

SUPPLEMENTAL MATERIALS

Best practices of multi-ancestry, meta-analytic transcriptome-wide associations: lessons from the Global Biobank Meta-Initiative

SUPPLEMENTAL FIGURES

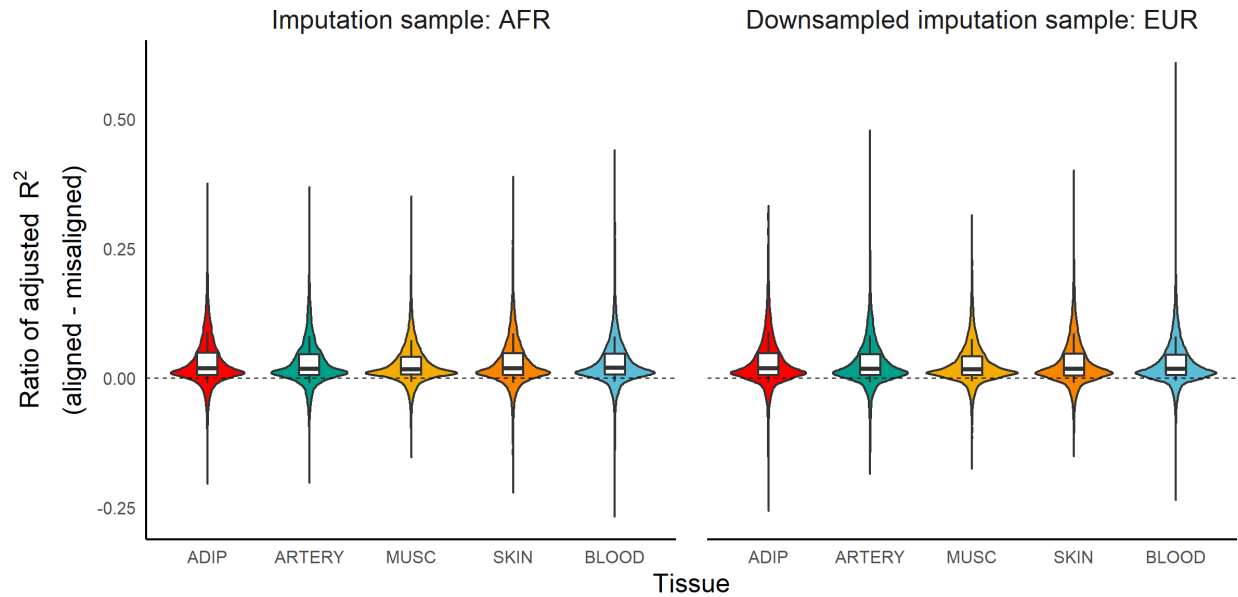


Figure S1: Ratio of predictive performance of expression models in aligned versus misaligned imputation samples across AFR (left) and EUR (right) ancestry in the imputation sample. Here, we down-sample the EUR imputation sample to match the sample size of the AFR imputation sample.

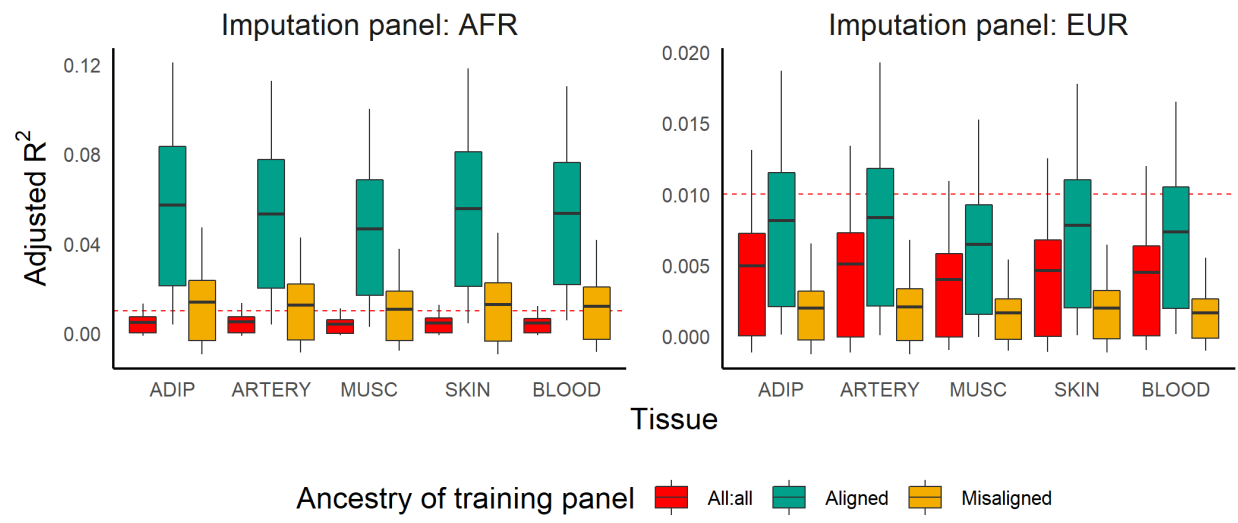


Figure S2: Predictive performance of expression models in aligned and misaligned imputation samples.

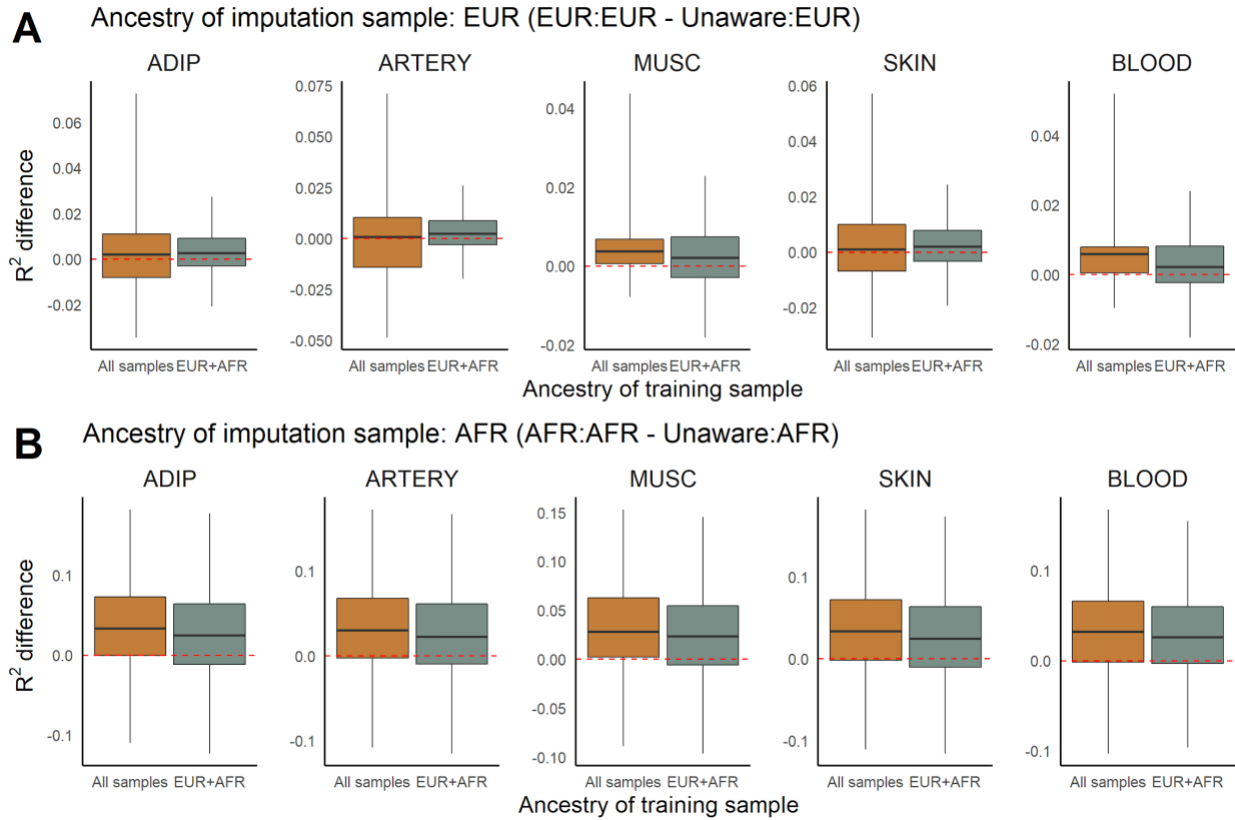


Figure S3: Predictive performance of ancestry-unaware expression models compared to ancestry-specific models across 5 tissues. Boxplot of difference in predictive performance in EUR (A) and AFR (B) samples between ancestry-aligned models and ancestry-unaware models. We consider (1) individuals of all ancestry in the training sample of the ancestry-unaware model (gold) or only EUR and AFR individuals in the training sample (grey). The red line indicates a difference of 0.

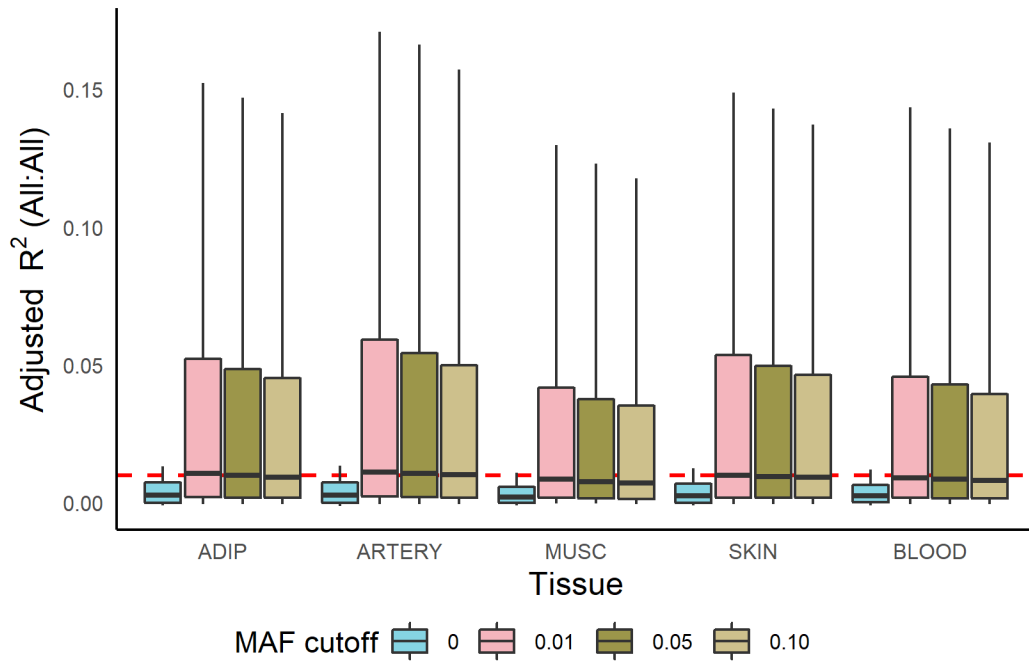


Figure S4: Predictive performance of ancestry-unaware expression models across minor allele frequency thresholds.

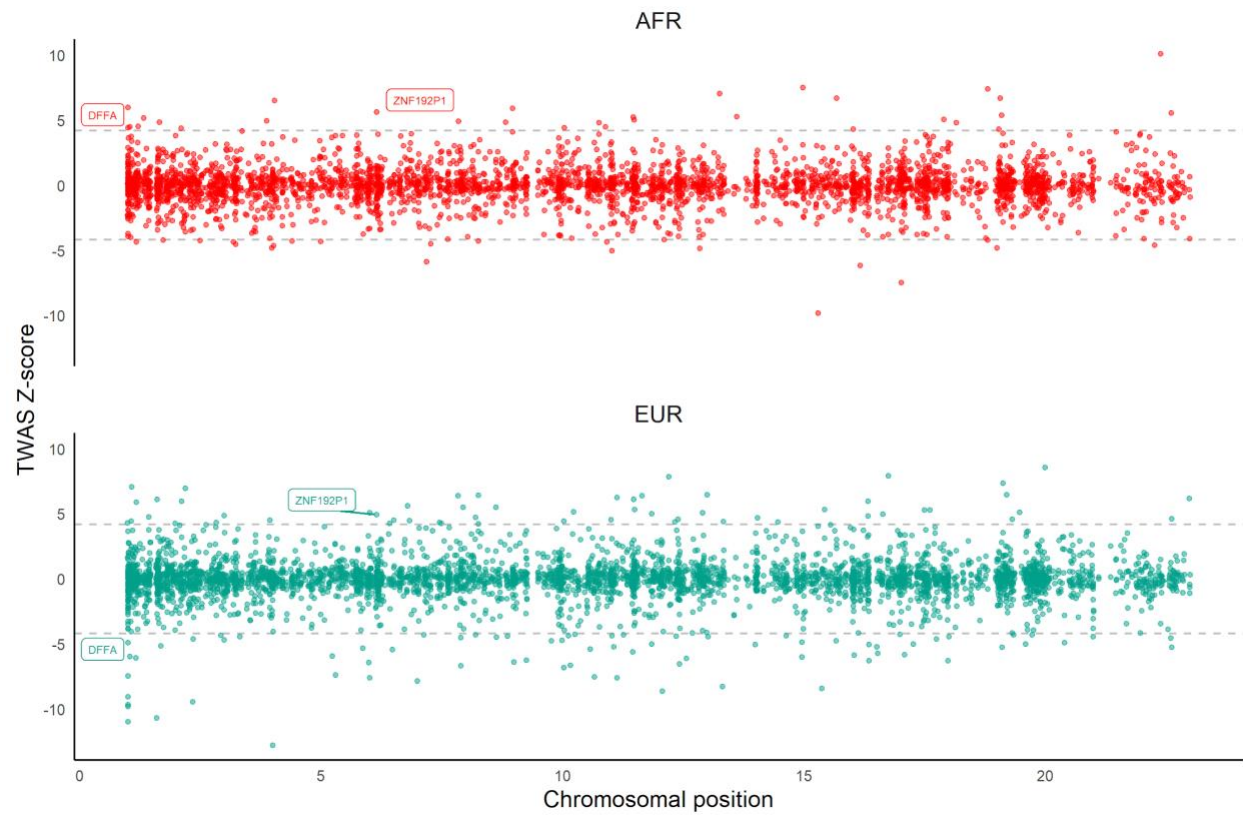


Figure S5: TWAS Miami plots across AFR and EUR ancestry groups for asthma using whole blood gene expression models.

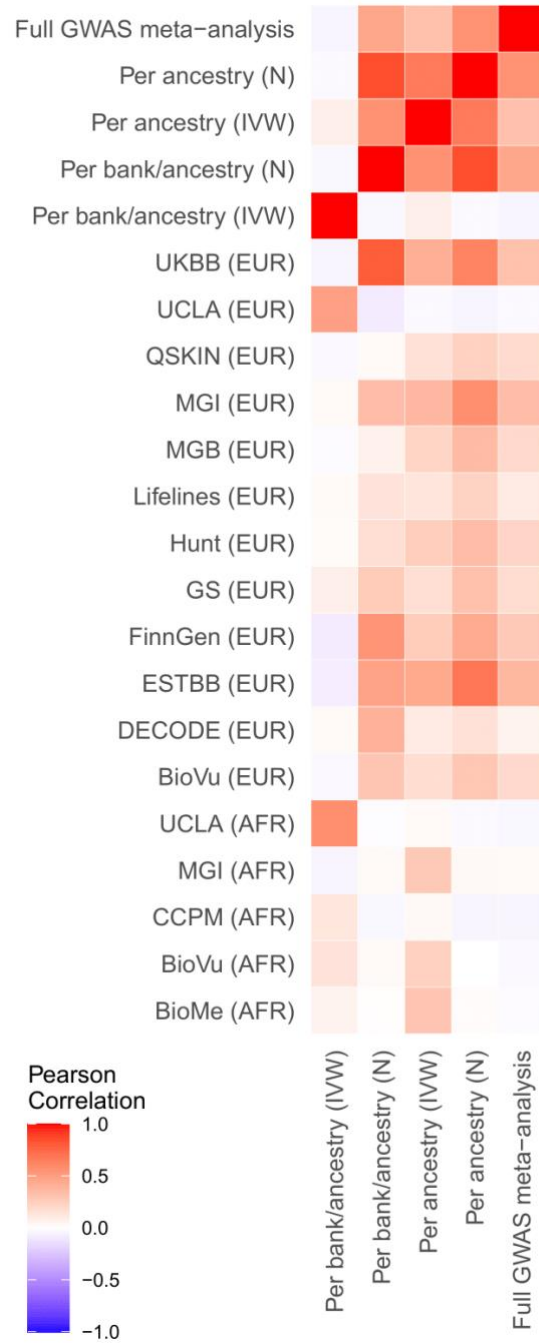


Figure S6: Correlation of TWAS Z-scores across ancestry-specific, individual biobank GWAS cohorts and 5 meta-analytic strategies.

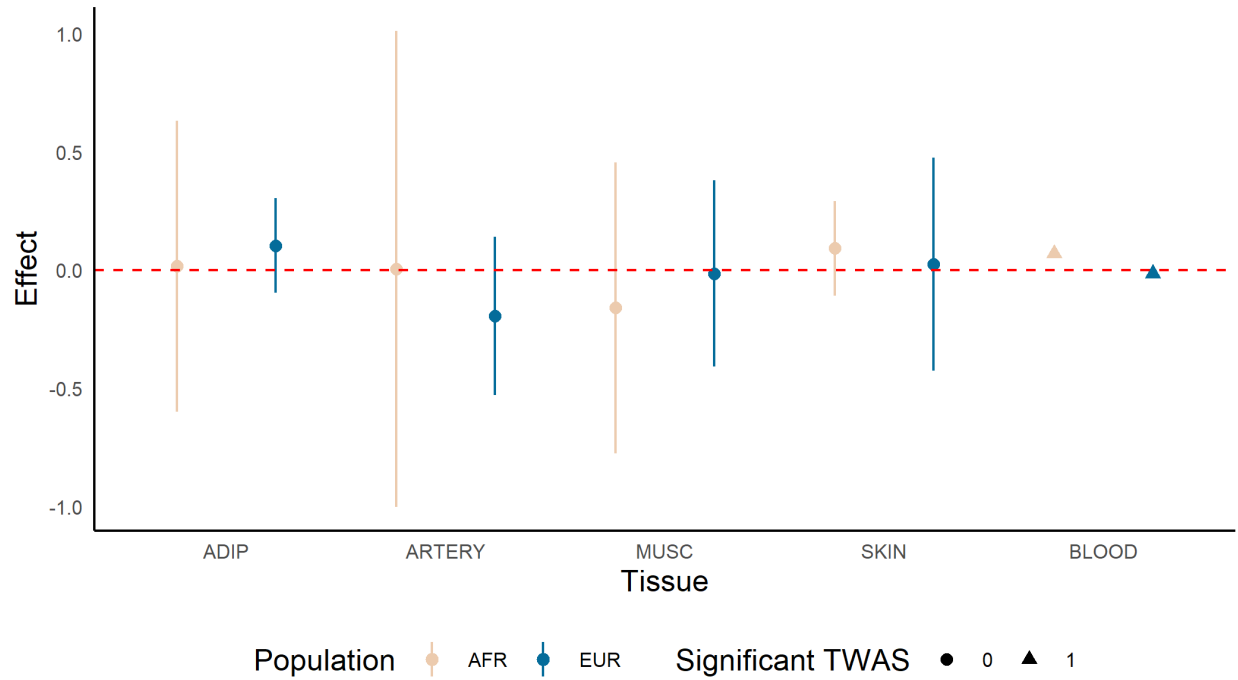


Figure S7: TWAS associations across EUR and AFR ancestry groups for DFFA across 5 tissues. The effect size is given with the point (triangle if association is transcriptome-wide significant) with a 95% confidence interval provided.

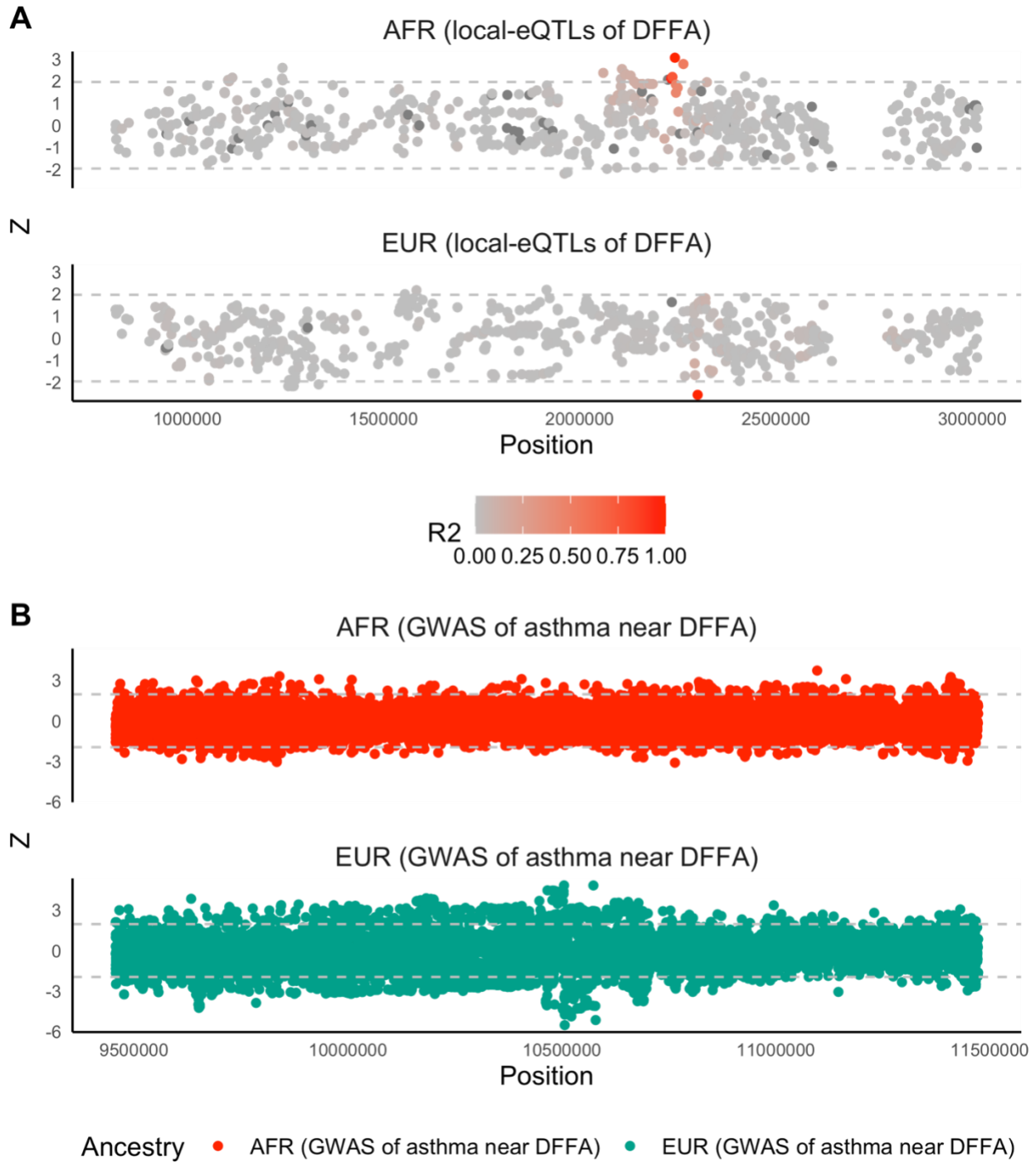


Figure S8: Miami plots of DFFA local-eQTLs and GWAS signal for SNPs around DFFA. In (A), color shows linkage disequilibrium R^2 to lead eQTL SNP. Grey line shows a nominal P-value cutoff of 0.05 ($|Z| = 1.96$).

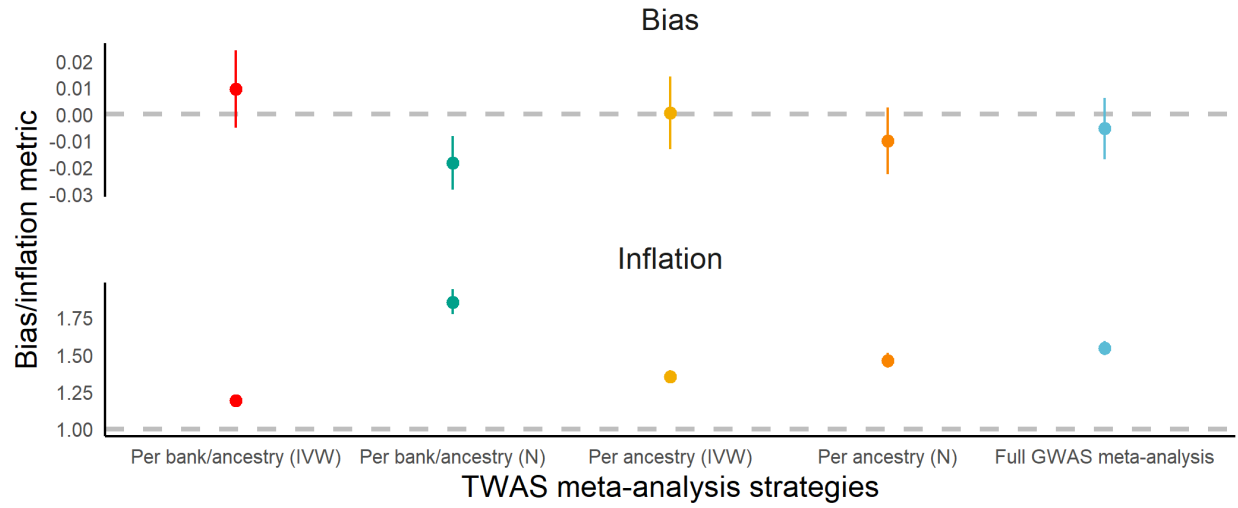


Figure S9: Empirical Bayes estimates of bias and inflation in TWAS Z-scores across meta-analysis strategies. Estimates of bias (top) and bottom (inflation) with one standard error width around the estimate are given across meta-analysis strategies. The dotted lines provide a reference for the null (0 for bias and 1 for inflation).

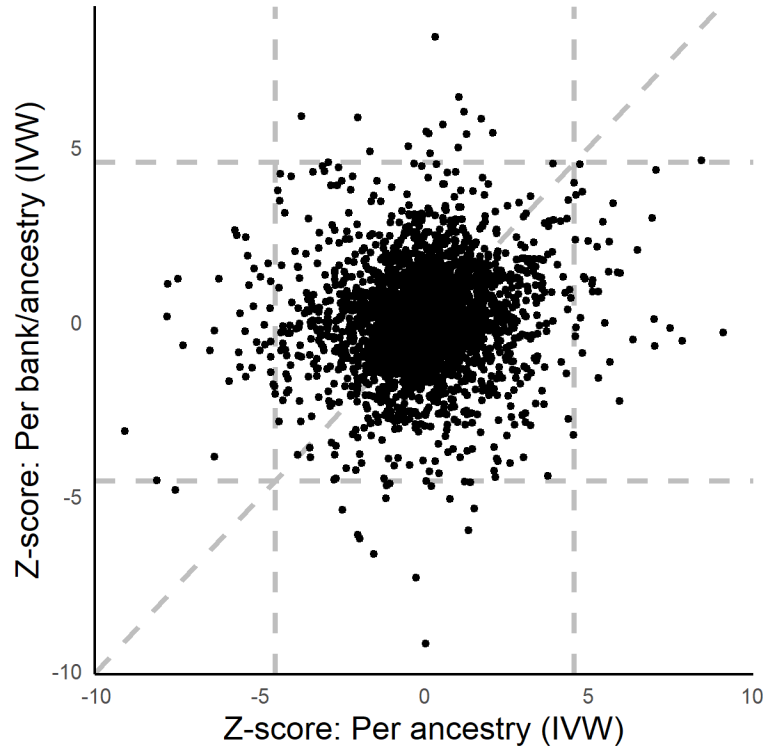


Figure S10: Comparison of two IVW meta-analyzed Z-scores. Vertical and horizontal dotted lines give a reference for the Bonferroni-corrected threshold for transcriptome-significance. A diagonal line is provided from reference.

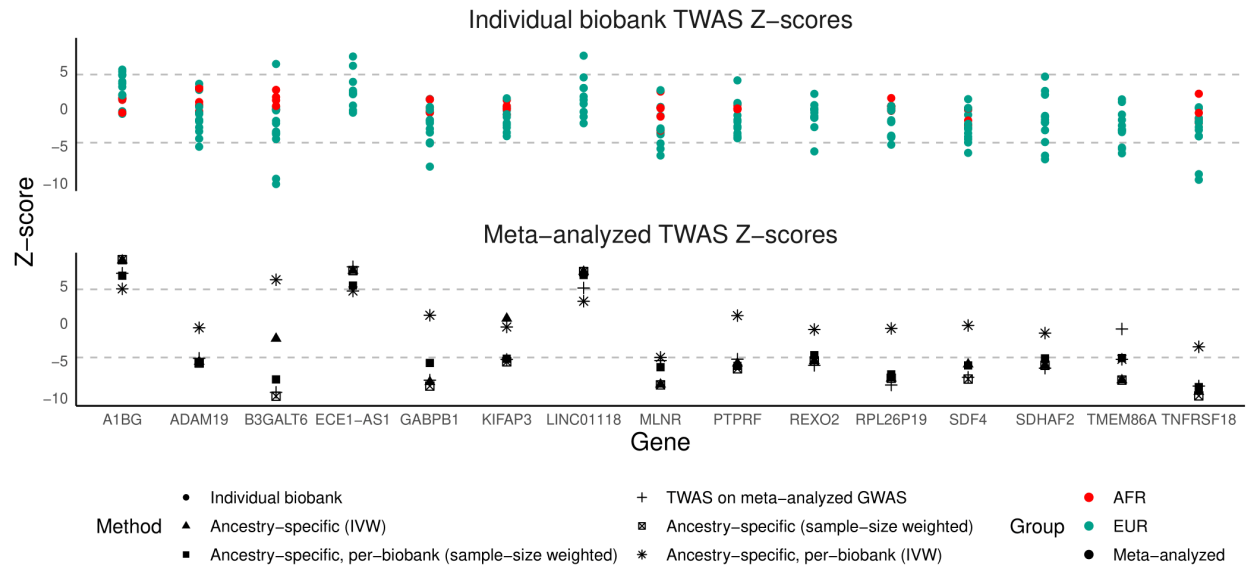


Figure S11: Comparison of meta-analyzed Z-scores with individual biobank TWAS Z-scores. Ancestry-specific TWAS Z-scores for individual biobanks are shown in the top panel, colored by ancestry. Meta-analyzed Z-scores are shown in the bottom panel with shapes reflecting the different strategies. Dotted lines provide a reference for transcriptome-wide significance.

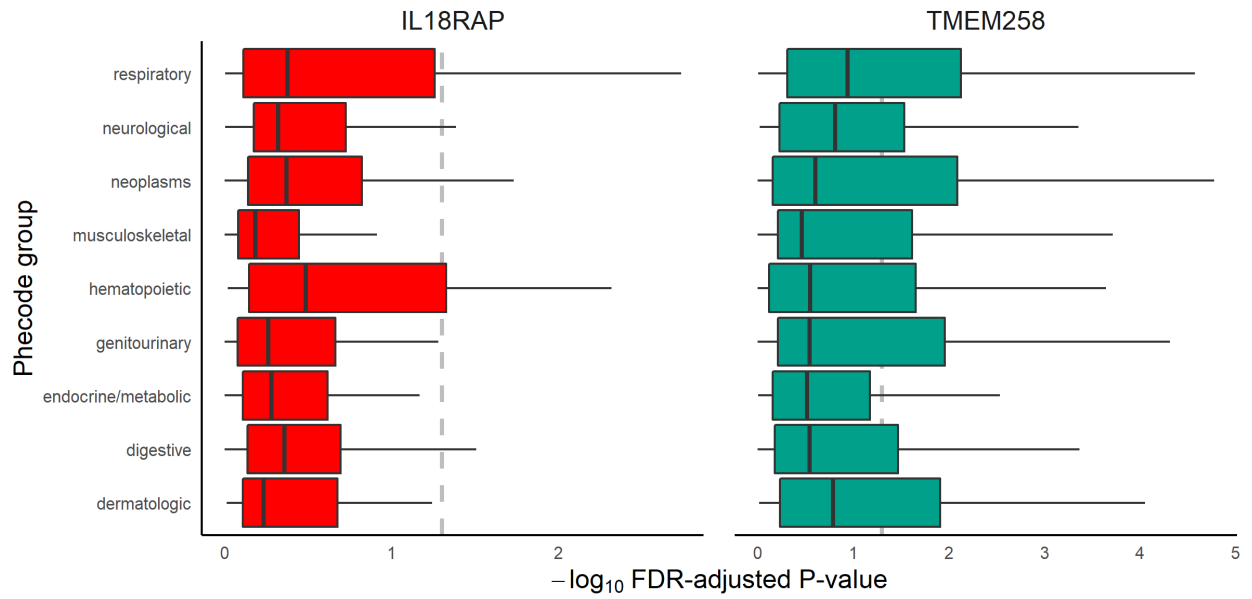


Figure S12: UKBB T-PheWAS associations across 5 representative asthma-associated genes through European-only meta-analytic TWAS, grouped by phecode group. The horizontal dotted line shows FDR-adjusted $P = 0.05$.

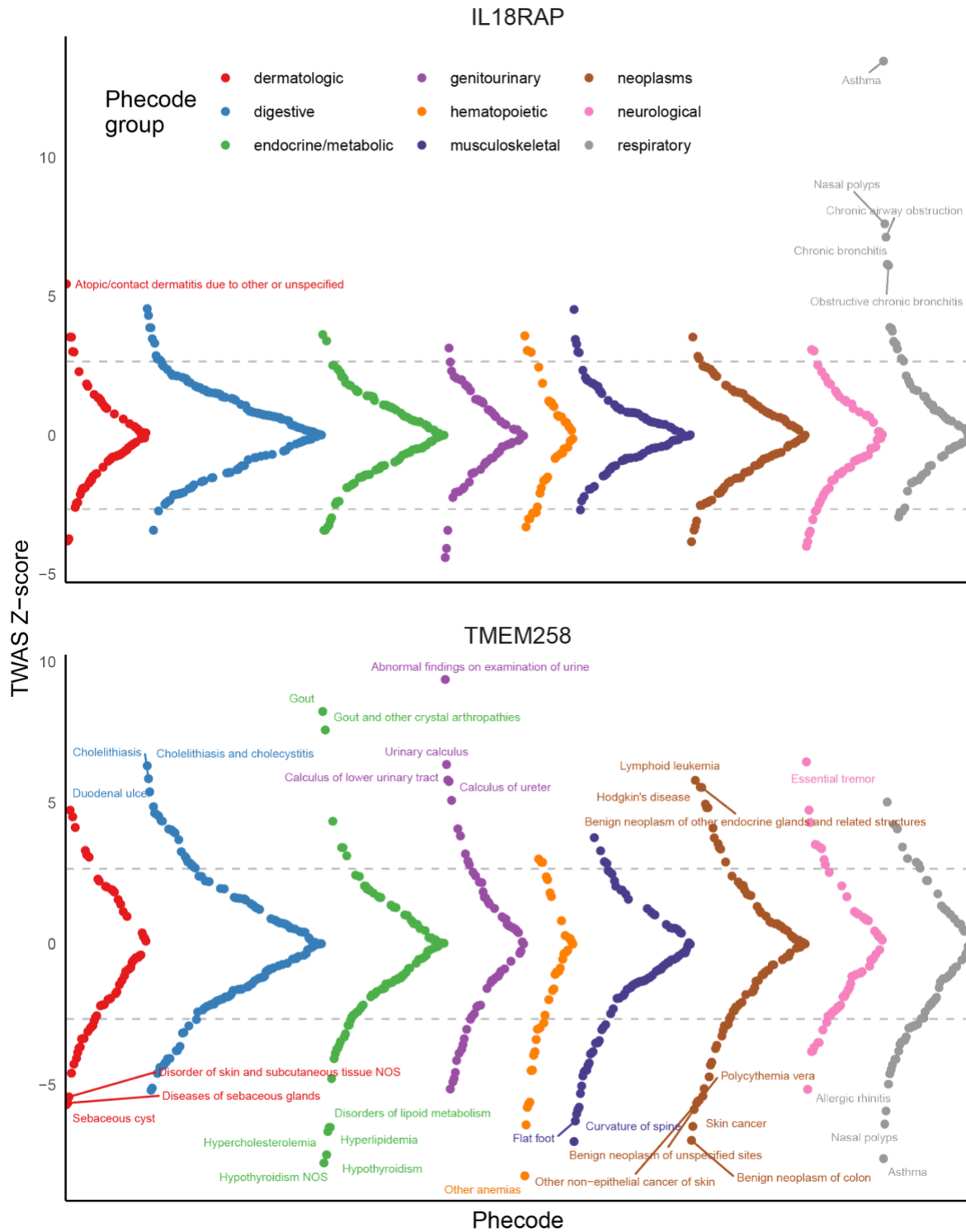


Figure S13: Miami plots of UKBB T-PheWAS associations across 2 genes previously implicated through GWAS and detected in European-only meta-analytic TWAS in GBMI.

SUPPLEMENTAL INFORMATION

The following represents the full author list under the Global Biobank Meta-analysis Initiative (GBMI) banner:

Wei Zhou^{1,2,3}, Masahiro Kanai^{1,2,3,4,5}, Kuan-Han H Wu⁶, Humaira Rasheed^{7,8,9}, Kristin Tsuo^{1,2,3}, Jibril B Hirbo^{10,11}, Ying Wang^{1,2,3}, Arjun Bhattacharya¹², Huiling Zhao⁹, Shinichi Namba⁵, Ida Surakka¹³, Brooke N Wolford^{6,7}, Valeria Lo Faro^{14,15,16}, Esteban A Lopera-Maya¹⁷, Kristi Läll¹⁸, Marie-Julie Favé¹⁹, Sinéad B Chapman^{2,3}, Juha Karjalainen^{1,2,3,20}, Mitja Kurki^{1,2,3,20}, Maasha Mutaamba^{1,2,3,20}, Juulia Partanen²⁰, Ben M Brumpton^{7,21,22}, Sameer Chavan²³, Tzu-Ting Chen²⁴, Michelle Daya²³, Yi Ding^{12,25}, Yen-Chen A Feng^{26,27}, Christopher R Gignoux²³, Sarah E Graham¹³, Whitney E Hornsby¹³, Nathan Ingold^{28,29}, Ruth Johnson^{12,30}, Triin Laisk¹⁸, Kuang Lin³¹, Jun Lv³², Iona Y Millwood^{31,33}, Priit Palta^{18,20}, Anita Pandit³⁴, Michael H Preuss³⁵, Unnur Thorsteinsdottir³⁶, Jasmina Uzunovic¹⁹, Matthew Zawistowski³⁴, Xue Zhong^{10,11}, Archie Campbell³⁷, Kristy Crooks²³, Geertruida H de Bock³⁸, Nicholas J Douville^{39,40}, Sarah Finer⁴¹, Lars G Fritsche³⁴, Christopher J Griffiths⁴¹, Yu Guo⁴², Karen A Hunt⁴³, Takahiro Konuma^{5,44}, Riccardo E Marioni³⁷, Jansonius Nomdo¹⁴, Snehal Patil³⁴, Nicholas Rafaels²³, Anne Richmond⁴⁵, Jonathan A Shortt²³, Peter Straub^{10,11}, Ran Tao¹¹, Brett Vanderwerff³⁴, Kathleen C Barnes²³, Marike Boezen³⁸, Zhengming Chen^{31,33}, Chia-Yen Chen⁴⁶, Judy Cho³⁵, George Davey Smith^{9,47}, Hilary K Finucane^{1,2,3}, Lude Franke¹⁷, Eric R Gamazon^{10,11,48}, Andrea Ganna^{1,2,20}, Tom R Gaunt⁹, Tian Ge^{27,49}, Hailiang Huang^{1,2}, Jennifer Huffman⁵⁰, Jukka T. Koskela²⁰, Clara Lajonchere^{51,52}, Matthew H Law^{28,29}, Liming Li³², Cecilia M Lindgren⁵³, Ruth JF Loos^{35,54}, Stuart MacGregor²⁸, Koichi Matsuda⁵⁵, Catherine M Olsen²⁸, David J Porteous³⁷, Jordan A Shavit⁵⁶, Harold Snieder³⁸, Richard C Trembath⁵⁷, Judith M Vonk³⁸, David Whiteman²⁸, Stephen J Wicks²³, Cisca Wijmenga¹⁷, John Wright⁵⁸, Jie Zheng⁹, Xiang Zhou³⁴, Philip Awadalla^{19,59}, Michael Boehnke³⁴, Nancy J Cox^{10,11}, Daniel H Geschwind^{51,60,61}, Caroline Hayward⁴⁵, Kristian Hveem^{7,21}, Eimear E Kenny⁶², Yen-Feng Lin^{24,63,64}, Reedik Mägi¹⁸, Hilary C Martin⁶⁵, Sarah E Medland²⁸, Yukinori Okada^{5,66,67,68,69}, Aarno V Palotie^{1,2,20}, Bogdan Pasaniuc^{12,25,51,60,70}, Serena Sanna^{17,71}, Jordan W Smoller²⁷, Kari Stefansson³⁶, David A van Heel⁴³, Robin G Walters^{31,33}, Sebastian Zöllner³⁴, Biobank Japan, BioMe, BioVU, Canadian Partnership for Tomorrow's Health/Ontario Health Study, China Kadoorie Biobank Collaborative Group, Colorado Center for Personalized Medicine, deCODE Genetics, Estonian Biobank, FinnGen, Generation Scotland, Genes & Health, LifeLines, Mass General Brigham Biobank, Michigan Genomics Initiative, QIMR Berghofer Biobank, Taiwan Biobank, The HUNT Study, UCLA ATLAS Community Health Initiative, UK Biobank, Alicia R Martin^{1,2,3}, Cristen J Willer^{6,13,72*}, Mark J Daly^{1,2,3,20*}, Benjamin M Neale^{1,2,3*}

‡Deceased

*These authors jointly supervised this work

¹Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA, ²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA, ³Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA, ⁴Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA, ⁵Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan, ⁶Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA, ⁷K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway, ⁸Division of Medicine and Laboratory Sciences, University of Oslo, Norway, ⁹MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Bristol, UK, ¹⁰Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA, ¹¹Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA, ¹²Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ¹³Department of Internal Medicine, Division of Cardiology, University

of Michigan, Ann Arbor, MI, USA, ¹⁴University of Groningen, UMCG, Department of Ophthalmology, Groningen, the Netherlands, ¹⁵Department of Clinical Genetics, Amsterdam University Medical Center (AMC), Amsterdam, the Netherlands, ¹⁶Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden, ¹⁷University of Groningen, UMCG, Department of Genetics, Groningen, the Netherlands, ¹⁸Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia, ¹⁹Ontario Institute for Cancer Research, Toronto, ON, Canada, ²⁰Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland, ²¹HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, Norway, ²²Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, ²³University of Colorado - Anschutz Medical Campus, Aurora, CO, USA, ²⁴Center for Neuropsychiatric Research, National Health Research Institutes, Miaoli, Taiwan, ²⁵Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA, ²⁶Division of Biostatistics, Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taiwan, ²⁷Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA, ²⁸QIMR Berghofer Medical Research Institute, Brisbane, Australia, ²⁹Faculty of Health, School of Biomedical Sciences, Queensland University of Technology, Australia, ³⁰Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA, ³¹Nuffield Department of Population Health, University of Oxford, Oxford, UK, ³²Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China, ³³MRC Population Health Research Unit, University of Oxford, Oxford, UK, ³⁴Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, ³⁵The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA, ³⁶deCODE Genetics/Amgen inc., 101, Reykjavik, Iceland, ³⁷Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK, ³⁸University of Groningen, UMCG, Department of Epidemiology, Groningen, the Netherlands, ³⁹Department of Anesthesiology, Michigan Medicine, Ann Arbor, MI, USA, ⁴⁰Institute of Healthcare Policy & Innovation, University of Michigan, Ann Arbor, MI, USA, ⁴¹Wolfson Institute of Population Health, Queen Mary University of London, London, UK, ⁴²Chinese Academy of Medical Sciences, Beijing, China, ⁴³Blizard Institute, Queen Mary University of London, London, UK, ⁴⁴Central Pharmaceutical Research Institute, JAPAN TOBACCO INC., Takatsuki 569-1125, Japan, ⁴⁵Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK, ⁴⁶Biogen, Cambridge, MA, USA, ⁴⁷NIHR Bristol Biomedical Research Centre, Bristol, UK, ⁴⁸MRC Epidemiology Unit, University of Cambridge, Cambridge, UK, ⁴⁹Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, USA, ⁵⁰Centre for Population Genomics, VA Boston Healthcare System, Boston, MA, USA, ⁵¹Institute of Precision Health, University of California, Los Angeles, Los Angeles, CA, USA, ⁵²Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ⁵³Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK, ⁵⁴Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Medicine and Health Sciences, University of Copenhagen, Copenhagen, Denmark, ⁵⁵Department of Computational Biology and Medical Sciences, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo, Japan, ⁵⁶University of Michigan, Department of Pediatrics, Ann Arbor MI 48109, ⁵⁷School of Basic and Medical Biosciences, Faculty of Life Sciences and Medicine, King's College London, London, UK, ⁵⁸Bradford Institute for Health Research, Bradford Teaching Hospitals National Health Service (NHS) Foundation Trust, Bradford, UK, ⁵⁹Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada, ⁶⁰Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ⁶¹Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ⁶²Institute for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA, ⁶³Department of Public Health & Medical Humanities, School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan, ⁶⁴Institute of Behavioral Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan, ⁶⁵Medical and Population Genomics, Wellcome Sanger Institute, Hinxton, UK, ⁶⁶Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita 565-0871, Japan, ⁶⁷Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan, ⁶⁸Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, ⁶⁹Integrated Frontier Research for Medical Science Division,

Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Suita 565-0871, Japan, ⁷⁰Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA, ⁷¹Institute for Genetics and Biomedical Research (IRGB), National Research Council (CNR), Cagliari, Italy, ⁷²Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA

Furthermore, we acknowledge each individual biobank:

Biobank Japan Project

The BioBank Japan Project was supported by the Tailor-Made Medical Treatment program of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), the Japan Agency for Medical Research and Development (AMED). S.N. was supported by Takeda Science Foundation.

Y.O. was supported by JSPS KAKENHI (19H01021, 20K21834), and AMED (JP21km0405211, JP21ek0109413, JP21ek0410075, JP21gm4010006, and JP21km0405217), JST Moonshot R&D (JPMJMS2021, JPMJMS2024), Takeda Science Foundation, and Bioinformatics Initiative of Osaka University Graduate School of Medicine, Osaka University.

BioMe - The Mount Sinai BioMe Biobank

The Mount Sinai BioMe Biobank has been supported by The Andrea and Charles Bronfman Philanthropies and in part by Federal funds from the NHLBI and NHGRI (U01HG00638001; U01HG007417; X01HL134588). We thank all participants in the Mount Sinai Biobank. We also thank all our recruiters who have assisted and continue to assist in data collection and management and are grateful for the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

BioVU

The BioVU projects at Vanderbilt University Medical Center are supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH-funded Shared Instrumentation Grant S10OD017985 and S10RR025141; CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975 from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the National Institutes of Health. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at <https://victr.vumc.org/biovu-funding/>.

Canadian Partnership for Tomorrow's Health/Ontario Health Study

This research was supported by the Canadian Institute of Health Research (CIHR) grant #EC3-144623

Colorado Center for Personalized Medicine (CCPM)

The Colorado Center for Personalized Medicine (CCPM) would like to thank Richard Zane, Steve Hess, Sarah White, Emily Hearst, Emily Roberts and the entire Health Data Compass team. CCPM was developed with support from UHealth, Children's Hospital Colorado, CU Medicine, CU Department of Medicine and CU School of Medicine.

China Kadoorie Biobank collaborative group

International Steering Committee: Junshi Chen, Zhengming Chen (PI), Robert Clarke, Rory Collins, Yu Guo, Liming Li (PI), Jun Lv, Richard Peto, Robin Walters, Chen Wang.

International Co-ordinating Centre, Oxford: Daniel Avery, Fiona Bragg, Derrick Bennett, Ruth Boxall, Ka Hung Chan, Yumei Chang, Yiping Chen, Zhengming Chen, Johnathan Clarke; Robert Clarke, Huaidong Du, Zамmy Fairhurst-Hunter, Hannah Fry, Simon Gilbert, Alex Hacker, Parisa Hariri, Mike Hill, Michael Holmes, Pek Kei Im, Andri Iona, Maria Kakkoura, Christiana Kartsonaki, Rene Kerosi, Kuang Lin, Mohsen Mazidi, Iona Millwood, Qunhua Nie, Alfred Pozarickij, Paul Ryder, Sam Sansome, Dan Schmidt, Paul Sherliker, Rajani Sohoni, Becky Stevens, Iain Turnbull, Robin Walters, Lin Wang, Neil Wright, Ling Yang, Xiaoming Yang, Pang Yao.

National Co-ordinating Centre, Beijing: Yu Guo, Xiao Han, Can Hou, Chun Li, Chao Liu, Jun Lv, Pei Pei, Canqing Yu.

Regional Coordinating Centres:

Guangxi Provincial CDC: Naying Chen, Duo Liu, Zhenzhu Tang. **Liuzhou CDC:** Ningyu Chen, Qilian Jiang, Jian Lan, Mingqiang Li, Yun Liu, Fanwen Meng, Jinhui Meng, Rong Pan, Yulu Qin, Ping Wang, Sisi Wang, Liuping Wei, Liyuan Zhou. **Gansu Provincial CDC:** Caixia Dong, Pengfei Ge, Xiaolan Ren. **Maiji CDC:** Zhongxiao Li, Enke Mao, Tao Wang, Hui Zhang, Xi Zhang. **Hainan Provincial CDC:** Jinyan Chen, Ximin Hu, Xiaohuan Wang. **Meilan CDC:** Zhendong Guo, Huimei Li, Yilei Li, Min Weng, Shukuan Wu. **Heilongjiang Provincial CDC:** Shichun Yan, Mingyuan Zou, Xue Zhou. **Nangang CDC:** Ziyang Guo, Quan Kang, Yanjie Li, Bo Yu, Qinai Xu. **Henan Provincial CDC:** Liang Chang, Lei Fan, Shixian Feng, Ding Zhang, Gang Zhou. **Huixian CDC:** Yulian Gao, Tianyou He, Pan He, Chen Hu, Huarong Sun, Xukui Zhang. **Hunan Provincial CDC:** Biyun Chen, Zhongxi Fu, Yuelong Huang, Huilin Liu, Qiaohua Xu, Li Yin. **Liuyang CDC:** Huajun Long, Xin Xu, Hao Zhang, Libo Zhang. **Jiangsu Provincial CDC:** Jian Su, Ran Tao, Ming Wu, Jie Yang, Jinyi Zhou, Yonglin Zhou. **Suzhou CDC:** Yihe Hu, Yujie Hua, Jianrong Jin Fang Liu, Jingchao Liu, Yan Lu, Liangcai Ma, Aiyu Tang, Jun Zhang. **Qingdao CDC:** Liang Cheng, Ranran Du, Ruqin Gao, Feifei Li, Shanpeng Li, Yongmei Liu, Feng Ning, Zengchang Pang, Xiaohui Sun, Xiaocao Tian, Shaojie Wang, Yaoming Zhai, Hua Zhang. **Licang CDC:** Wei Hou, Silu Lv, Junzheng Wang. **Sichuan Provincial CDC:** Xiaofang Chen, Xianping Wu, Ningmei Zhang, Weiwei Zhou. **Pengzhou CDC:** Xiaofang Chen, Jianguo Li, Jiaqiu Liu, Guojin Luo, Qiang Sun, Xunfu Zhong. **Zhejiang Provincial CDC:** Weiwei Gong, Ruying Hu, Hao Wang, Meng Wan, Min Yu. **Tongxiang CDC:** Lingli Chen, Qijun Gu, Dongxia Pan, Chunmei Wang, Kaixu Xie, Xiaoyi Zhang.

CKB Acknowledgements and Funding:

China Kadoorie Biobank gratefully acknowledges the participants, project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices. China's National Health Insurance provides electronic linkage to all hospital treatment. Funding sources: Baseline survey and first re-survey – Kadoorie Charitable Foundation, Hong Kong; long-term follow-up – UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), National Natural Science Foundation of China (91843302), National Key Research and Development Program of China (2016YFC 0900500, 0900501, 0900504, 1303904); DNA extraction and genotyping – GlaxoSmithKline, UK Medical Research Council (MC-PC-13049, MC-PC-14135); core funding for the project to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University – British Heart Foundation (CH/1996001/9454), UK Medical Research Council (MC-UU-00017/1, MC-UU-12026/2, MC_U137686851), Cancer Research UK (C16077/A29186, C500/A16896).

deCODE Genetics

We thank participants in deCODE cardiovascular and obesity studies and collaborators for their cooperation.

Estonian Biobank

This research was supported by the European Union through Horizon 2020 research and innovation programme under grant no 810645 and through the European Regional Development Fund project no. MOBEC008, by the Estonian Research Council grant PUT (PRG1291, PRG687 and PRG184) and by the European Union through the European Regional Development Fund project no. MOBERA21 (ERA-CVD project DETECT ARRHYTHMIAS, GA no JTC2018-009), Project No. 2014-2020.4.01.15-0012 and Project No.

2014-2020.4.01.16-0125.

We would like to acknowledge Dr. Tõnu Esko; Dr. Lili Milani; Dr. Reedik Mägi, Dr. Mari Nelis and Dr. Andres Metspalu, all from the Institute of Genomics, University of Tartu, Tartu, Estonia.

FinnGen

The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, and Novartis AG. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (<https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshP.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB.

Generation Scotland

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “STratifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/Z/14/Z).”

Genes and Health

Genes & Health is/has recently been core-funded by Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017), Higher Education Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London substantive site), and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). We thank Social Action for Health, Centre of The Cell, members of our Community Advisory Group, and staff who have recruited and collected data from volunteers. We thank the NIHR National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry Centre (King's College London), Wellcome Sanger Institute, and Broad Institute for sample processing, genotyping, sequencing and variant annotation. We thank: Barts Health NHS Trust, NHS Clinical Commissioning Groups (City and Hackney, Waltham Forest, Tower Hamlets, Newham, Redbridge, Havering, Barking and Dagenham), East London NHS Foundation Trust, Bradford Teaching Hospitals NHS Foundation Trust, Public Health England (especially David Wyllie), Discovery Data Service/Endeavour Health Charitable Trust (especially David Stables) - for GDPR-compliant data sharing backed by individual written informed consent. Most of all we thank all of the volunteers participating in Genes & Health.

Genes & Health Research Team (in alphabetical order by surname): Shaheen Akhtar, Mohammad Anwar, Elena Arciero, Samina Ashraf, Gerome Breen, Raymond Chung, Charles J Curtis, Maharun Chowdhury, Grainne Colligan, Panos Deloukas, Ceri Durham, Sarah Finer, Chris Griffiths, Qin Qin Huang, Matt Hurles, Karen A Hunt, Shapna Hussain, Kamrul Islam, Ahsan Khan, Amara Khan, Cath Lavery, Sang Hyuck Lee, Robin Lerner, Daniel MacArthur, Bev MacLaughlin, Hilary Martin, Dan Mason, Shefa Miah, Bill Newman, Nishat Safa, Farah Tahmasebi, Richard C Trembath, Bhavi Trivedi, David A van Heel, John Wright.

The HUNT Study

A special thanks to all the HUNT participants for donating their time, samples and information to help others.

The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping in HUNT was financed by the National Institutes of Health; University of Michigan; the Research Council of Norway; the Liaison Committee for Education, Research and Innovation in Central Norway; and the Joint Research Committee between St Olavs hospital and the Faculty of Medicine and Health Sciences, NTNU. The genetic investigations of the HUNT Study, is a collaboration between researchers from the K.G. Jebsen Center for Genetic Epidemiology, NTNU and the University of Michigan Medical School and the University of Michigan School of Public Health. The K.G. Jebsen Center for Genetic Epidemiology is financed by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU, Norway.

We want to thank clinicians and other employees at Nord-Trøndelag Hospital Trust for their support and for contributing to data collection in this research project.

We also acknowledge; **HUNT-MI Leadership:** Kristian Hveem, Cristen Willer, Oddgeir Lingaas Holmen, Mike Boehnke, Goncalo Abecasis, Bjorn Olav Åsvold, Ben Brumpton; **Scientific Advisory Committee:** Ele Zeggini, Mark Daly, Bjørn Pasternak; **HUNT Research Centre:** Jørn Søberg Fenstad, Anne Jorunn Vikdal, Marit Næss; **HUNT Cloud:** Oddgeir Lingaas Holmen, Sandor Zeestraten, Tom Erik Røberg; **Data applications and registry linkages:** Maiken E. Gabrielsen, Anne Heidi Skogholt; **Low-pass whole sequencing genome bioinformatics and statistical analysis:** He Zhang, Hyun Min Kang, Jin Chen; **Array genotyping:** Sten Even Erlandsen, Vidar Beisvåg; **GWAS bioinformatics, QC, imputation and statistical analysis:** Wei Zhou, Jonas Nielsen, Lars Fritsche, Hyun Min Kang, Oddgeir Holmen, Ben Brumpton, Laurent Thomas; **CNV calling:** Ellen Schmidt, Ryan Mills; **Statistical methods development for analyzing HUNT data:** Wei Zhou, Shawn Lee

Funding:

The K. G. Jebsen Centre for Genetic Epidemiology is financed by Stiftelsen Kristian Gerhard Jebsen. The genotyping in HUNT was financed by the National Institutes of Health; University of Michigan; the Research Council of Norway; Stiftelsen Kristian Gerhard Jebsen; the Liaison Committee for Education, Research and Innovation in Central Norway; and the Joint Research Committee between St Olav's hospital and the Faculty of Medicine and Health Sciences, NTNU.

LifeLines

The Lifelines Biobank initiative has been made possible by funding from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG the Netherlands), University of Groningen and the Northern Provinces of the Netherlands. The generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the UMCG Genetics Lifelines Initiative (UGLI). UGLI is partly supported by a Spinoza Grant from NWO, awarded to Cisca Wijmenga.

The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

UMCG Genetics Lifelines Initiative (UGLI) team: Raul Aguirre-Gamboa (1), Patrick Deelen (1), Lude Franke (1), Jan A Kuivenhoven (2), Esteban A Lopera Maya (1), Ilja M Nolte (3), Serena Sanna (1), Harold Snieder (3), Morris A Swertz (1), Peter M. Visscher (3,4), Judith M Vonk (3), Cisca Wijmenga (1)

- (1) *Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands*
- (2) *Department of Pediatrics, University of Groningen, University Medical Center Groningen, The Netherlands*
- (3) *Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands*
- (4) *Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia.*

Mass General Brigham (MGB) Biobank

Samples, genomic data, and health information were obtained from the Mass General Brigham Biobank, a biorepository of consented patients samples at Mass General Brigham (parent organization of Massachusetts General Hospital and Brigham and Women's Hospital). We are grateful to all of the participants and clinical and research teams who made this work possible. Support for genotyping was provided through MGB Personalized Medicine.

MGB Biobank Leadership: Elizabeth W. Karlson, MD; Shawn N. Murphy, MD, PhD; Susan A Slauchaupt, PhD; Jordan W. Smoller, MD, ScD; Scott T. Weiss, MD, MSc

Michigan Genomics Initiative

The authors acknowledge the Michigan Genomics Initiative participants, Precision Health at the University of Michigan, the University of Michigan Medical School Central Biorepository, and the University of Michigan Advanced Genomics Core for providing data and specimen storage, management, processing, and distribution services, and the Center for Statistical Genetics in the Department of Biostatistics at the School of Public Health for genotype data curation, imputation, and management in support of the research reported in this publication.

QIMR Berghofer Biobank

This work is supported by a project grant (APP1063061) and a Program grant (APP1073898) from the Australian National Health and Medical Research Council (NHMRC). SM is supported by a Research Fellowship from the NHMRC (Aust). NI received scholarship support from the University of Queensland and QIMR Berghofer Medical Research Institute. We thank research staff (David Whiteman, Catherine Olsen, Rachel Neale) and participants from the Australian QSKIN study (<https://www.qimrberghofer.edu.au/study/qskin/>).

SEM is supported in part by APP1172917 from the Australian National Health and Medical Research Council (NHMRC).

Taiwan Biobank

This research has been conducted using the Taiwan Biobank resource. We thank all the participants and investigators of the Taiwan Biobank. We thank the National Center for Genome Medicine of Taiwan for the technical support in genotyping. We thank the National Core Facility for Biopharmaceuticals (NCFB, MOST 106-2319-B-492-002) and National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) of Taiwan for providing computational and storage resources.

UCLA ATLAS Community Health Initiative (UCLA)

We gratefully acknowledge the resources provided by the Institute for Precision Health (IPH) and participating UCLA ATLAS Community Health Initiative patients. The UCLA ATLAS Community Health Initiative in collaboration with UCLA ATLAS Precision Health Biobank, is a program of IPH, which directs and supports the biobanking and genotyping of biospecimen samples from participating UCLA patients in collaboration with the David Geffen School of Medicine, UCLA CTSI and UCLA Health. Members of the UCLA ATLAS Community Health Initiative include Ruth Johnson, Yi Ding, Vidhya Venkateswaran, Arjun Bhattacharya, Alec Chiu, Tommer Schwarz, Malika Freund, Lingyu Zhan, Kathryn S. Burch, Christa Caggiano, Brian Hill, Nadav Rakocz, Brunilda Balliu, Jae Hoon Sul, Noah Zaitlen, Valerie A. Arboleda, Eran Halperin, Sriram Sankararaman, Manish J. Butte, Clara Lajonchere, Daniel H. Geschwind, and Bogdan Pasaniuc, on behalf of the UCLA Precision Health Data Discovery Repository Working Group and UCLA Precision Health ATLAS Working Group.

UK Biobank

Access to data from the UK BioBank was obtained through Application #31063 PI: Ben Neale, Claire Churchhouse

Overview: "Methodological extensions to estimate genetic heritability and shared risk factors for phenotypes of the UK Biobank".

Website for Pan-UKBB results can be found: <https://pan.ukbb.broadinstitute.org/>

Other

G.D.S.; T.R.G.; and J.Z. are supported by a grant from the Medical Research Council for the Integrative Epidemiology Unit at the University of Bristol MC_UU_00011/1 & 4.

J.Z. is supported by the Academy of Medical Sciences (AMS) Springboard Award, the Wellcome Trust, the Government Department of Business, Energy and Industrial Strategy (BEIS), the British Heart Foundation and Diabetes UK (SBF006\1117). J.Z. is funded by the Vice-Chancellor Fellowship from the University of Bristol. W.Z. was supported by the National Human Genome Research Institute of the National Institutes of Health under award number T32HG010464.

ICDA

The authors would like to acknowledge the organizing committee of the International Common Disease Alliance for intellectual contributions on the set up of the GBMI as a nascent activity to the larger effort. We also thank them for the use of their slack platform. Website for ICDA can be found here: <https://www.icda.bio/>

The Hail Team and Data Management at the Stanley Center for Psychiatric Research Hail is an open-source Python library that simplifies genomic data analysis in the cloud. It provides powerful, easy-to-use data science tools that can be used to interrogate biobank-scale genomic data and was used in the analysis of the data for this paper. We would especially like to thank Daniel King from the Hail team and Sam Bryant from the Stanley Center Data Management team for helping with the Google bucket set up and data sharing. Website for Hail can be found here: <https://hail.is/>