BP2, JAZF1, KCNK17, MACF1, MADD, NKX6-3, PDE8B, PLIN1, SGSM2, SPRY2, UBE2E2, VPS13C</i>
Possible	<i>ANK1, ASCC2, CALCOCO2, FADS1, HMG20A, IL17REL, MRPS30, PRC1, PTRF, SCD5, SNAPC4, ST6GAL1, TP53INP1</i>
Weak	<i>ABO, CARD9, CDK2AP1, CTNNA1, DNZL, ITGB6</i>
Related Traits	<i>ADRA2A, AKT2, APPL1, BLK, BSCL2, CAV1, CEL, EIF2AK3, ERAP2, FOXP3, G6PC2, G6PD, GATA4, GATA6, GCG, GRB10, IER3IP1, IGF1, KLF11, NAT2, NEUROD1, PAX6, PCBD1, PCSK1, POLD1, PPP1R15B, PTF1A, RFX6, SIX2, SIX3, SLC19A2, TRMT10A, WARS, ZFP57</i>

Supplementary Table S3: List of curated genes and strengths of evidence for causal association to type 2 diabetes.

This list is curated by the Accelerating Medicines Partnership: Type 2 Diabetes Knowledge Portal.

Gene	Chromosome	p-value	Locus RCP	Number of UK Biobank GWAS	Number of non- UK Biobank GWAS	Number of clinical or functional studies focused on sleep or chronotype mechanisms
<i>TRAF3IP1</i>	2q37.3	1.724e-20	0.40	4	1	0
<i>RASA4B</i>	7q22.1	1.660e-19	0.63	1	0	1
<i>CLN5</i>	13q22.3	5.248e-18	0.33	4	1	3
<i>VAMP3</i>	1p36.23	7.317e-18	0.63	0	0	0
<i>VIP</i>	6q25.2	1.812e-17	0.26	0	1	7
<i>FBXL3</i>	13q22.3	1.545e-16	0.35	4	1	29
<i>LINC01470</i>	5q33.1	4.161e-16	0.10	3	0	0
<i>TNRC6B</i>	22q13.1	8.441e-14	0.20	6	1	0
<i>RASD1</i>	17p11.2	1.246e-12	0.22	4	1	0
<i>ZCCHC7</i>	9p13.2	4.282e-11	0.25	2	0	0
<i>RP11-220I1.5</i>	9	6.427e-11	0.20	0	0	0

<i>EBLN3P</i>	9p13.2	6.853e-11	0.91	2	0	0
<i>RASL10B</i>	17q12	1.098e-10	0.15	0	0	0
<i>PMFBP1</i>	16q22.2	1.413e-10	0.85	4	0	0
<i>DDI2</i>	1p36.21	2.156e-10	0.26	4	0	0

Supplementary Table S4: Summary of genes and evidence associated with UK Biobank trait “Morning/evening person (chronotype)”. Genes are sorted by PrediXcan p-value for the best tissue expression, with locus regional colocalization probability (locus RCP) higher than 0.1. Higher p-values and locus RCP scores suggest greater likelihood of causal association to the trait. Evidence is organized by gene reports in GWAS using UK Biobank subjects, GWAS not using UK Biobank subjects, and clinical/functional studies. GWAS were identified using the NHGRI-EBI GWAS catalog (10/21/2019), and functional/clinical studies were identified from PubMed using searches for the gene name as well as the gene name/trait category and gene name/trait.

Category	Trait	Abbreviation in PhenomeXcan	Sample Size
Psychiatric-neurologic	CNCR Insomnia all	INSOMN	113006
Psychiatric-neurologic	IGAP Alzheimer	AD	54162
Psychiatric-neurologic	Jones et al 2016 Chronotype	CHRONO	128266
Psychiatric-neurologic	Jones et al 2016 SleepDuration	SLEEP	128266
Psychiatric-neurologic	PGC ADHD EUR 2017	ADHD	53293
Psychiatric-neurologic	pgc.scz2	SCZ	150064

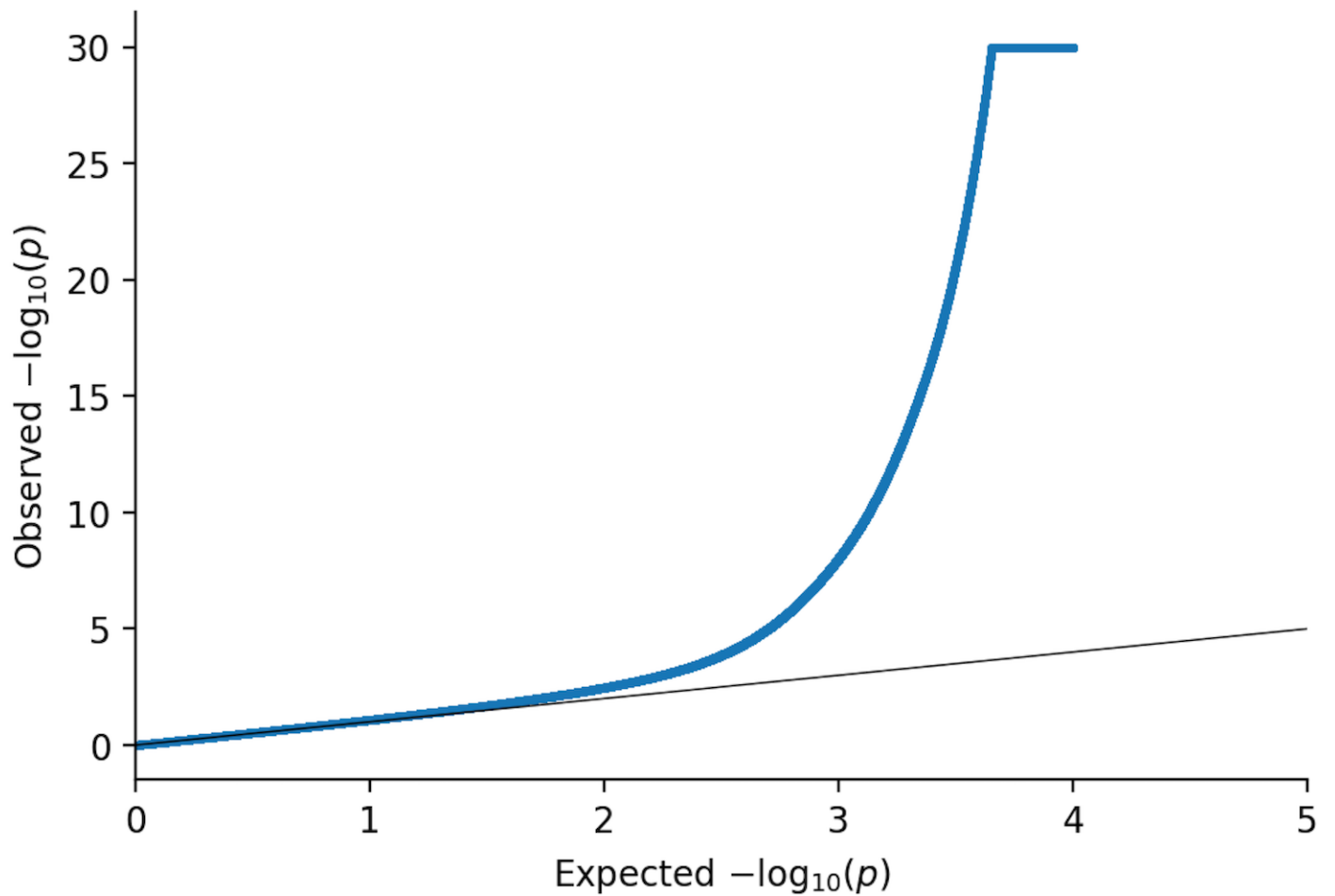
Psychiatric-neurologic	SSGAC Depressive Symptoms	DEPR	180866
Psychiatric-neurologic	SSGAC Education Years Pooled	EDU	293723
Anthropometric	EGG BW3 EUR	BW	143677
Anthropometric	ENIGMA Intracranial Volume	ICV	30717
Anthropometric	GEFOS Forearm	BMD	49988
Anthropometric	Giant Height	HEIGHT	253288
Cardiometabolic	CARDIoGRAM C4D CAD ADDITIVE	CAD	184305
Cardiometabolic	MAGIC FastingGlucose	FG	46186
Cardiometabolic	MAGIC In FastingInsulin	INSUL	38238
Cardiometabolic	MAGNETIC CH2.DB.ratio	CH2	24154
Cardiometabolic	MAGNETIC HDL.C	HDL.C	19270
Cardiometabolic	MAGNETIC IDL.TG	IDL	21559
Cardiometabolic	MAGNETIC LDL.C	LDL.C	13527
Blood	Astle et al 2016 Eosinophil counts	EC	173480
Blood	Astle et al 2016 Granulocyte count	GC	173480
Blood	Astle et al 2016 High light scatter reticulocyte count	HRET	173480

Blood	Astle et al 2016 Lymphocyte counts	LC	173480
Blood	Astle et al 2016 Monocyte count	MC	173480
Blood	Astle et al 2016 Myeloid white cell count	MWBC	173480
Blood	Astle et al 2016 Neutrophil count	NC	173480
Blood	Astle et al 2016 Platelet count	PLT	173480
Blood	Astle et al 2016 Red blood cell count	RBC	173480
Blood	Astle et al 2016 Reticulocyte count	RET	173480
Blood	Astle et al 2016 Sum basophil neutrophil counts	BNC	173480
Blood	Astle et al 2016 Sum eosinophil basophil counts	EBC	173480
Blood	Astle et al 2016 Sum neutrophil eosinophil counts	NEC	173480
Blood	Astle et al 2016 White blood cell count	WBC	173480
Cancer	BCAC ER negative BreastCancer EUR	ERNBC	120000
Cancer	BCAC ER positive BreastCancer EUR	ERPBC	120000
Cancer	BCAC Overall BreastCancer EUR	BC	120000
Allergy	EAGLE Eczema	ECZ	116863

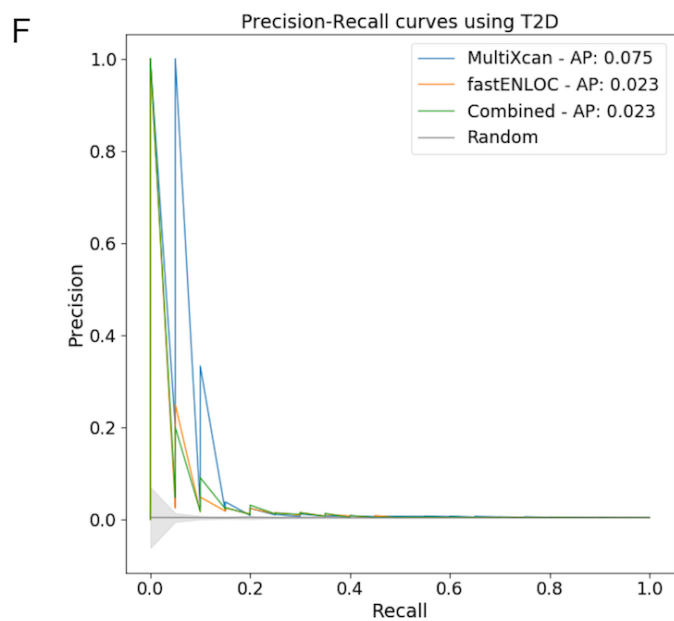
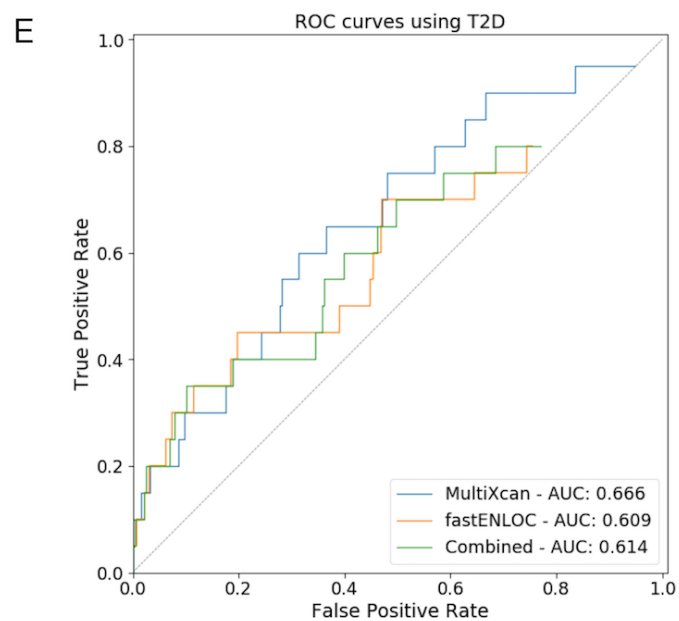
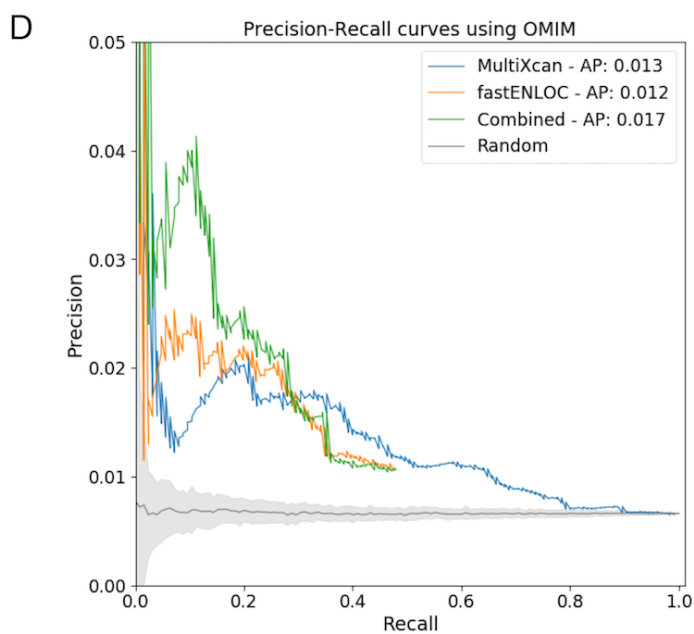
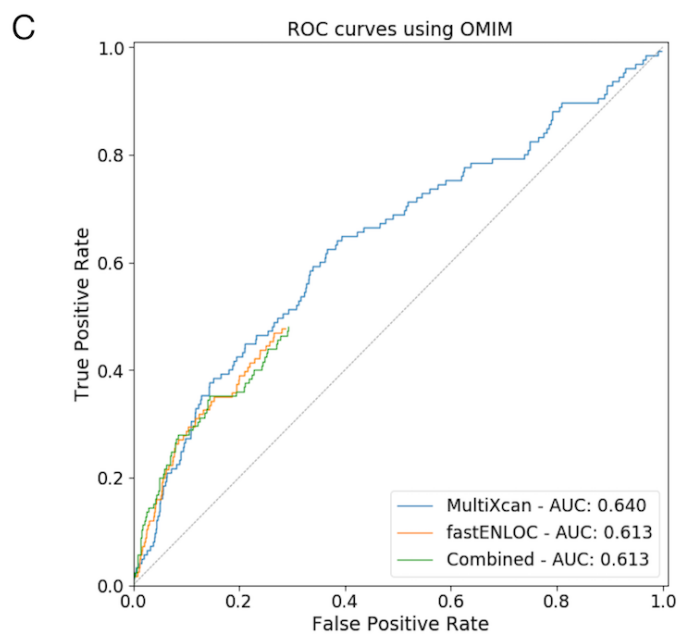
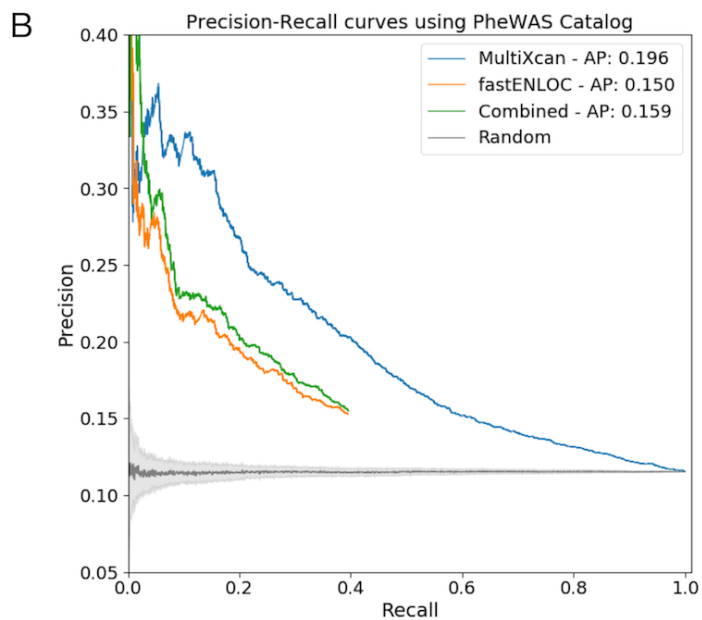
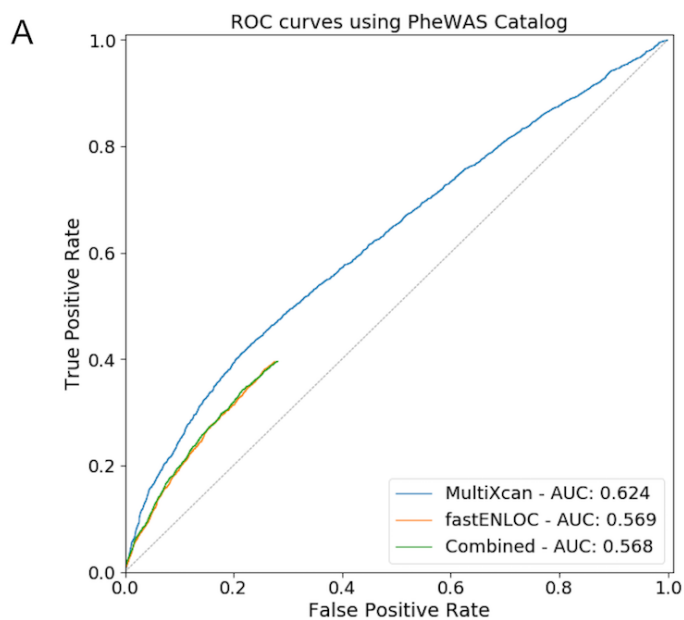
Immune	IBD.EUR.Crohns Disease	CD	20833
Immune	IBD.EUR.Inflammatory Bowel Disease	IBD	34652
Immune	IBD.EUR.Ulcerative Colitis	UC	27432
Immune	IMMUNOBASE Systemic lupus erythematosus hg19	SLE	23210
Immune	RA OKADA TRANS ETHNIC	RA	80799

Supplementary Table S5: 42 additional traits taken from GWAS studies for the development of PhenomeXcan.

Traits are organized by trait category, data source, abbreviation in PhenomeXcan and number of subjects in the dataset.

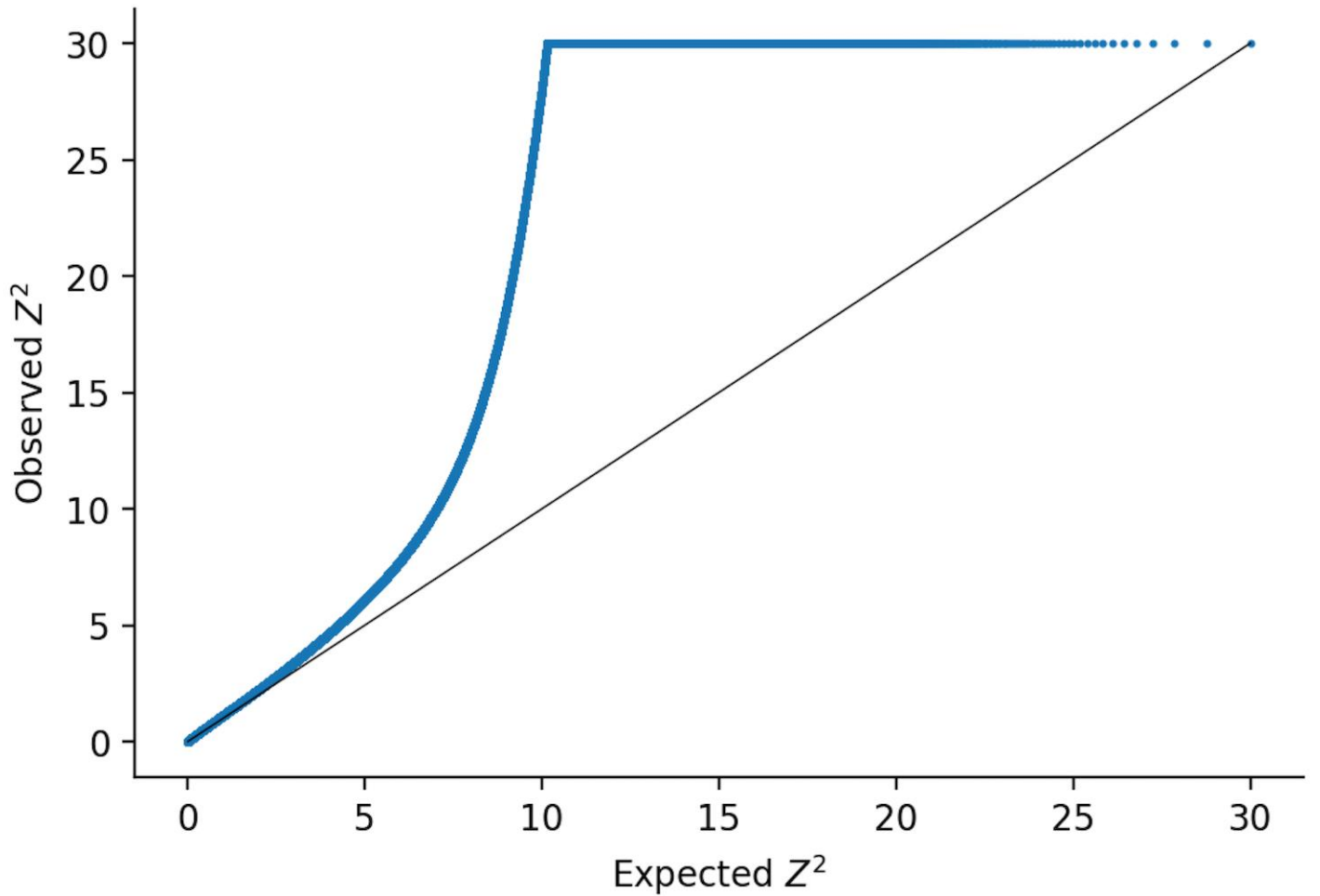


Supplementary Fig. S1: Quantile-quantile (QQ) plot of all associations in PhenomeXcan. The expected null distribution is plotted along the black diagonal, and the entire distribution of observed p-values is plotted in blue. We do not see evidence of systematic inflation given the initial consistency in expected and observed p-values. (To improve visualization, p-values are thresholded at $-\log_{10}(\text{p-value})=30$.) The increase in the QQ plot for observed p-values can be seen with the extremely large number of associations tested.

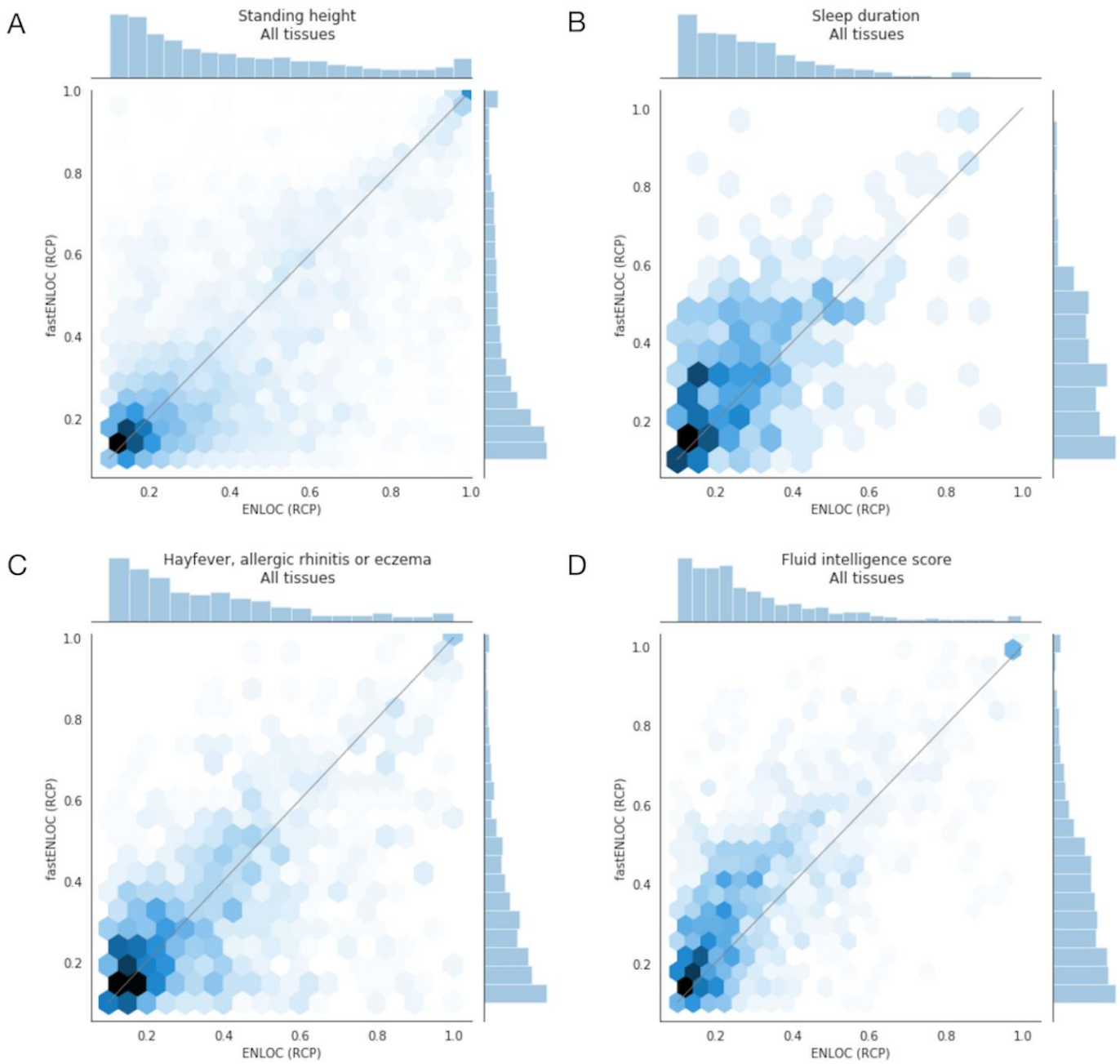


Supplementary Fig. S2: Combined performance of PhenomeXcan methods (S-MultiXcan and fastENLOC) across the PheWAS Catalog, OMIM data sets and T2D reported genes. MultiXcan refers to the version of PrediXcan designed to take GWAS summary statistics and aggregate results across tissues. (A, B) ROC curve and PR curve of PrediXcan significance scores (blue), fastENLOC (orange) and combined methods (green) to predict PheWAS catalog gene-trait associations. (C, D) ROC curve and PR curve of PrediXcan significance scores (blue), fastENLOC (orange) and combined methods (green) to predict OMIM catalog gene-trait associations. (E, F) ROC curve and PR curve of PrediXcan significance scores (blue), fastENLOC (orange) and combined methods (green) to predict T2D catalog gene-trait associations. AUC refers to the area under the curve, AP refers to average precision. The maximum fastENLOC colocalization probability across tissues was used for all figures. Points with a minimum colocalization probability of 0.1 were included for the combined analysis.

The variations in performance of the combined approach reflect the conservative nature of colocalization analysis, which can lead to increased false negatives. However, we do see that the combined approach improves the precision (proportion of “true causal” genes, i.e. the true discovery rate) for OMIM genes. Precision-recall curves better reflect relative performance in this dataset given the class imbalance between “true” associations and “negative” associations. For the PheWAS-based silver standard, precision is better than MultiXcan only at the very low end of the recall range. This reflects the correlation between S-MultiXcan and PheWAS criterion to call a causal gene and should not be over-interpreted.



Supplementary Fig. S3: Quantile-quantile (QQ) plot of all associations in PhenomeXcan and Clinvar traits. The expected χ^2 null distribution is plotted along the black diagonal, and the entire distribution of observed Z^2 is plotted in blue. We do not see evidence of systematic inflation given the initial consistency in expected and observed p-values. (To improve visualization, Z^2 are thresholded at 30.) The increase in the QQ plot for observed p-values can be seen with the extremely large number of associations tested (20.6 million) as well as the pleiotropy we identify with trait-trait associations in which multiple genes are involved. Z^2 correspondence to percentiles were as follows: 95th percentile: $Z^2=4.45$, 99th percentile: $Z^2=9.07$, 99.9th percentile: $Z^2=214.45$. A Z^2 of 6 represents a Bonferroni-adjusted p-value of 0.05.



Supplementary Fig. S4: Joint histograms using hexagonal bins for the regional colocalization probability (RCP) agreement between fastENLOC and ENLOC. We analyzed the regional colocalization probabilities across traits between fastENLOC and ENLOC to assess their agreement. We found largely strong correlation between these methods, with the Spearman correlation coefficient for **(A)** “Standing height” = 0.61, **(B)** “Sleep duration” = 0.52, **(C)**, “Hayfever, allergic rhinitis or eczema” = 0.57 and **(D)** “Fluid intelligence score” = 0.65.

List of GTex Consortium Authors:

Laboratory and Data Analysis Coordinating Center (LDACC): François Aguet¹, Shankara Anand¹, Kristin G Ardlie¹, Stacey Gabriel¹, Gad Getz^{1,2}, Aaron Graubert¹, Kane Hadley¹, Robert E Handsaker^{3,4,5}, Katherine H Huang¹, Seva Kashin^{3,4,5}, Xiao Li¹, Daniel G MacArthur^{4,6}, Samuel R Meier¹, Jared L Nedzel¹, Duyen Y Nguyen¹, Ayellet V Segrè^{1,7}, Ellen Todres¹

Analysis Working Group (funded by GTEx project grants): François Aguet¹, Shankara Anand¹, Kristin G Ardlie¹, Brunilda Balliu⁸, Alvaro N Barbeira⁹, Alexis Battle^{10,11}, Rodrigo Bonazzola⁹, Andrew Brown^{12,13}, Christopher D Brown¹⁴, Stephane E Castel^{15,16}, Don Conrad^{17,18}, Daniel J Cotter¹⁹, Nancy Cox²⁰, Sayantan Das²¹, Olivia M de Goede¹⁹, Emmanouil T Dermitzakis^{12,22,23}, Barbara E Engelhardt^{24,25}, Eleazar Eskin²⁶, Tiany Y Eulalio²⁷, Nicole M Ferraro²⁷, Elise Flynn^{15, 16}, Laure Fresard²⁸, Eric R Gamazon^{20, 29, 30, 31}, Diego Garrido-Martín³², Nicole R Gay¹⁹, Gad Getz^{1,2}, Aaron Graubert¹, Roderic Guigó^{32, 33}, Kane Hadley¹, Andrew R Hamel^{1, 7}, Robert E Handsaker^{3,4,5}, Yuan He¹⁰, Paul J Homan¹⁵, Farhad Hormozdiari^{1,34}, Lei Hou^{1, 35}, Katherine H Huang¹, Hae Kyung Im⁹, Brian Jo^{24, 25}, Silva Kasela^{15, 16}, Seva Kashin^{3,4,5}, Manolis Kellis^{1,35}, Sarah Kim-Hellmuth^{15, 16, 36}, Alan Kwong²¹, Tuuli Lappalainen^{15, 16}, Xiao Li¹, Xin Li²⁸, Yanyu Liang⁹, Daniel G MacArthur^{4,6}, Serghei Mangul^{26, 37}, Samuel R Meier¹, Pejman Mohammadi^{15, 16, 38, 39}, Stephen B Montgomery^{19, 28}, Manuel Muñoz-Aguirre^{32, 40}, Daniel C Nachun²⁸, Jared L Nedzel¹, Duyen Y Nguyen¹, Andrew B Nobel⁴¹, Meritxell Oliva^{9,42}, YoSon Park^{14,43}, Yongjin Park^{1,35}, Princy Parsana¹¹, Ferran Reverter⁴⁴, John M Rouhana^{1,7}, Chiara Sabatti⁴⁵, Ashis Saha¹¹, Ayellet V Segrè^{1,7}, Andrew D Skol^{9,46}, Matthew Stephens⁴⁷, Barbara E Stranger^{9,48}, Benjamin J Strober¹⁰, Nicole A Teran²⁸, Ellen Todres¹, Ana Viñuela^{12,22,23,49}, Gao Wang⁴⁷, Xiaoquan Wen²¹, Fred Wright⁵⁰, Valentin Wucher³², Yuxin Zou⁵¹

Analysis Working Group (not funded by GTEx project grants): Pedro G Ferreira^{52,53,54}, Gen Li⁵⁵, Marta Melé⁵⁶, Esti Yeger-Lotem^{57,58}, Leidos Biomedical - Project Management: Mary E Barcus⁵⁹, Debra Bradbury⁶⁰, Tanya Krubit⁶⁰, Jerrey A McLean⁶⁰, Liquan Qi⁶⁰, Karna Robinson⁶⁰, Nancy V Roche⁶⁰, Anna M Smith⁶⁰, Leslie Sobin⁶⁰, David E Tabor⁶⁰, Anita Undale⁶⁰

Biospecimen collection source sites: Jason Bridge⁶¹, Lori E Brigham⁶², Barbara A Foster⁶³, Bryan M Gillard⁶³, Richard

Hasz⁶⁴, Marcus Hunter⁶⁵, Christopher Johns⁶⁶, Mark Johnson⁶⁷, Ellen Karasik⁶³, Gene Kopen⁶⁸, William F Leinweber⁶⁸, Alisa McDonald⁶⁸, Michael T Moser⁶³, Kevin Myer⁶⁵, Kimberley D Ramsey⁶³, Brian Roe⁶⁵, Saboor Shad⁶⁸, Jerey A Thomas^{67,68}, Gary Walters⁶⁷, Michael Washington⁶⁷, Joseph Wheeler⁶⁶

Biospecimen core resource: Scott D Jewell⁶⁹, Daniel C Rohrer⁶⁹, Dana R Valley⁶⁹

Brain bank repository: David A Davis⁷⁰, Deborah C Mash⁷⁰

Pathology: Mary E Barcus⁵⁹, Philip A Branton⁷¹, Leslie Sobin⁶⁰

ELSI study: Laura K Barker⁷², Heather M Gardiner⁷², Maghboeba Mosavel⁷³, Laura A Simino⁷²

Genome Browser Data Integration & Visualization: Paul Flicek⁷⁴, Maximilian Haeussler⁷⁵, Thomas Juettemann⁷⁴, W James Kent⁷⁵, Christopher M Lee⁷⁵, Conner C Powell⁷⁵, Kate R Rosenbloom⁷⁵, Magali Ru-er⁷⁴, Dan Sheppard⁷⁴, Kieron Taylor⁷⁴, Stephen J Trevanion⁷⁴, Daniel R Zerbino⁷⁴

eGTEx groups: Nathan S Abell¹⁹, Joshua Akey⁷⁶, Lin Chen⁴², Kathryn Demanelis⁴², Jennifer A Doherty⁷⁷, Andrew P Feinberg⁷⁸, Kasper D Hansen⁷⁹, Peter F Hickey⁸⁰, Lei Hou^{1,35}, Farzana Jasmine⁴², Lihua Jiang¹⁹, Rajinder Kaul^{81,82}, Manolis Kellis^{1,35}, Muhammad G Kibriya⁴², Jin Billy Li¹⁹, Qin Li¹⁹, Shin Lin⁸³, Sandra E Linder¹⁹, Stephen B Montgomery^{19,28}, Meritxell Oliva^{9,42}, Yongjin Park^{1,35}, Brandon L Pierce⁴², Lindsay F Rizzardi⁸⁴, Andrew D Skol^{9,46}, Kevin S Smith²⁸, Michael Snyder¹⁹, John Stamatoyannopoulos^{81,85}, Barbara E Stranger^{9,48}, Hua Tang¹⁹, Meng Wang¹⁹

NIH program management: Philip A Branton⁷¹, Latarsha J Carithers^{71,86}, Ping Guan⁷¹, Susan E Koester⁸⁷, A. Roger Little⁸⁸, Helen M Moore⁷¹, Concepcion R Nierras⁸⁹, Abhi K Rao⁷¹, Jimmie B Vaught⁷¹, Simona Volpi⁹⁰

Affiliations:

1. The Broad Institute of MIT and Harvard, Cambridge, MA, USA
2. Cancer Center and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA

3. Department of Genetics, Harvard Medical School, Boston, MA, USA
4. Program in Medical and Population Genetics, The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA
5. Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA
6. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
7. Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA
8. Department of Biomathematics, University of California, Los Angeles, Los Angeles, CA, USA
9. Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL, USA
10. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
11. Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
12. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
13. Population Health and Genomics, University of Dundee, Dundee, Scotland, UK
14. Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
15. New York Genome Center, New York, NY, USA
16. Department of Systems Biology, Columbia University, New York, NY, USA
17. Department of Genetics, Washington University School of Medicine, St. Louis, Missouri, USA
18. Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, Missouri, USA
19. Department of Genetics, Stanford University, Stanford, CA, USA
20. Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
21. Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
22. Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland
23. Swiss Institute of Bioinformatics, Geneva, Switzerland
24. Department of Computer Science, Princeton University, Princeton, NJ, USA
25. Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA
26. Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA
27. Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA, USA
28. Department of Pathology, Stanford University, Stanford, CA, USA

29. Data Science Institute, Vanderbilt University, Nashville, TN, USA
30. Clare Hall, University of Cambridge, Cambridge, UK
31. MRC Epidemiology Unit, University of Cambridge, Cambridge, UK
32. Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain
33. Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain
34. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
35. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
36. Statistical Genetics, Max Planck Institute of Psychiatry, Munich, Germany
37. Department of Clinical Pharmacy, School of Pharmacy, University of Southern California, Los Angeles, CA, USA
38. Scripps Research Translational Institute, La Jolla, CA, USA
39. Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA
40. Department of Statistics and Operations Research, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain
41. Department of Statistics and Operations Research and Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA
42. Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA
43. Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA
44. Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona. Spain.
45. Departments of Biomedical Data Science and Statistics, Stanford University, Stanford, CA, USA
46. Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
47. Department of Human Genetics, University of Chicago, Chicago, IL, USA
48. Center for Genetic Medicine, Department of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

49. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
50. Bioinformatics Research Center and Departments of Statistics and Biological Sciences, North Carolina State University, Raleigh, NC, USA
51. Department of Statistics, University of Chicago, Chicago, IL, USA
52. Department of Computer Sciences, Faculty of Sciences, University of Porto, Porto, Portugal
53. Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal
54. Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal
55. Columbia University Mailman School of Public Health, New York, NY, USA
56. Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
57. Department of Clinical Biochemistry and Pharmacology, Ben-Gurion University of the Negev, Beer-Sheva, Israel
58. National Institute for Biotechnology in the Negev, Beer-Sheva, Israel
59. Leidos Biomedical, Frederick, MD, USA
60. Leidos Biomedical, Rockville, MD, USA
61. UNYTS, Buffalo, NY, USA
62. Washington Regional Transplant Community, Annandale, VA, USA
63. Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA
64. Gift of Life Donor Program, Philadelphia, PA, USA
65. LifeGift, Houston, TX, USA
66. Center for Organ Recovery and Education, Pittsburgh, PA, USA
67. LifeNet Health, Virginia Beach, VA, USA
68. National Disease Research Interchange, Philadelphia, PA, USA
69. Van Andel Research Institute, Grand Rapids, MI, USA
70. Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA
71. Biorepositories and Biospecimen Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD, USA
72. Temple University, Philadelphia, PA, USA
73. Virginia Commonwealth University, Richmond, VA, USA

74. European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom v 75. Genomics Institute, UC Santa Cruz, Santa Cruz, CA, USA
76. Carl Icahn Laboratory, Princeton University, Princeton, NJ, USA
77. Department of Population Health Sciences, The University of Utah, Salt Lake City, Utah, USA
78. Schools of Medicine, Engineering, and Public Health, Johns Hopkins University, Baltimore, MD, USA
79. Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
80. Department of Medical Biology, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia
81. Altius Institute for Biomedical Sciences, Seattle, WA, USA
82. Division of Genetics, University of Washington, Seattle, WA, University of Washington, Seattle, WA, USA
83. Department of Cardiology, University of Washington, Seattle, WA, USA
84. HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
85. Genome Sciences, University of Washington, Seattle, WA, USA
86. National Institute of Dental and Craniofacial Research, Bethesda, MD, USA
87. Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA
88. National Institute on Drug Abuse, Bethesda, MD, USA
89. Office of Strategic Coordination, Division of Program Coordination, Planning and Strategic Initiatives, Office of the Director, National Institutes of Health, Rockville, MD, USA
90. Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD, USA

Supplementary Methods:

Preprocessing of Traits and Quality Control of Variants

We developed PhenomeXcan with 4,091 traits from publicly available GWAS summary statistics. Summary statistics from GWAS performed for 4,049 traits from the UK Biobank (on 361,194 samples) were obtained from the publicly available dataset compiled by the Neale Lab at the Broad Institute (61); we did not use individual-level data. The UK Biobank is a prospective cohort of approximately 500,000 subjects between 40 and 69 years of age, recruited from 2006-2010 in the United Kingdom (62). Traits characterized by the Neale lab include 2,891 auto-curated traits using PHESANT (63), of which 274 are continuous, 271 ordinal and 2,346 binary. 633 binary traits were extracted from hospital-level data (ICD-10 codes). 559 traits were manually curated in collaboration with the FinnGen Consortium. Traits available cover a range of categories, from lifestyle traits and socio-demographic questions to clinical biomarkers and diagnoses. Separate sex-specific summary statistics and sex chromosome analyses were not included in this project. More details on the GWAS derivations and quality control is provided in the website of the project: <http://www.nealelab.is/uk-biobank>. We do note that for these GWAS, 361,194 individuals were selected for inclusion based on quality of genotypes, white British ancestry (based on both self-report and principal components analysis). Only those variants with an imputation quality score (INFO) > 80%, a minor allele frequency (MAF) > 0.1%, call rate > 95% and a Hardy-Weinberg equilibrium p-value > 1×10^{-10} were selected.

We also compiled 42 additional traits from summary statistics from publicly available GWAS and GWAS-meta analyses external to the UK Biobank study both to validate synthesis of additional GWAS data and to overcome limitations related to poor sample sizes in the UK Biobank for specific diseases (e.g. breast cancer). These GWAS and traits represent a broad array of disease-related categories, including immunological response, psychiatric and neurologic traits, cardiometabolic diseases and syndromes and cancer. We have previously described the harmonization and imputation process (24) (Supplementary Table S5).

ClinVar is a publicly available archive of clinically reported human genetic variants and associations with disease maintained by the National Institutes of Health (<https://www.ncbi.nlm.nih.gov/clinvar/>). Variant associations with disease are identified by manual review of submitted interpretations from “clinical testing laboratories, research laboratories, locus-

specific databases, Online Mendelian Inheritance of Man (OMIM), GeneReviews, UniProt, expert panels and practice guidelines” (30, 48). Traits can be reported to ClinVar as a single concept or set of clinical features. When possible, traits are mapped manually to standardized terms from databases including OMIM and the Human Phenotype Ontology (HPO) (29). All gene-trait associations published by ClinVar for 7/2019 were used for integration with PhenomeXcan.

fastENLOC

fastENLOC is a novel computational method built upon the existing colocalization analysis framework of ENLOC (23). As in the original ENLOC, fastENLOC estimates the relative enrichment of causal eQTLs in the GWAS hits. This is a distinct feature from other existing colocalization approaches. For example, eCAVIAR assumes that the probability of a variant being causal for a trait is independent of the probability of the variant causally affecting gene expression, which results in rapid processing but can be too conservative (64); whereas “coloc” allows users to specify enrichment parameter subjectively (65). However, the inappropriate parameter setting can introduce systematic false-positive and false-negative colocalization findings (64, 65). fastENLOC also follows the ENLOC approach to compute and report both SNP-level colocalization probabilities (SCPs) and region/locus-level colocalization probabilities (RCPs), although the probability computation is improved and the selection of genetic loci becomes fully automatic in the new method.

The main innovation in the fastENLOC lies in its explicit usage of signal clusters/Bayesian credible sets computed from the Bayesian fine-mapping analysis. A signal cluster is a group of SNPs in LD and represents the same underlying genetic association signal. Currently, we note that fine-mapping algorithms DAP-G and SUSIE-R can generate the required signal clusters. The utilization of pre-computed signal cluster information from both GWAS and eQTL data greatly speeds up the overall colocalization analysis and achieves higher accuracy through better probability calibration and explicit accounting of allelic heterogeneity. This computation saving is particularly important in this application, where each of the 4000+ traits is tested for colocalization with eQTLs in 49 different tissues. In a benchmark test, we record the computational time of ENLOC and fastENLOC for colocalization analysis of whole blood eQTL data from the GTEx and the high-density cholesterol GWAS data from the Global Lipid Genetic Consortium (GLGC) (66). In this experiment, we provide fine-mapped eQTL results from DAP-G and single-SNP z-scores from GLGC. On a Linux system with Xeon 2.13

GHz CPUs and 48 GB memory, the total processing times for fastENLOC and ENLOC were 6 minutes 44 seconds and 219 minutes 20 seconds, respectively. While the methodological innovation contributes to this performance improvement, the new computational implementation of fastENLOC (e.g., by reducing redundant I/O operations) also has a significant impact. Finally, we keep fastENLOC backwards compatible with the original ENLOC: in the simplest case, single-SNP association summary statistics from GWAS and eQTL analysis are sufficient to start the fastENLOC analysis with a fast pre-processing step via software package TORUS. The details are provided here: <https://github.com/xqwen/fastenloc>.

Next, we provide an overview of the fastENLOC analysis procedure. The overall procedure includes two interconnected steps: enrichment estimation and colocalization computation. The enrichment estimation aims to estimate the enrichment level of causal eQTLs in causal GWAS hits, or equivalently, the conditional prior probability that a SNP being a causal GWAS hit given its causal eQTL status. Because the true causal SNPs for both eQTL and GWAS are not observed, and the results from the association analysis carry a great deal of uncertainty due to LD, we adopt a multiple imputation (MI) approach to sample the true causal SNPs from different traits then average the enrichment estimates over multiple independent sets of imputed causal status. As a unique feature, fastENLOC implements a novel hierarchical sampling procedure based on the signal cluster information: It first draws a Bernoulli random variable based on the cumulative SNP-level posterior inclusion probabilities (PIPs) for a given signal cluster to determine if a cluster contains a causal variant; conditional on a positive outcome, the actual causal SNP is subsequently drawn from the member SNPs within the cluster. In comparison, the original ENLOC imputation procedure performs independent sampling based on SNP-level PIPs. This unstructured sampling procedure has some obvious caveats that are remedied in the new fastENLOC procedure. For example, the fastENLOC scheme ensures at most a single causal variant can be drawn from each signal cluster, whereas the ENLOC procedure does not enforce such desirable constraint. As a result of this new imputation procedure, we observe that the fastENLOC yields more accurate enrichment estimate and the sampling variance from the multiple imputation procedure is also lower than the original ENLOC.

In summary, the faster speed and the improved performance of fastENLOC are achieved with the following changes:

- Pre-computation of posterior inclusion probabilities (fine-mapping) of eQTLs, which could be reused for any trait

- Pre-computation of posterior inclusion probabilities (fine-mapping) of GWAS, which could be re-used for each of the 49 tissues
- Faster calculation of enrichment parameters (α_0, α_1) by taking advantage of the pre-computed credible sets
- Analytic update of the posterior probability of fine-mapping and colocalization using the posterior probability input from both GWAS and eQTL fine-mapping
- A new computational implementation using C++
- Structured sampling procedure based on the signal clustering

When a GWAS identifies few association signals, the enrichment estimate can be highly unstable, i.e., the point estimate is outside the normal range and the standard error is extremely large. In such cases, a desired outcome is to shrink the estimate to near 0. Through analyzing 4000+ traits with 49 GTEx tissues, we find the adaptive shrinkage procedure proposed in the ENLOC occasionally is insufficient to stabilize the point estimate. Thus, we implement a new shrinkage approach in fastENLOC to deal with these scenarios.

Briefly, we propose a normal prior, $N(0, 1/\lambda)$, for the enrichment parameter α_1 , and report its posterior mean for the downstream colocalization computation. Specifically,

$$\widehat{\alpha}_1^{\text{shrinkage}} = \frac{\widehat{\alpha}_1}{1 + \lambda s_1^2},$$

$$s_1^{\text{shrinkage}} = \sqrt{\frac{s_1^2}{1 + \lambda s_1^2}},$$

Where $\widehat{\alpha}_1$ and s_1 are estimates from the multiple imputation procedure. The shrinkage parameter λ , which can be customized by the users, quantifies the strength of the shrinkage: $\lambda \rightarrow 0$ represents no shrinkage and $\lambda \rightarrow \infty$ shrinks any estimate to exactly 0. By default, we find set $\lambda = 1$ achieves a good balance in practice, where we find the enrichment parameters are well stabilized with standard error ≤ 1 .

Despite the new shrinkage scheme, the locus RCPs calculated by ENLOC and fastENLOC show good agreement (Supplementary Figure S4). This is likely because the instability of the enrichment estimates typically indicate a lack of colocalized signals. (On the contrary, the noteworthy colocalization probabilities are typically computed from those traits

where enrichment parameters can be stably computed even with weak or no shrinkages.)

Given the enrichment parameters $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2)$, we derive the analytic formula to calculate the colocalization probability using the pre-computed posterior inclusion probabilities and the signal cluster information from fine-mapping results of GWAS and eQTL traits.

Let \mathbf{D}, \mathbf{E} denote the association data from GWAS and eQTL analyses, respectively. We consider a signal cluster inferred from the fine-mapping analysis of either eQTLs or GWAS and use latent binary indicator p -vectors \mathbf{d}, \mathbf{y} to represent the causal association status of its p member single-nucleotide polymorphisms (SNPs) with the complex trait and the gene expression level of interest, respectively. (Recall that a signal cluster, by definition, contains a set of SNPs in LD and represents the same underlying genetic association signal.) Furthermore, we use γ_0 to denote the configuration of no causal eQTLs in the cluster and \mathbf{y}_1 to denote the i^{th} SNP is the true causal eQTL SNP (i.e., the i^{th} entry is set to 1 and 0 for the remaining SNPs).

Assuming GWAS data are originally analyzed using an exchangeable prior $\widetilde{\pi}_1$, i.e.,

$$\Pr(\mathbf{d}_i) = \widetilde{\pi}_1(1 - \widetilde{\pi}_1)^{p-1},$$

and

$$\Pr(\mathbf{d}_0) = (1 - \widetilde{\pi}_1)^p$$

By the nature of a signal cluster, it follows from the Bayes rule that

$$\Pr(\mathbf{d}_i | \mathbf{D}) = \frac{\text{BF}_i}{(1 - \widetilde{\pi}_1)/\widetilde{\pi}_1 + \sum_j \text{BF}_j}, \quad (1)$$

where BF_i denotes the marginal likelihood ratio,

$$\text{BF}_i = \frac{P(\mathbf{D} | \mathbf{d}_i)}{P(\mathbf{D} | \mathbf{d}_0)}$$

Note that in case that the GWAS posterior probability is derived from a multi-SNP analysis, BF_i may not be well-approximated by single SNP testing statistics. Nevertheless, given $\widetilde{\pi}_1$ and note that $\Pr(\mathbf{y}_i | \mathbf{D})$ coincides with the posterior inclusion probability (PIP) of the i th SNP in the signal cluster, BF_i 's can be straightforwardly computed from

Eq. (1). Additionally, $\widetilde{\pi}_1$ can be obtained by averaging the PIPs from all interrogated SNPs.

Given the enrichment information, the GWAS prior differs for eQTL and non-eQTL SNPs. Specifically, for eQTL SNP,

$$\pi_1^e := \Pr(d = 1 \mid \gamma = 1, \hat{\alpha}) = \frac{\exp(\widehat{\alpha}_0 + \widehat{\alpha}_1)}{1 + \exp(\widehat{\alpha}_0 + \widehat{\alpha}_1)},$$

and for non-eQTL SNP,

$$\pi_1^{\bar{e}} := \Pr(d = 1 \mid \gamma = 0, \hat{\alpha}) = \frac{\exp(\widehat{\alpha}_0)}{1 + \exp(\widehat{\alpha}_0)}.$$

Using the eQTL-informed priors, the GWAS posterior probability can be updated analytically, i.e.,

$$\Pr(\mathbf{d}_i \mid \mathbf{D}, \hat{\alpha}, \gamma_i)$$

$$\begin{aligned} &= \frac{\pi_1^e (1 - \pi_1^{\bar{e}})_i^{p-1} \text{BF}_i}{(1 - \pi_1^e)(1 - \pi_1^{\bar{e}})^{p-1} + (1 - \pi_1^e)(1 - \pi_1^{\bar{e}})^{p-2} \pi_1^{\bar{e}} \sum_{j \neq i} \text{BF}_j + \pi_1^e (1 - \pi_1^{\bar{e}})^{p-1} \text{BF}_i} \\ &= \frac{\pi_1^e (1 - \pi_1^{\bar{e}}) \text{BF}_i}{(1 - \pi_1^e)(1 - \pi_1^{\bar{e}}) + (1 - \pi_1^e) \pi_1^{\bar{e}} \sum_{j \neq i} \text{BF}_j + \pi_1^e (1 - \pi_1^{\bar{e}}) \text{BF}_i}. \end{aligned}$$

Subsequently, the colocalization probability at the i th SNP is computed by

$$\Pr(\mathbf{d}_i, \gamma_i \mid \mathbf{D}, \mathbf{E}, \hat{\alpha}) = \Pr(\mathbf{d}_i \mid \mathbf{D}, \hat{\alpha}, \gamma_i) \Pr(\gamma_i \mid \mathbf{E}, \mathbf{D}),$$

where we approximate $\Pr(\gamma_i \mid \mathbf{E}, \mathbf{D})$ with the eQTL PIP for the i th SNP. The regional colocalization probability, RCP, for the signal cluster of interest is given by

$$\text{RCP} = \sum_i \Pr(\mathbf{d}_i, \gamma_i \mid \mathbf{D}, \mathbf{E}, \hat{\alpha}),$$

because events $\{\gamma_i, \mathbf{d}_i\}$ and $\{\gamma_j, \mathbf{d}_j\}$ for $i \neq j$ are mutually exclusive within a signal cluster.

The software and its source code are freely available at <https://github.com/xqwen/fastenloc/>.

REFERENCES AND NOTES

1. P. M. Visscher, N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, J. Yang, 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
2. K. Musunuru, D. Bernstein, F. S. Cole, M. K. Khokha, F. S. Lee, S. Lin, T. V. McDonald, I. P. Moskowitz, T. Quertermous, V. G. Sankaran, D. A. Schwartz, E. K. Silverman, X. Zhou, A. A. K. Hasan, X. J. Luo, Functional assays to screen and dissect genomic hits: Doubling down on the national investment in genomic research. *Circ. Genom. Precis. Med.* **11**, e002178 (2018).
3. A. Buniello, J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A. McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousitou, P. L. Whetzel, R. Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett, L. A. Hindorff, F. Cunningham, H. Parkinson, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
4. T. Stoeger, M. Gerlach, R. I. Morimoto, L. A. Nunes Amaral, Large-scale investigation of the reasons why potentially important genes are ignored. *PLOS Biol.* **16**, e2006643 (2018).
5. W. A. Haynes, A. Tomczak, P. Khatri, Gene annotation bias impedes biomedical research. *Sci. Rep.* **8**, 1362 (2018).
6. D. M. Jordan, M. Verbanck, R. Do, Pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. *bioRxiv* 311332 (2019).
7. O. Canela-Xandri, K. Rawlik, A. Tenesa, An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
8. K. Watanabe, S. Stringer, O. Frei, M. Umićević Mirkov, C. de Leeuw, T. J. C. Polderman, S. van der Sluis, O. A. Andreassen, B. M. Neale, D. Posthuma, A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
9. M. A. Kamat, J. A. Blackshaw, R. Young, P. Surendran, S. Burgess, J. Danesh, A. S. Butterworth, J. R. Staley, PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
10. L. G. Fritsche, S. B. Gruber, Z. Wu, E. M. Schmidt, M. Zawistowski, S. E. Moser, V. M. Blanc, C. M. Brummett, S. Kheterpal, G. R. Abecasis, B. Mukherjee, Association of

polygenic risk scores for multiple cancers in a phenome-wide study: Results from The Michigan Genomics Initiative. *Am. J. Hum. Genet.* **102**, 1048–1061 (2018).

11. Y. Tanigawa, J. Li, J. M. Justesen, H. Horn, M. Aguirre, C. DeBoever, C. Chang, B. Narasimhan, K. Lage, T. Hastie, C. Y. Park, G. Bejerano, E. Ingelsson, M. A. Rivas, Components of genetic associations across 2,138 phenotypes in the UK Biobank highlight adipocyte biology. *Nat. Commun.* **10**, 4064 (2019).
12. J. Zhao, F. Cheng, P. Jia, N. Cox, J. C. Denny, Z. Zhao, An integrative functional genomics framework for effective identification of novel regulatory variants in genome-phenome studies. *Genome Med.* **10**, 7 (2018).
13. A. Brodie, J. R. Azaria, Y. Ofran, How far from the SNP may the causative genes be? *Nucleic Acids Res.* **44**, 6046–6054 (2016).
14. G. Pei, H. Sun, Y. Dai, X. Liu, Z. Zhao, P. Jia, Investigation of multi-trait associations using pathway-based analysis of GWAS summary statistics. *BMC Genomics* **20**, 79 (2019).
15. A. Khosravi, M. Kouhsar, B. Goliaei, B. Jayaram, A. Masoudi-Nejad, Systematic analysis of genes and diseases using PheWAS-Associated networks. *Comput. Biol. Med.* **109**, 311–321 (2019).
16. A. Verma, L. Bang, J. E. Miller, Y. Zhang, M. T. M. Lee, Y. Zhang, M. Byrska-Bishop, D. J. Carey, M. D. Ritchie, S. A. Pendergrass, D. Kim; DiscovEHR Collaboration, Human-disease phenotype map derived from PheWAS across 38,682 individuals. *Am. J. Hum. Genet.* **104**, 55–64 (2019).
17. M. R. Nelson, H. Tipney, J. L. Painter, J. Shen, P. Nicoletti, Y. Shen, A. Floratos, P. C. Sham, M. J. Li, J. Wang, L. R. Cardon, J. C. Whittaker, P. Sanseau, The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
18. E. R. Gamazon, H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny; GTEx Consortium, D. L. Nicolae, N. J. Cox, H. K. Im, A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
19. Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, J. Yang, Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

20. A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B. W. Penninx, R. Jansen, E. J. de Geus, D. I. Boomsma, F. A. Wright, P. F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A. J. Lusis, T. Lehtimäki, E. Raitoharju, M. Kahonen, I. Seppälä, O. T. Raitakari, J. Kuusisto, M. Laakso, A. L. Price, P. Pajukanta, B. Pasaniuc, Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
21. A. François, A. N. Barbeira, R. Bonazzola, A. Brown, S. E. Castel, B. Jo, S. Kasela, S. Kim-Hellmuth, Y. Liang, M. Oliva, P. E. Parsana, E. Flynn, L. Fresard, E. R. Gamazon, A. R. Hamel, Y. He, F. Hormozdiari, P. Mohammadi, M. Muñoz-Aguirre, Y. Park, A. Saha, A. Y. Segré, B. J. Strobe, X. Wen, V. Wucher, S. Das, D. Garrido-Martín, N. R. Gay, R. E. Handsaker, P. J. Hoffman, S. Kashin, A. Kwong, X. Li, D. MacArthur, J. M. Rouhana, M. Stephens, E. Todres, A. Viñuela, G. Wang, Y. Zou; The GTEx Consortium, C. D. Brown, N. Cox, E. Dermitzakis, B. E. Engelhardt, G. Getz, R. Guigo, S. B. Montgomery, B. E. Stranger, H. K. Im, A. Battle, K. G. Ardlie, T. Lappalainen, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv* 787903 (2019).
22. A. N. Barbeira, S. P. Dickinson, R. Bonazzola, J. Zheng, H. E. Wheeler, J. M. Torres, E. S. Torstenson, K. P. Shah, T. Garcia, T. L. Edwards, E. A. Stahl, L. M. Huckins; GTEx Consortium, D. L. Nicolae, N. J. Cox, H. K. Im, Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
23. X. Wen, R. Pique-Regi, F. Luca, Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genet.* **13**, e1006646 (2017).
24. A. N. Barbeira, R. Bonazzola, E. R. Gamazon, Y. Liang, Y. Park, S. Kim-Hellmuth, G. Wang, Z. Jiang, D. Zhou, F. Hormozdiari, B. Liu, A. Rao, A. Hamel, M. Pividori, F. Aguet; GTEx GWAS Working Group, L. Bastarache, D. M. Jordan, M. Verbanck, R. Do; GTEx Consortium, M. Stephens, K. Ardlie, M. McCarthy, S. B. Montgomery, A. Segré, C. D. Brown, T. Lappalainen, X. Wen, H. K. Im, Widespread dose-dependent effects of RNA expression and splicing on complex diseases and traits. *bioRxiv* 814350 (2019).
25. Y. Lee, F. Luca, R. Pique-Regi, X. Wen, Bayesian multi-SNP genetic association analysis: Control of FDR and use of summary statistics. *bioRxiv* 316471 (2018).

26. T. Berisa, J. K. Pickrell, Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).
27. X. Wen, Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann. Appl. Stat.* **10**, 1619–1638 (2016).
28. J. C. Denny, L. Bastarache, M. D. Ritchie, R. J. Carroll, R. Zink, J. D. Mosley, J. R. Field, J. M. Pulley, A. H. Ramirez, E. Bowton, M. A. Basford, D. S. Carrell, P. L. Peissig, A. N. Kho, J. A. Pacheco, L. V. Rasmussen, D. R. Crosslin, P. K. Crane, J. Pathak, S. J. Bielinski, S. A. Pendergrass, H. Xu, L. A. Hindorff, R. Li, T. A. Manolio, C. G. Chute, R. L. Chisholm, E. B. Larson, G. P. Jarvik, M. H. Brilliant, C. A. McCarty, I. J. Kullo, J. L. Haines, D. C. Crawford, D. R. Masys, D. M. Roden, Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1111 (2013).
29. S. Kohler, L. Carmody, N. Vasilevsky, J. O. B. Jacobsen, D. Danis, J. P. Gouridine, M. Gargano, N. L. Harris, N. Matentzoglou, J. A. McMurry, D. Osumi-Sutherland, V. Cipriani, J. P. Balhoff, T. Conlin, H. Blau, G. Baynam, R. Palmer, D. Gratian, H. Dawkins, M. Segal, A. C. Jansen, A. Muaz, W. H. Chang, J. Bergerson, S. J. F. Laulederkind, Z. Yuksel, S. Beltran, A. F. Freeman, P. I. Sergouniotis, D. Durkin, A. L. Storm, M. Hanauer, M. Brudno, S. M. Bello, M. Sincan, K. Rageth, M. T. Wheeler, R. Oegema, H. Loughi, M. G. Della Rocca, R. Thompson, F. Castellanos, J. Priest, C. Cunningham-Rundles, A. Hegde, R. C. Lovering, C. Hajek, A. Olry, L. Notarangelo, M. Similuk, X. A. Zhang, D. Gomez-Andres, H. Lochmuller, H. Dollfus, S. Rosenzweig, S. Marwaha, A. Rath, K. Sullivan, C. Smith, J. D. Milner, D. Leroux, C. F. Boerkoel, A. Klion, M. C. Carter, T. Groza, D. Smedley, M. A. Haendel, C. Mungall, P. N. Robinson, Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* **47**, D1018–D1027 (2019).
30. J. S. Amberger, C. A. Bocchini, A. F. Scott, A. Hamosh, OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2019).
31. Type 2 Diabetes Knowledge Portal. Accelerating Medicine Partnership: T2D; type2diabetesgenetics.org.
32. Y. Hu, A. Shmygelska, D. Tran, N. Eriksson, J. Y. Tung, D. A. Hinds, GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7**, 10448 (2016).

33. J. R. Jones, T. Simon, L. Lones, E. D. Herzog, SCN VIP neurons are essential for normal light-mediated resetting of the circadian system. *J. Neurosci.* **38**, 7986–7995 (2018).
34. J. L. Bedont, T. A. LeGates, E. Buhr, A. Bathini, J. P. Ling, B. Bell, M. N. Wu, P. C. Wong, R. N. Van Gelder, V. Mongrain, S. Hattar, S. Blackshaw, An LHX1-regulated transcriptional network controls sleep/wake coupling and thermal resistance of the central circadian clockworks. *Curr. Biol.* **27**, 128–136 (2017).
35. E. Kirveskari, M. Partinen, P. Santavuori, Sleep and its disturbance in a variant form of late infantile neuronal ceroid lipofuscinosis (CLN5). *J. Child Neurol.* **16**, 707–713 (2001).
36. S. M. Siepka, S. H. Yoo, J. Park, W. Song, V. Kumar, Y. Hu, C. Lee, J. S. Takahashi, Circadian mutant overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression. *Cell* **129**, 1011–1023 (2007).
37. S. N. Archer, C. Schmidt, G. Vandewalle, D.-J. Dijk, Phenotyping of PER3 variants reveals widespread effects on circadian preference, sleep regulation, and health. *Sleep Med. Rev.* **40**, 109–126 (2018).
38. A. Jagannath, L. Taylor, Z. Wakaf, S. R. Vasudevan, R. G. Foster, The genetics of circadian rhythms, sleep and health. *Hum. Mol. Genet.* **26**, R128–R138 (2017).
39. J. Ruf, P. Carayon, Structural and functional aspects of thyroid peroxidase. *Arch. Biochem. Biophys.* **445**, 269–277 (2006).
40. S. S. Birring, A. J. Morgan, B. Prudon, T. M. McKeever, S. A. Lewis, J. F. Falconer Smith, R. J. Robinson, J. R. Britton, I. D. Pavord, Respiratory symptoms in patients with treated hypothyroidism and inflammatory bowel disease. *Thorax* **58**, 533–536 (2003).
41. L. Y. Geer, A. Marchler-Bauer, R. C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, S. H. Bryant, The NCBI BioSystems database. *Nucleic Acids Res.* **38**, D492–D496 (2010).
42. M. A. Beydoun, H. A. Beydoun, O. S. Rostant, G. A. Dore, M. T. Fanelli-Kuczmarski, M. K. Evans, A. B. Zonderman, Thyroid hormones are associated with longitudinal cognitive change in an urban adult population. *Neurobiol. Aging* **36**, 3056–3066 (2015).
43. M. Barbuti, A. F. Carvalho, C. A. Kohler, A. Murru, N. Verdolini, G. Guiso, L. Samalin, M. Maes, B. Stubbs, G. Perugi, E. Vieta, I. Pacchiarotti, Thyroid autoimmunity in bipolar disorder: A systematic review. *J. Affect. Disord.* **221**, 97–106 (2017).
44. E. Berta, I. Lengyel, S. Halmi, M. Zrinyi, A. Erdei, M. Harangi, D. Pall, E. V. Nagy, M. Bodor, Hypertension in thyroid disorders. *Front. Endocrinol.* **10**, 482 (2019).

45. The Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
46. D. J. Stumpo, H. E. Broxmeyer, T. Ward, S. Cooper, G. Hangoc, Y. J. Chung, W. C. Shelley, E. K. Richfield, M. K. Ray, M. C. Yoder, P. D. Aplan, P. J. Blackshear, Targeted disruption of *Zfp36l2*, encoding a CCCH tandem zinc finger RNA-binding protein, results in defective hematopoiesis. *Blood* **114**, 2401–2410 (2009).
47. E. Iio, K. Matsuura, N. Nishida, S. Maekawa, N. Enomoto, M. Nakagawa, N. Sakamoto, H. Yatsushashi, M. Kurosaki, N. Izumi, Y. Hiasa, N. Masaki, T. Ide, K. Hino, A. Tamori, M. Honda, S. Kaneko, S. Mochida, H. Nomura, S. Nishiguchi, C. Okuse, Y. Itoh, H. Yoshiji, I. Sakaida, K. Yamamoto, H. Watanabe, S. Hige, A. Matsumoto, E. Tanaka, K. Tokunaga, Y. Tanaka, Genome-wide association study identifies a PSMD3 variant associated with neutropenia in interferon-based therapy for chronic hepatitis C. *Hum. Genet.* **134**, 279–289 (2015).
48. M. J. Landrum, J. M. Lee, M. Benson, G. R. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Karapetyan, K. Katz, C. Liu, Z. Maddipatla, A. Malheiro, K. McDaniel, M. Ovetsky, G. Riley, G. Zhou, J. B. Holmes, B. L. Kattman, D. R. Maglott, ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
49. Z. D. Zhou, J. C. T. Lee, E. K. Tan, Pathophysiological mechanisms linking F-box only protein 7 (FBXO7) and Parkinson's disease (PD). *Mutat. Res.* **778**, 72–78 (2018).
50. N. Soranzo, T. D. Spector, M. Mangino, B. Kuhnel, A. Rendon, A. Teumer, C. Willenborg, B. Wright, L. Chen, M. Li, P. Salo, B. F. Voight, P. Burns, R. A. Laskowski, Y. Xue, S. Menzel, D. Altshuler, J. R. Bradley, S. Bumpstead, M. S. Burnett, J. Devaney, A. Doring, R. Elosua, S. E. Epstein, W. Erber, M. Falchi, S. F. Garner, M. J. Ghorri, A. H. Goodall, R. Gwilliam, H. H. Hakonarson, A. S. Hall, N. Hammond, C. Hengstenberg, T. Illig, I. R. Konig, C. W. Knouff, R. McPherson, O. Melander, V. Mooser, M. Nauck, M. S. Nieminen, C. J. O'Donnell, L. Peltonen, S. C. Potter, H. Prokisch, D. J. Rader, C. M. Rice, R. Roberts, V. Salomaa, J. Sambrook, S. Schreiber, H. Schunkert, S. M. Schwartz, J. Serbanovic-Canic, J. Sinisalo, D. S. Siscovick, K. Stark, I. Surakka, J. Stephens, J. R. Thompson, U. Volker, H. Volzke, N. A. Watkins, G. A. Wells, H. E. Wichmann, D. A. Van Heel, C. Tyler-Smith, S. L. Thein, S. Kathiresan, M. Perola, M. P. Reilly, A. F. Stewart, J. Erdmann, N. J. Samani, C.

Meisinger, A. Greinacher, P. Deloukas, W. H. Ouwehand, C. Gieger, A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat. Genet.* **41**, 1182–1190 (2009).

51. K. Ding, K. Shameer, H. Jouni, D. R. Masys, G. P. Jarvik, A. N. Kho, M. D. Ritchie, C. A. McCarty, C. G. Chute, T. A. Manolio, I. J. Kullo, Genetic Loci implicated in erythroid differentiation and cell cycle regulation are associated with red blood cell traits. *Mayo Clin. Proc.* **87**, 461–474 (2012).
52. P. van der Harst, W. Zhang, I. Mateo Leach, A. Rendon, N. Verweij, J. Sehmi, D. S. Paul, U. Elling, H. Allayee, X. Li, A. Radhakrishnan, S. T. Tan, K. Voss, C. X. Weichenberger, C. A. Albers, A. Al-Hussani, F. W. Asselbergs, M. Ciullo, F. Danjou, C. Dina, T. Esko, D. M. Evans, L. Franke, M. Gogele, J. Hartiala, M. Hersch, H. Holm, J. J. Hottenga, S. Kanoni, M. E. Kleber, V. Lagou, C. Langenberg, L. M. Lopez, L. P. Lyytikainen, O. Melander, F. Murgia, I. M. Nolte, P. F. O'Reilly, S. Padmanabhan, A. Parsa, N. Pirastu, E. Porcu, L. Portas, I. Prokopenko, J. S. Ried, S. Y. Shin, C. S. Tang, A. Teumer, M. Traglia, S. Ulivi, H. J. Westra, J. Yang, J. H. Zhao, F. Anni, A. Abdellaoui, A. Attwood, B. Balkau, S. Bandinelli, F. Bastardot, B. Benyamin, B. O. Boehm, W. O. Cookson, D. Das, P. I. de Bakker, R. A. de Boer, E. J. de Geus, M. H. de Moor, M. Dimitriou, F. S. Domingues, A. Doring, G. Engstrom, G. I. Eyjolfsson, L. Ferrucci, K. Fischer, R. Galanelli, S. F. Garner, B. Genser, Q. D. Gibson, G. Girotto, D. F. Gudbjartsson, S. E. Harris, A. L. Hartikainen, C. E. Hastie, B. Hedblad, T. Illig, J. Jolley, M. Kahonen, I. P. Kema, J. P. Kemp, L. Liang, H. Lloyd-Jones, R. J. Loos, S. Meacham, S. E. Medland, C. Meisinger, Y. Memari, E. Mihailov, K. Miller, M. F. Moffatt, M. Nauck, M. Novatchkova, T. Nutile, I. Olafsson, P. T. Onundarson, D. Parracciani, B. W. Penninx, L. Perseu, A. Piga, G. Pistis, A. Pouta, U. Puc, O. Raitakari, S. M. Ring, A. Robino, D. Ruggiero, A. Ruukonen, A. Saint-Pierre, C. Sala, A. Salumets, J. Sambrook, H. Schepers, C. O. Schmidt, H. H. Sillje, R. Sladek, J. H. Smit, J. M. Starr, J. Stephens, P. Sulem, T. Tanaka, U. Thorsteinsdottir, V. Tragante, W. H. van Gilst, L. J. van Pelt, D. J. van Veldhuisen, U. Volker, J. B. Whitfield, G. Willemsen, B. R. Winkelmann, G. Wirnsberger, A. Algra, F. Cucca, A. P. d'Adamo, J. Danesh, I. J. Deary, A. F. Dominiczak, P. Elliott, P. Fortina, P. Froguel, P. Gasparini, A. Greinacher, S. L. Hazen, M. R. Jarvelin, K. T. Khaw, T. Lehtimäki, W. Maerz, N. G. Martin, A. Metspalu, B. D. Mitchell, G. W. Montgomery, C. Moore, G. Navis, M. Pirastu, P. P. Pramstaller, R. Ramirez-Solis, E. Schadt,

- J. Scott, A. R. Shuldiner, G. D. Smith, J. G. Smith, H. Snieder, R. Sorice, T. D. Spector, K. Stefansson, M. Stumvoll, W. H. Tang, D. Toniolo, A. Tonjes, P. M. Visscher, P. Vollenweider, N. J. Wareham, B. H. Wolffenbuttel, D. I. Boomsma, J. S. Beckmann, G. V. Dedoussis, P. Deloukas, M. A. Ferreira, S. Sanna, M. Uda, A. A. Hicks, J. M. Penninger, C. Gieger, J. S. Kooner, W. H. Ouwehand, N. Soranzo, J. C. Chambers, Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369–375 (2012).
53. S. J. Randle, D. E. Nelson, S. P. Patel, H. Laman, Defective erythropoiesis in a mouse model of reduced Fbxo7 expression due to decreased p27 expression. *J. Pathol.* **237**, 263–272 (2015).
54. C. Ballesteros Reviriego, S. Clare, M. J. Arends, E. L. Cambridge, A. Swiatkowska, S. Caetano, B. Abu-Helil, L. Kane, K. Harcourt, D. A. Goulding, D. Gleeson, E. Ryder, B. Doe, J. K. White, L. van der Weyden, G. Dougan, D. J. Adams, A. O. Speak, FBXO7 sensitivity of phenotypic traits elucidated by a hypomorphic allele. *PLOS ONE* **14**, e0212481 (2019).
55. M. D. Shapiro, H. Tavori, S. Fazio, PCSK9: From basic science discoveries to clinical trials. *Circ. Res.* **122**, 1420–1438 (2018).
56. A. F. Schmidt, D. I. Swerdlow, M. V. Holmes, R. S. Patel, Z. Fairhurst-Hunter, D. M. Lyall, F. P. Hartwig, B. L. Horta, E. Hypponen, C. Power, M. Moldovan, E. van Iperen, G. K. Hovingh, I. Demuth, K. Norman, E. Steinhagen-Thiessen, J. Demuth, L. Bertram, T. Liu, S. Coassin, J. Willeit, S. Kiechl, K. Willeit, D. Mason, J. Wright, R. Morris, G. Wanamethee, P. Whincup, Y. Ben-Shlomo, S. McLachlan, J. F. Price, M. Kivimaki, C. Welch, A. Sanchez-Galvez, P. Marques-Vidal, A. Nicolaides, A. G. Panayiotou, N. C. Onland-Moret, Y. T. van der Schouw, G. Matullo, G. Fiorito, S. Guarrera, C. Sacerdote, N. J. Wareham, C. Langenberg, R. Scott, J. Luan, M. Bobak, S. Malyutina, A. Pajak, R. Kubinova, A. Tamosiunas, H. Pikhart, L. L. Husemoen, N. Grarup, O. Pedersen, T. Hansen, A. Linneberg, K. S. Simonsen, J. Cooper, S. E. Humphries, M. Brilliant, T. Kitchner, H. Hakonarson, D. S. Carrell, C. A. McCarty, H. L. Kirchner, E. B. Larson, D. R. Crosslin, M. de Andrade, D. M. Roden, J. C. Denny, C. Carty, S. Hancock, J. Attia, E. Holliday, M. O'Donnell, S. Yusuf, M. Chong, G. Pare, P. van der Harst, M. A. Said, R. N. Eppinga, N. Verweij, H. Snieder; LifeLines Cohort study group, T. Christen, D. O. Mook-Kanamori, S. Gustafsson, L. Lind, E. Ingelsson, R. Pazoki, O. Franco, A. Hofman, A. Uitterlinden, A. Dehghan, A. Teumer, S. Baumeister, M. Dorr, M. M. Lerch, U. Volker, H. Volzke, J. Ward, J. P. Pell, D. J. Smith, T.

- Meade, A. H. Maitland-van der Zee, E. V. Baranova, R. Young, I. Ford, A. Campbell, S. Padmanabhan, M. L. Bots, D. E. Grobbee, P. Froguel, D. Thuillier, B. Balkau, A. Bonnefond, B. Cariou, M. Smart, Y. Bao, M. Kumari, A. Mahajan, P. M. Ridker, D. I. Chasman, A. P. Reiner, L. A. Lange, M. D. Ritchie, F. W. Asselbergs, J. P. Casas, B. J. Keating, D. Preiss, A. D. Hingorani; UCLEB consortium, N. Sattar, PCSK9 genetic variants and risk of type 2 diabetes: A mendelian randomisation study. *Lancet Diabetes Endocrinol.* **5**, 97–105 (2017).
57. M. S. Sabatine, L. A. Leiter, S. D. Wiviott, R. P. Giugliano, P. Deedwania, G. M. De Ferrari, S. A. Murphy, J. F. Kuder, I. Gouni-Berthold, B. S. Lewis, Y. Handelsman, A. L. Pineda, N. Honarpour, A. C. Keech, P. S. Sever, T. R. Pedersen, Cardiovascular safety and efficacy of the PCSK9 inhibitor evolocumab in patients with and without diabetes and the effect of evolocumab on glycaemia and risk of new-onset diabetes: A prespecified analysis of the FOURIER randomised controlled trial. *Lancet Diabetes Endocrinol.* **5**, 941–950 (2017).
58. C. S. Paththinige, N. D. Sirisena, V. Dissanayake, Genetic determinants of inherited susceptibility to hypercholesterolemia - A comprehensive literature review. *Lipids Health Dis.* **16**, 103 (2017).
59. S. M. Uebachs, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).
60. E. Zeggini, A. L. Gloyn, A. C. Barton, L. V. Wain, Translational genomics and precision medicine: Moving from the lab to the clinic. *Science* **365**, 1409–1413 (2019).
61. B. M. Neale, Neale Lab - UK Biobank GWAS Results (2018); www.nealelab.is/uk-biobank/.
62. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O’Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
63. L. A. C. Millard, N. M. Davies, T. R. Gaunt, G. Davey Smith, K. Tilling, Software application profile: PHESANT: A tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* **47**, 29–35 (2017).

64. F. Hormozdiari, M. van de Bunt, A. V. Segre, X. Li, J. W. J. Joo, M. Bilow, J. H. Sul, S. Sankararaman, B. Pasaniuc, E. Eskin, Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
65. C. Wallace, Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLOS Genet.* **16**, e1008720 (2019).
66. Global Lipids Genetics Consortium, Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).