THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Global Biobank Meta-analysis Initiative: powering genetic discovery across human disease

Citation for published version:
Hayward, C, Campbell, A, Porteous, DJ, Richmond , A & Marioni, RE 2022, 'Global Biobank Meta-analysis Initiative: powering genetic discovery across human disease', *Cell Genomics*.

Link:
Link to publication record in Edinburgh Research Explorer

Published In:
Cell Genomics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 01. Nov. 2022

Acknowledgments

Biobank of the Americas
We are grateful to all our participants for being a part of this study

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 UCHealth, 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, Zammy 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, Jinhuai 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:** Ziyan 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 & Health Research Team

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). Genes & Health is/has recently been funded by Alnylam Pharmaceuticals, Genomics PLC; and a Life Sciences Industry Consortium of Bristol-Myers

Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc, Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc, Takeda Development Centre Americas Inc.

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.

Current 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 Slaugenhaupt, 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.

National Biobank of Korea

This research was supported by Brain Pool Plus (BP+, Brain Pool+) Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2020H1D3A2A03100666). Data in this study were from the Korean Genome and Epidemiology

Study (KoGES; 4851-302), National Research Institute of Health, Centers for Disease Control and Prevention, Ministry for Health and Welfare, Republic of Korea.

Penn Medicine Biobank
We acknowledge the Penn Medicine BioBank (PMBB) for providing data and thank the patient-participants of Penn Medicine who consented to participate in this research program. We would also like to thank the Penn Medicine BioBank team and Regeneron Genetics Center for providing genetic variant data for analysis. The PMBB is approved under IRB protocol# 813913 and supported by Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878.

Qatar Biobank
The Qatar Biobank (QBB) and The Qatar Genome Program (QGP) are both Research, Development and Innovations entities within Qatar Foundation for Education, Science and Community Development. We would like to thank all participants in Qatar Biobank. We also thank all Qatar Biobank staff who have assisted and continue to assist in participant recruitment, data collection, biological sample collection, processing and data management. We also thank QBB stakeholders, Hamad Medical Corporation and Sidra, for their contribution.

The QSkin Sun and Health Study
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). SM is a co-founder and holds stock in StratifEYE.

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 theUCLA 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

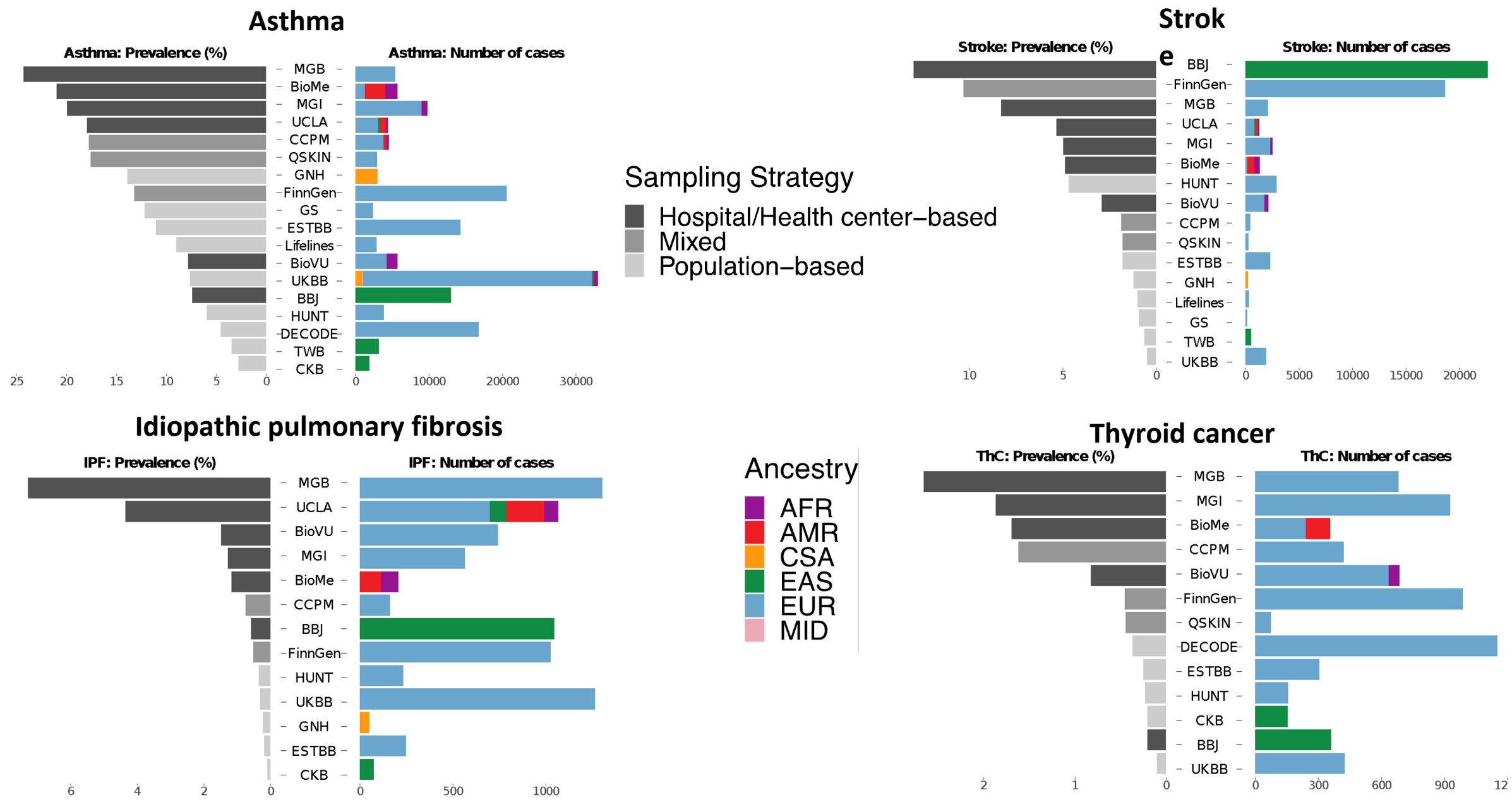
G.D.S.; T.R.G.; and J.Z. are supported by a grant from the Medical Research Council for the Integrative Epidemiology Unit at the University of Bristol MC_UU_00011/1 & 4. J.Z. is supported by the Academy of Medical Sciences (AMS) Springboard Award, the Wellcome Trust, the Government Department of Business, 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. EALM was sponsored by the Colciencias scholarship edition 783. W.Z. was supported by the National Human Genome Research Institute of the National Institutes of Health under award number T32HG010464 and K99HG012222-01.

ICDA

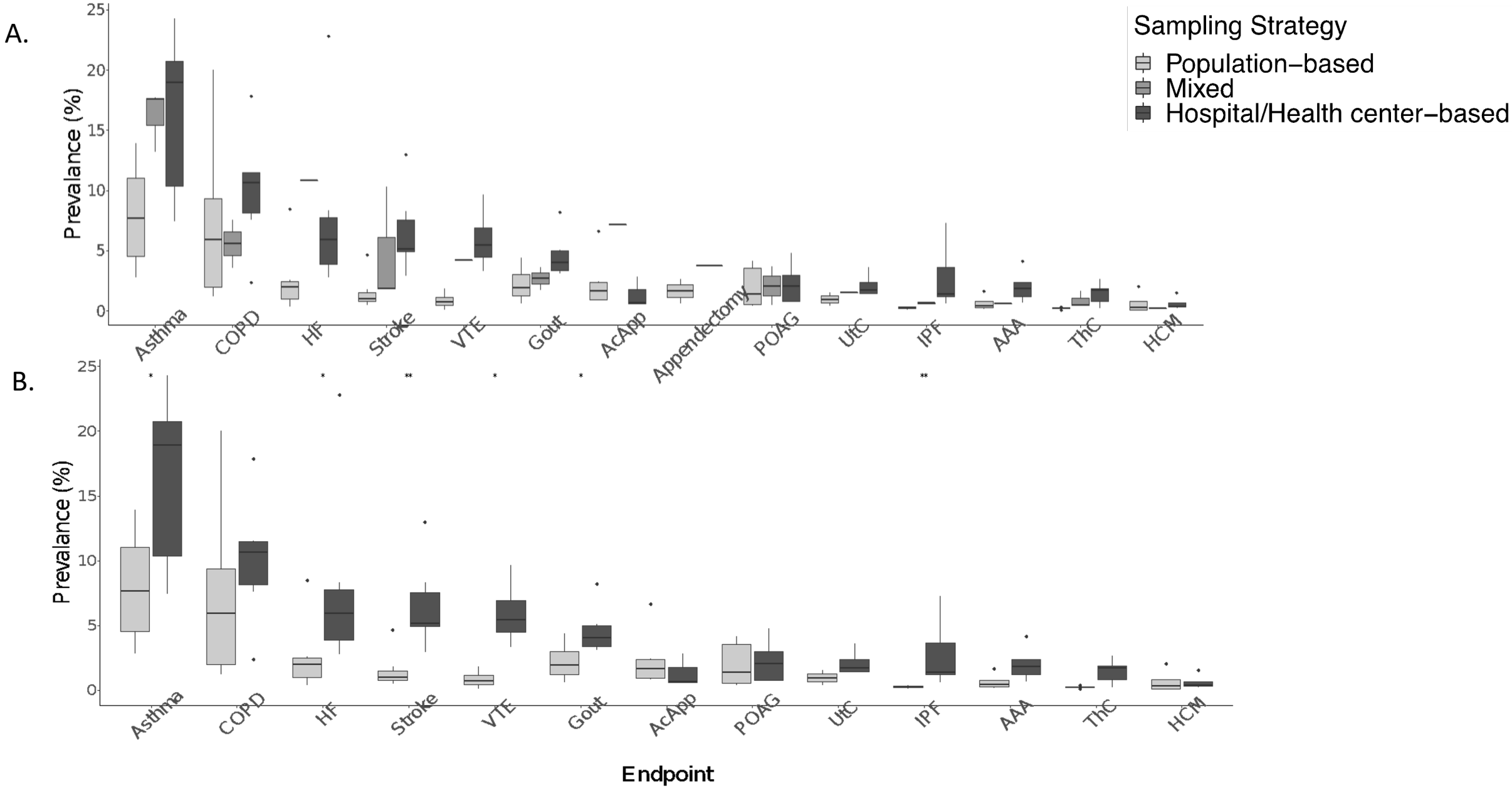
The authors would like to acknowledge the organizing committee of the International Common Disease Alliance for intellectual contributions on the set up of the GBMI as a nascent activity to the larger effort. We also thank them for the use of their slack platform.
Website for ICDA can be found here: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

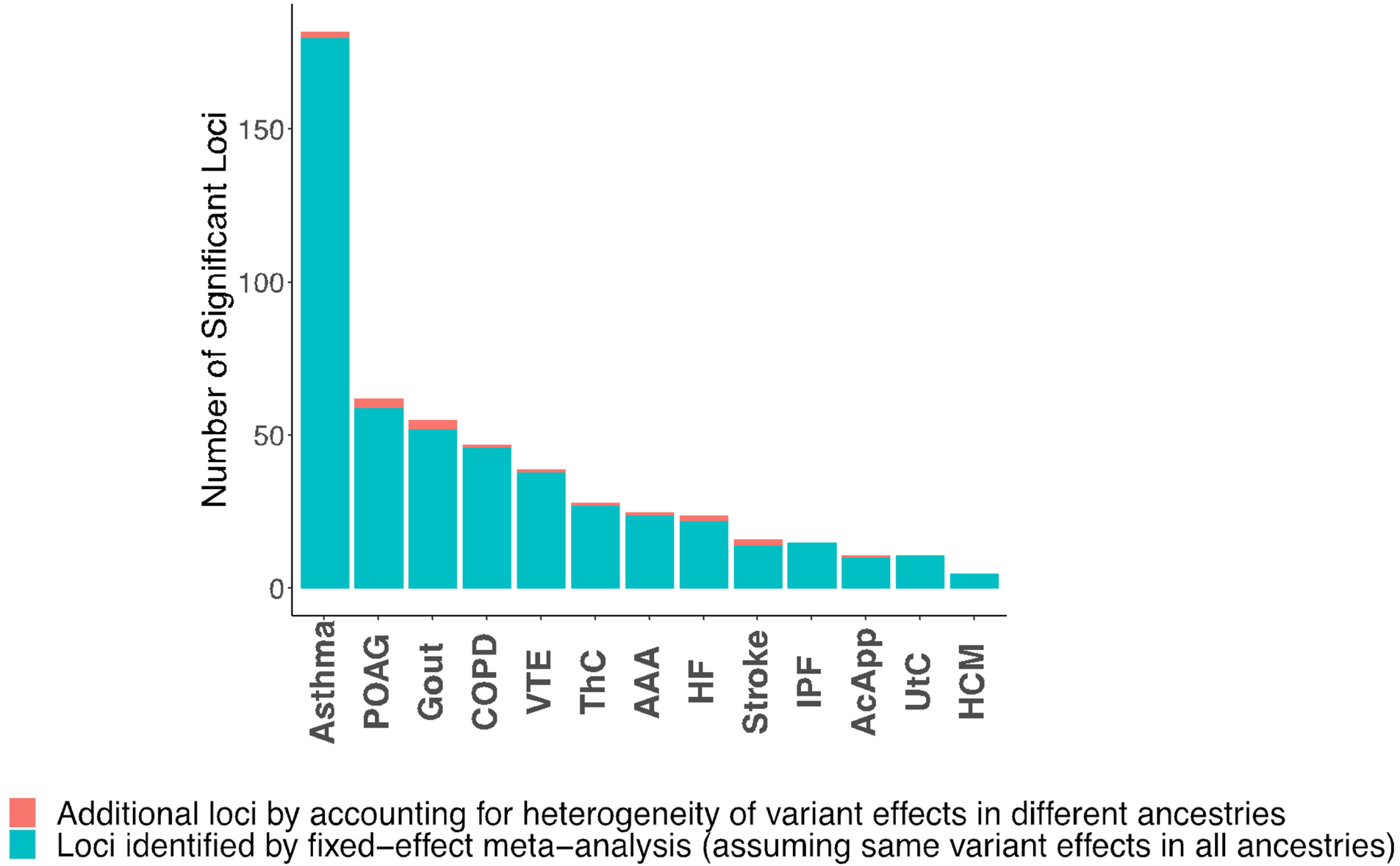
Supplementary Figure 1. Disease prevalence varies across biobanks



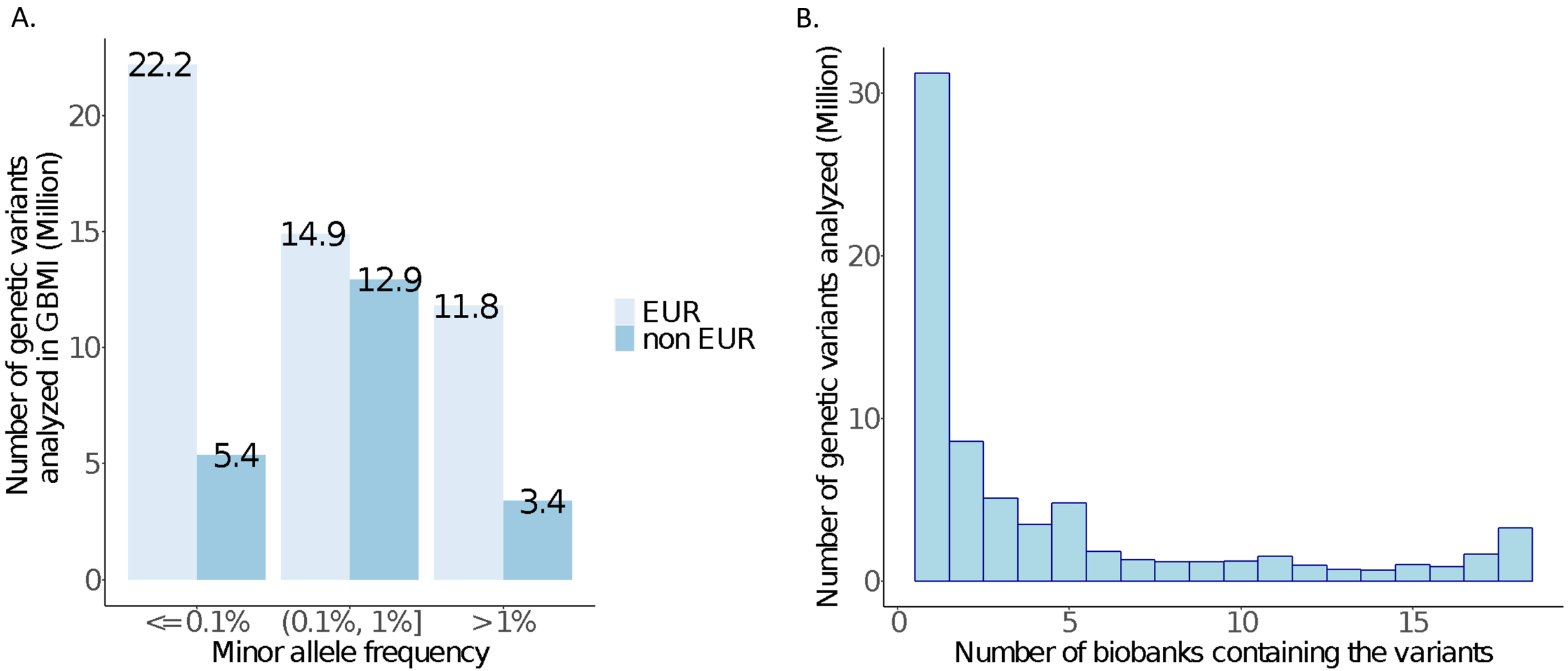
Supplementary Figure 2. Disease prevalence varies by difference sample recruiting strategies. A. Box plots for prevalence by three sampling strategies. B. Box plots to compare prevalence between population-based and hospital/health center-based biobanks. (**, P < 0.01, *, P < 0.05, unpaired Wilcoxon test)



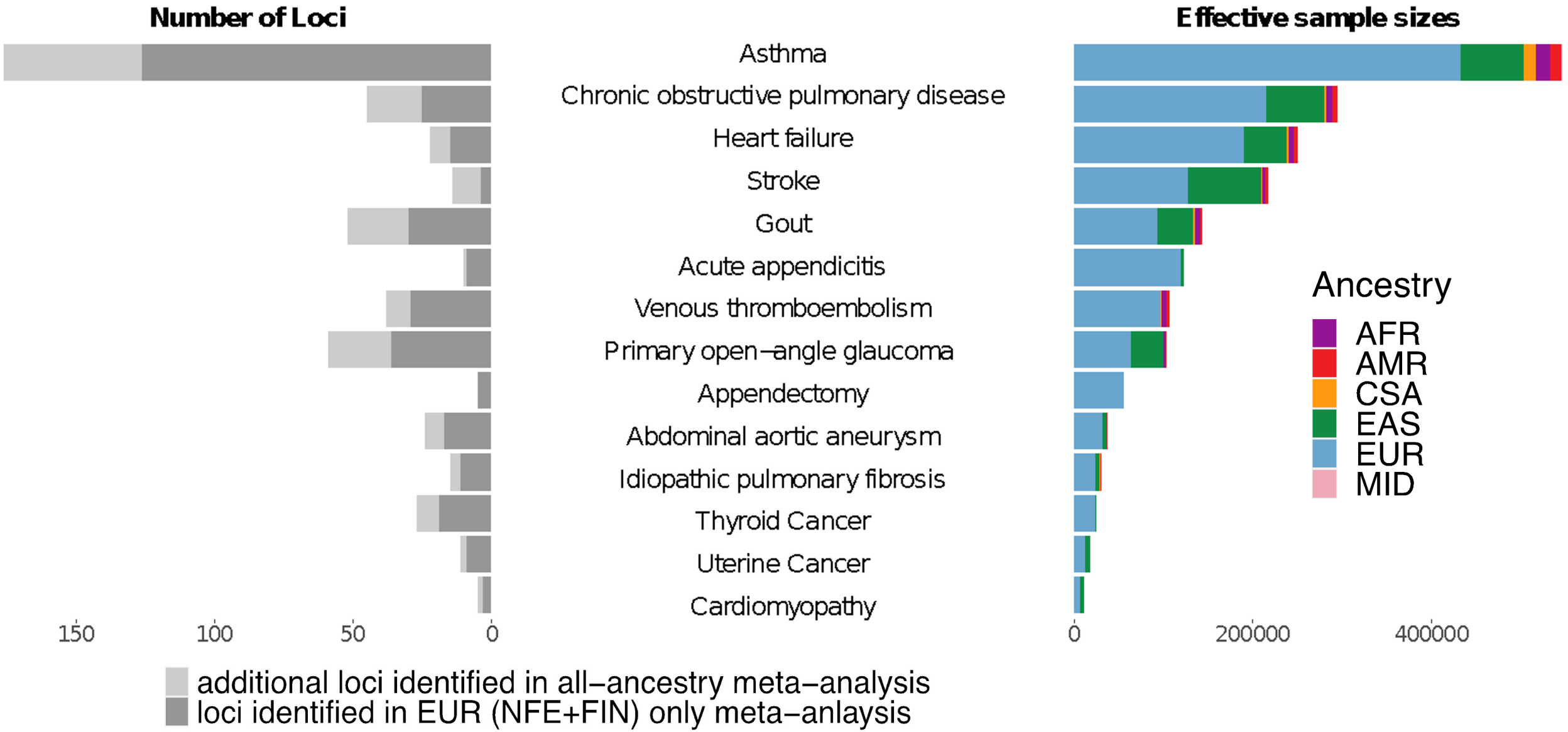
Supplementary Figure 3. Additional significant loci identified by the meta-regression approach implemented in MR-MEGA[30] to account for effect size heterogeneity across different data sets in meta-analyses compared to the fixed-effect meta-analyses



Supplementary Figure 4. A. Additional genetic variants analyzed due to incorporating non-European samples. EUR: genetic variants observed in samples with European ancestry. non EUR: genetic variants only observed in samples with non-European ancestry. The highest minor allele frequency (MAF) among non EUR ancestry was used in the plot. B. Distribution of the number of biobanks in which the genetic variants were tested



Supplementary Figure 5A. Additional significant loci identified when non-European samples were included in the meta-analysis



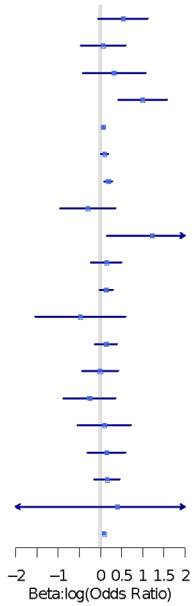
Supplementary Figure 5B. Additional significant loci identified when non-European samples were included in the meta-analysis

POAG

rs1355927 (intergenic)

MIR2054;INTU

Biobank	Ancestry	Freq	Beta	P–value
BioMe	AFR	0.0503	0.5416	7.55e–02
BioVU	AFR	0.0404	0.074	7.85e–01
UKBB	AFR	0.0492	0.3286	3.92e–01
BioMe	AMR	0.0541	1.007	7.80e–04
BBJ	EAS	0.3141	0.0731	3.40e–05
TWB	EAS	0.2112	0.103	4.16e–02
FinnGen	EUR	0.0554	0.1898	1.48e–04
BioVU	EUR	0.014	–0.288	3.94e–01
CCPM	EUR	0.0158	1.2276	2.57e–02
DECODE	EUR	0.0261	0.1497	4.31e–01
ESTBB	EUR	0.0194	0.1412	9.02e–02
GS	EUR	0.014	–0.47	3.92e–01
HUNT	EUR	0.031	0.1408	3.07e–01
Lifelines	EUR	0.0136	–0.0012	9.96e–01
MGB	EUR	0.0121	–0.2518	4.34e–01
MGI	EUR	0.0152	0.0988	7.61e–01
QSKIN	EUR	0.0149	0.153	5.18e–01
UKBB	EUR	0.0144	0.1647	2.91e–01
UKBB	CSA	0.0083	0.4025	7.45e–01
Meta–analysis		0.0723	0.0938	5.18e–10

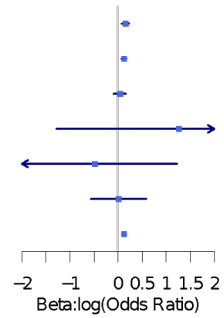


COPD

rs75550771 (intergenic)

PNPT1;EFEMP1

Biobank	Ancestry	Freq	Beta	P–value
BBJ	EAS	0.12	0.1568	1.03e–04
CKB	EAS	0.1047	0.1285	1.04e–06
TWB	EAS	0.1269	0.0395	5.59e–01
CCPM	EUR	0.0007	1.2617	3.30e–01
MGI	EUR	0.0004	–0.4863	5.80e–01
GNH	CSA	0.0326	0.0112	9.70e–01
Meta–analysis		0.0963	0.1269	1.24e–09

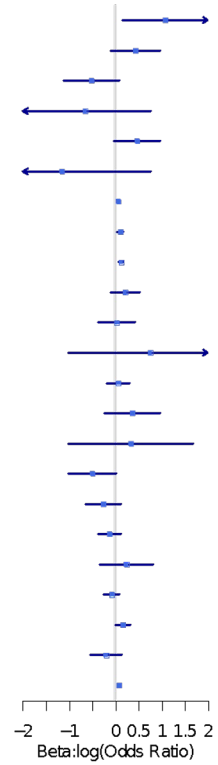


Asthma

rs74252294 (Intronic)

NAA38

Biobank	Ancestry	Freq	Beta	P–value
CCPM	AFR	0.0086	1.0801	2.32e–02
MGI	AFR	0.0124	0.4358	1.12e–01
UKBB	AFR	0.0112	–0.514	9.63e–02
UKBB	AMI	0.0092	–0.6576	3.68e–01
CCPM	AMR	0.0235	0.4719	6.71e–02
UKBB	AMR	0.0119	–1.154	2.40e–01
BBJ	EAS	0.1956	0.0656	1.24e–04
CKB	EAS	0.2272	0.1072	9.68e–03
TWB	EAS	0.1991	0.1218	3.53e–04
UKBB	EAS	0.1815	0.2073	2.07e–01
CCPM	EUR	0.0039	0.0251	9.04e–01
DECODE	EUR	0	0.7493	4.07e–01
ESTBB	EUR	0.0025	0.0584	6.52e–01
GS	EUR	0.0027	0.3664	2.37e–01
HUNT	EUR	0.0004	0.336	6.27e–01
Lifelines	EUR	0.0031	–0.4923	6.63e–02
MGB	EUR	0.0034	–0.2586	1.92e–01
MGI	EUR	0.0043	–0.1223	3.46e–01
QSKIN	EUR	0.0028	0.2378	4.21e–01
UKBB	EUR	0.0023	–0.079	3.95e–01
GNH	CSA	0.0314	0.1594	6.48e–02
UKBB	CSA	0.0226	–0.2023	2.46e–01
Meta–analysis		0.0481	0.0747	3.56e–08

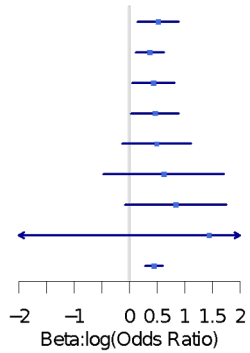


VTE

rs112106699 (intergenic)

VPS13D;DHRS3

Biobank	Ancestry	Freq	Beta	P–value
BioMe	AFR	0.0906	0.5228	5.66e–03
BioVU	AFR	0.0599	0.3738	3.75e–03
MGI	AFR	0.0806	0.4379	2.62e–02
UKBB	AFR	0.0986	0.4687	3.29e–02
BioMe	AMR	0.0258	0.4987	1.14e–01
MGI	EUR	0.0014	0.6274	2.59e–01
UKBB	EUR	0.0005	0.8431	7.03e–02
UKBB	CSA	0.0031	1.4447	5.31e–01
Meta–analysis		0.0045	0.4538	2.04e–08

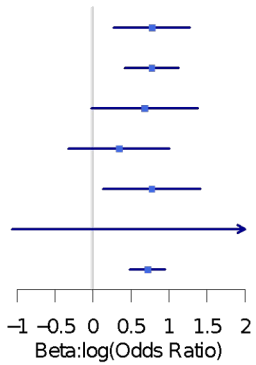


Heart failure

rs145478347 (intronic)

BCL2L12

Biobank	Ancestry	Freq	Beta	P–value
BioMe	AFR	0.033	0.7792	2.33e–03
BioVU	AFR	0.0238	0.7758	1.67e–05
MGI	AFR	0.0372	0.6809	5.66e–02
UKBB	AFR	0.0465	0.3459	3.07e–01
BioMe	AMR	0.0136	0.7766	1.70e–02
MGI	EUR	0.0003	2.0703	1.96e–01
Meta–analysis		0.0114	0.7217	8.17e–10

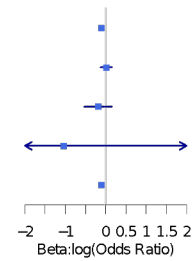


Stroke

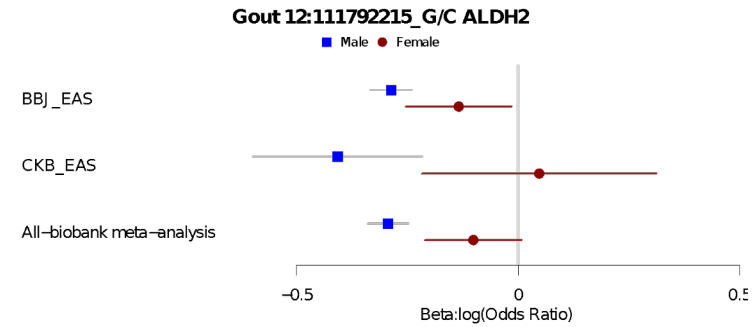
12:111803962 (nonsynonymous)

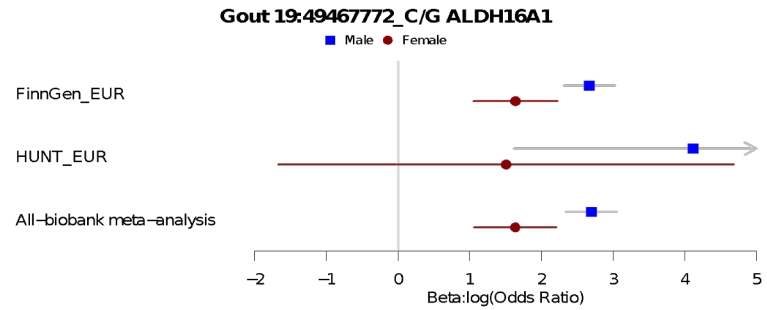
ALDH2

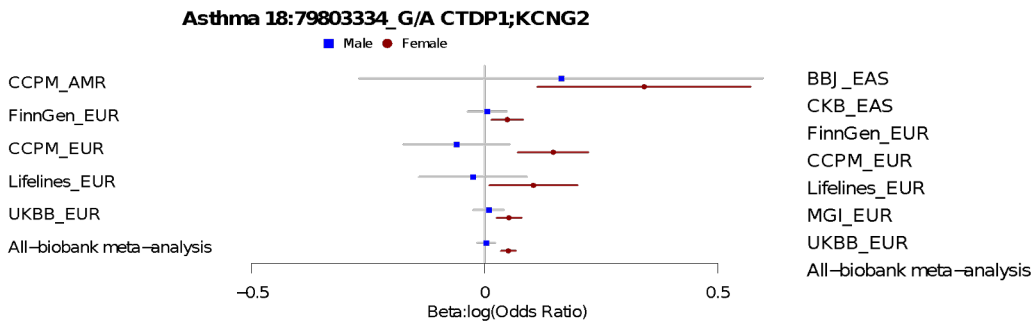
Biobank	Ancestry	Freq	Beta	P–value
BBJ	EAS	0.2469	–0.1098	1.99e–18
TWB	EAS	0.284	0.0164	8.23e–01
UCLA	EAS	0.2028	–0.1869	2.86e–01
CCPM	EUR	0.0008	–1.0403	7.85e–01
Meta–analysis		0.2376	–0.1066	5.16e–18

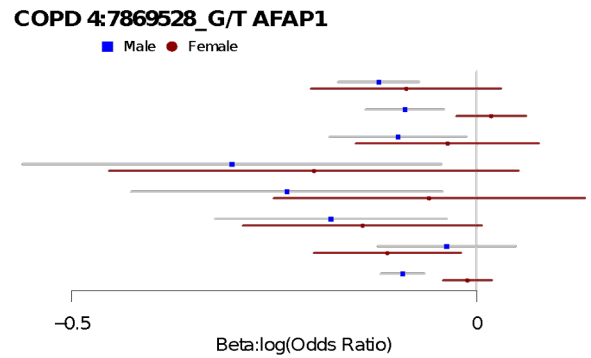


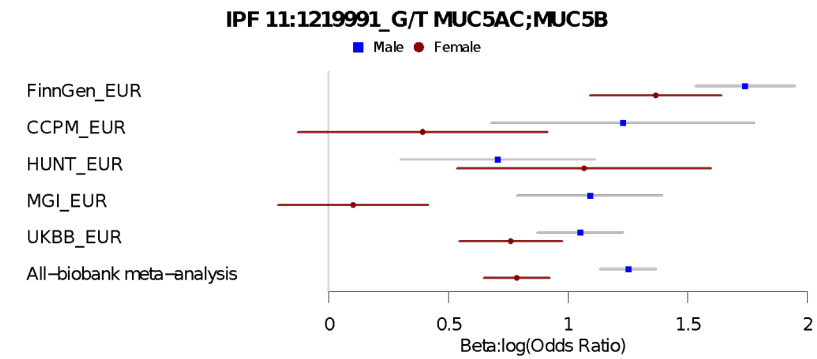
Supplementary Figure 6. LocusZoom plots of region showing differential association between sexes. All-biobank meta-analysis results stratified by sex were filtered to identify regions with different effect sizes in men and women (Phet for Cochran's Q test < 0.002).

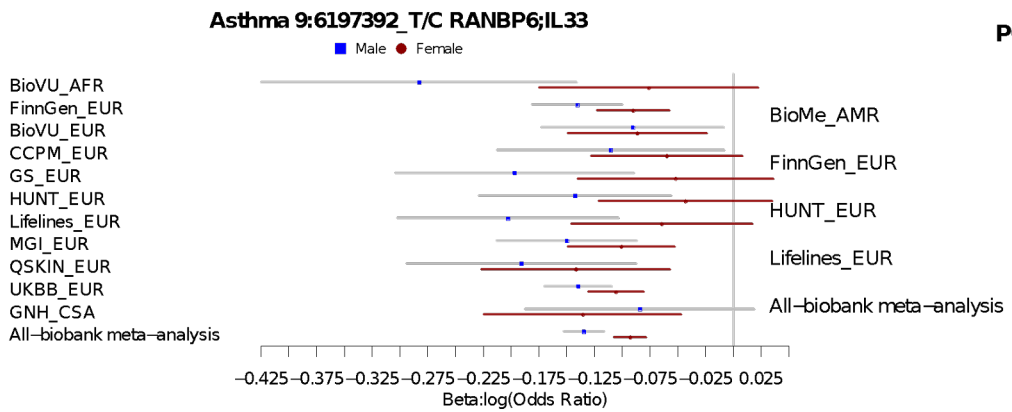


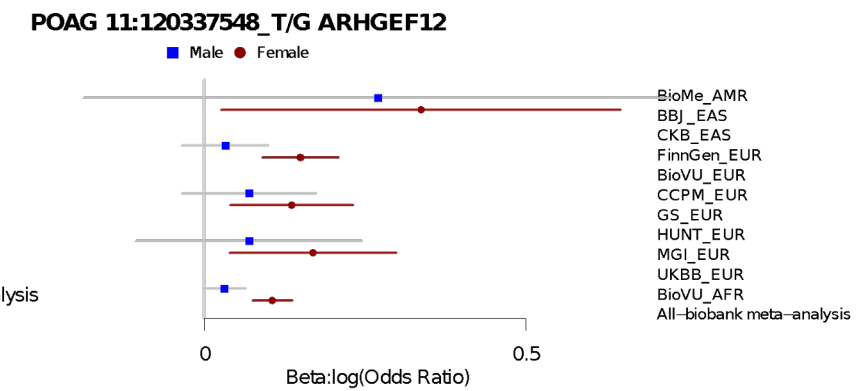


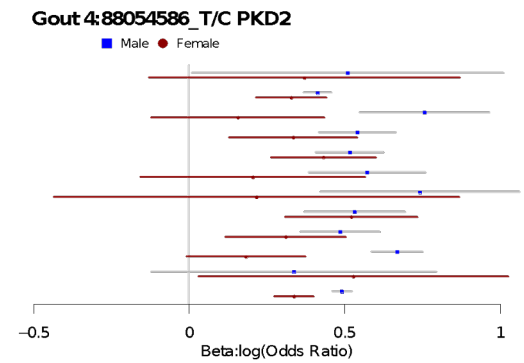






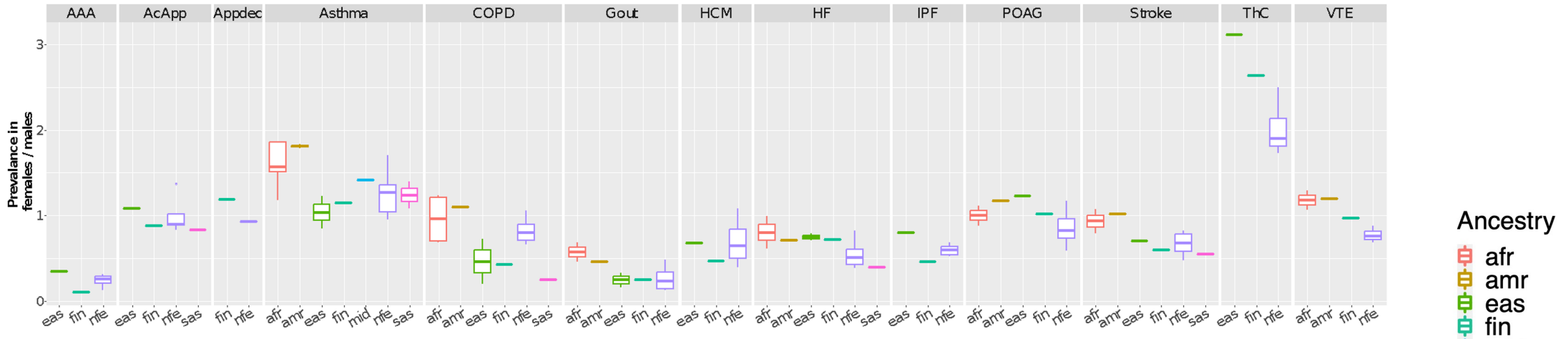




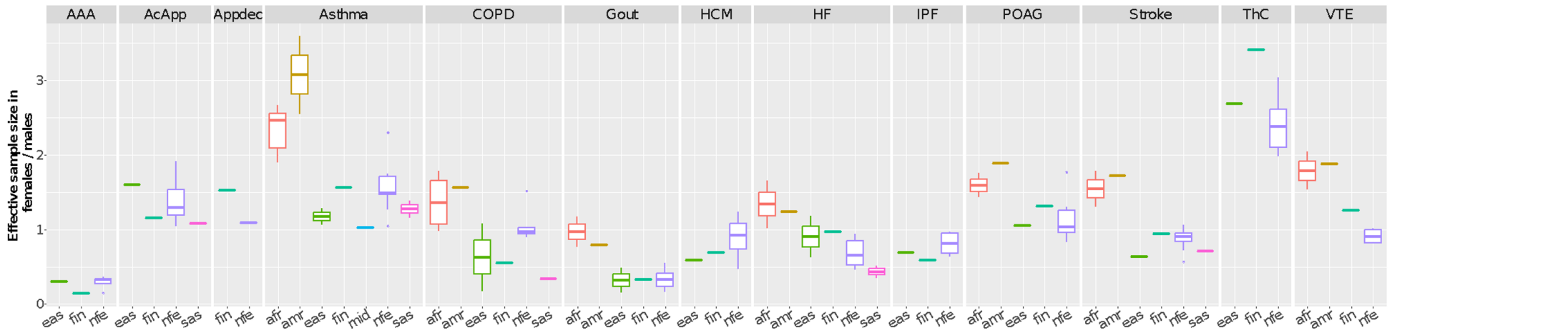


Supplementary Figure 7. Plots of A. prevalence ratios and B. effective sample sizes ratios in females and in males by ancestry across endpoints. The prevalence is calculated as number of cases /(number of cases + number of controls). The effective sample sizes is calculated as 4/(1/ number of cases + 1/ number of controls).

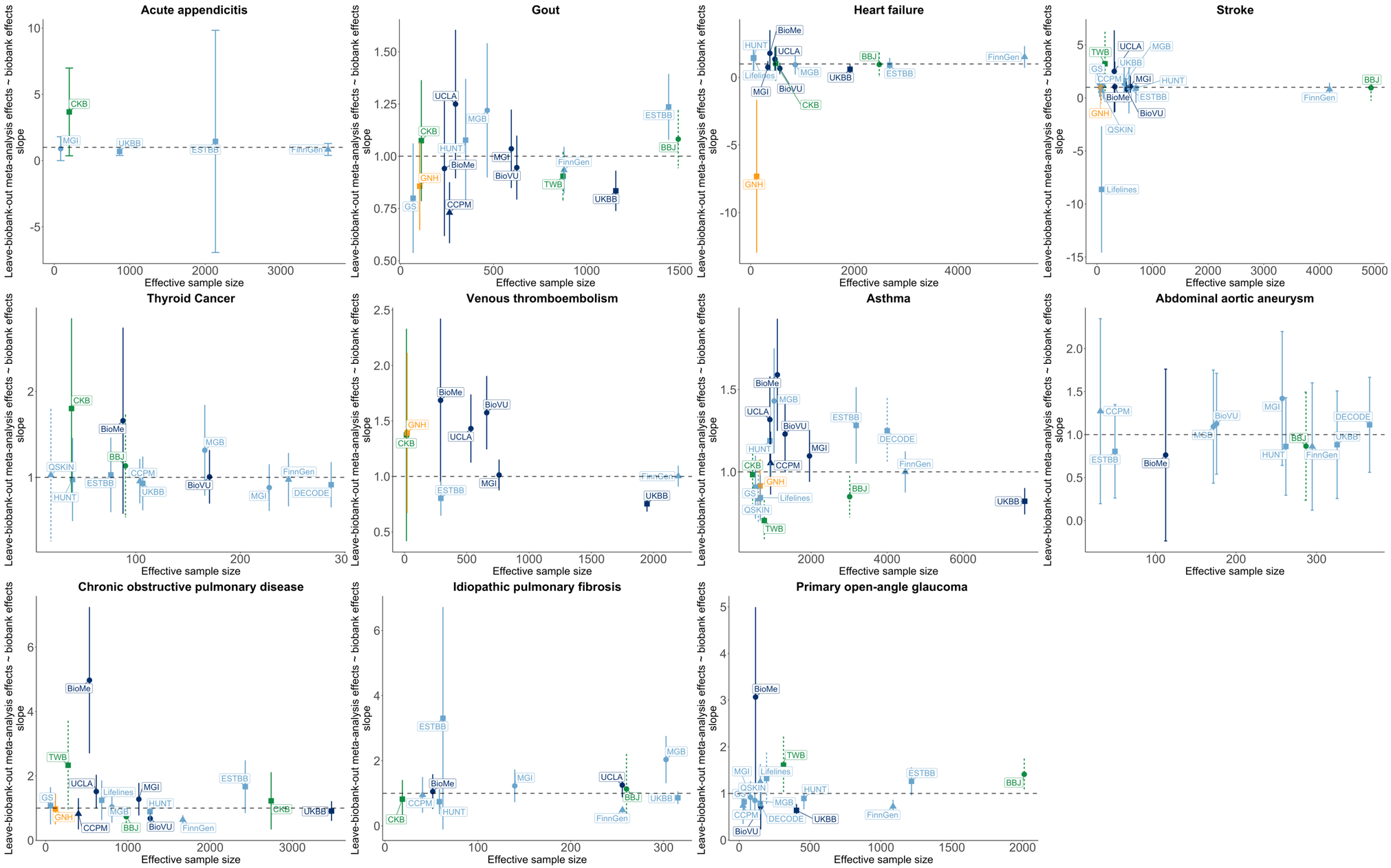
A.



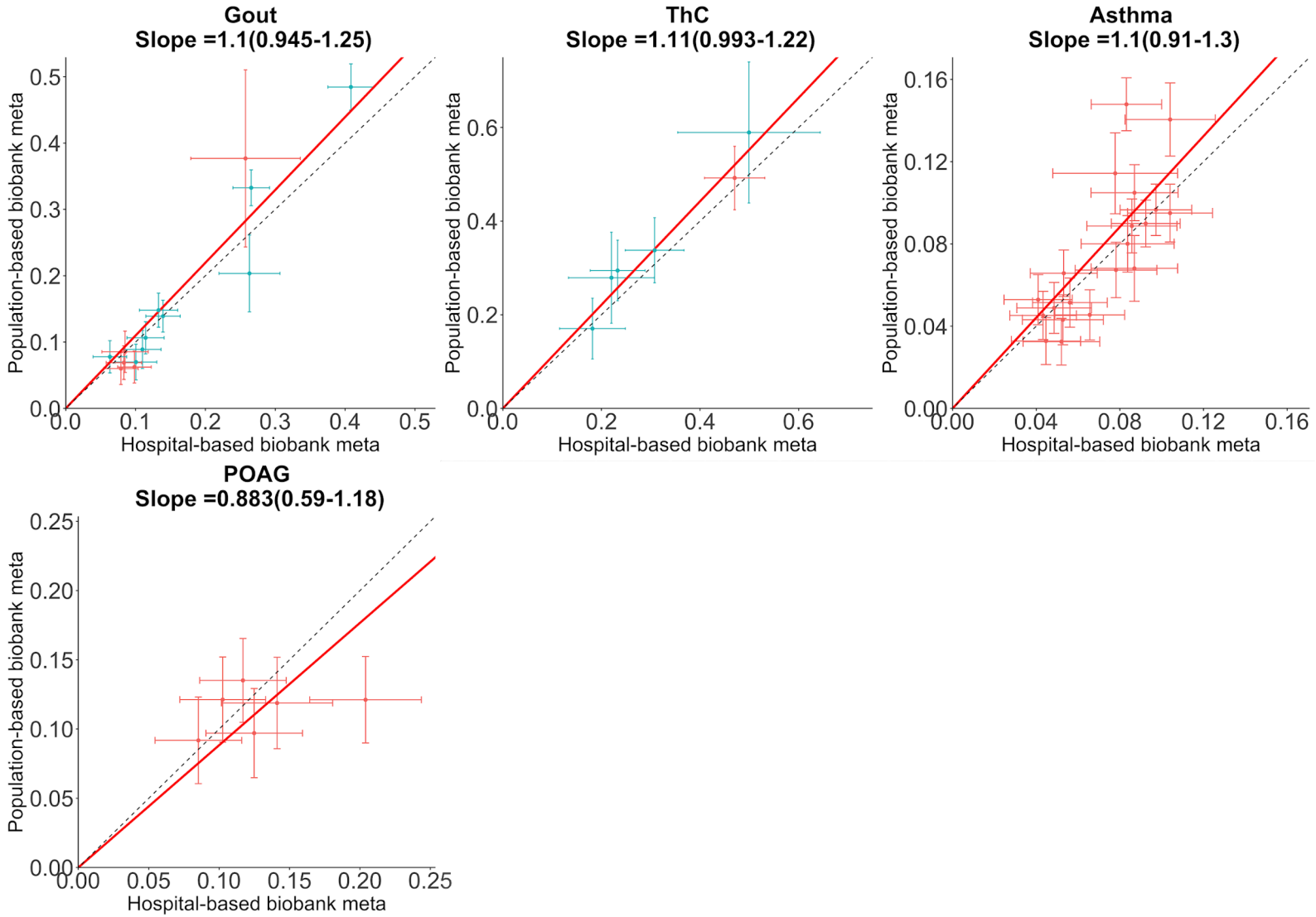
B.



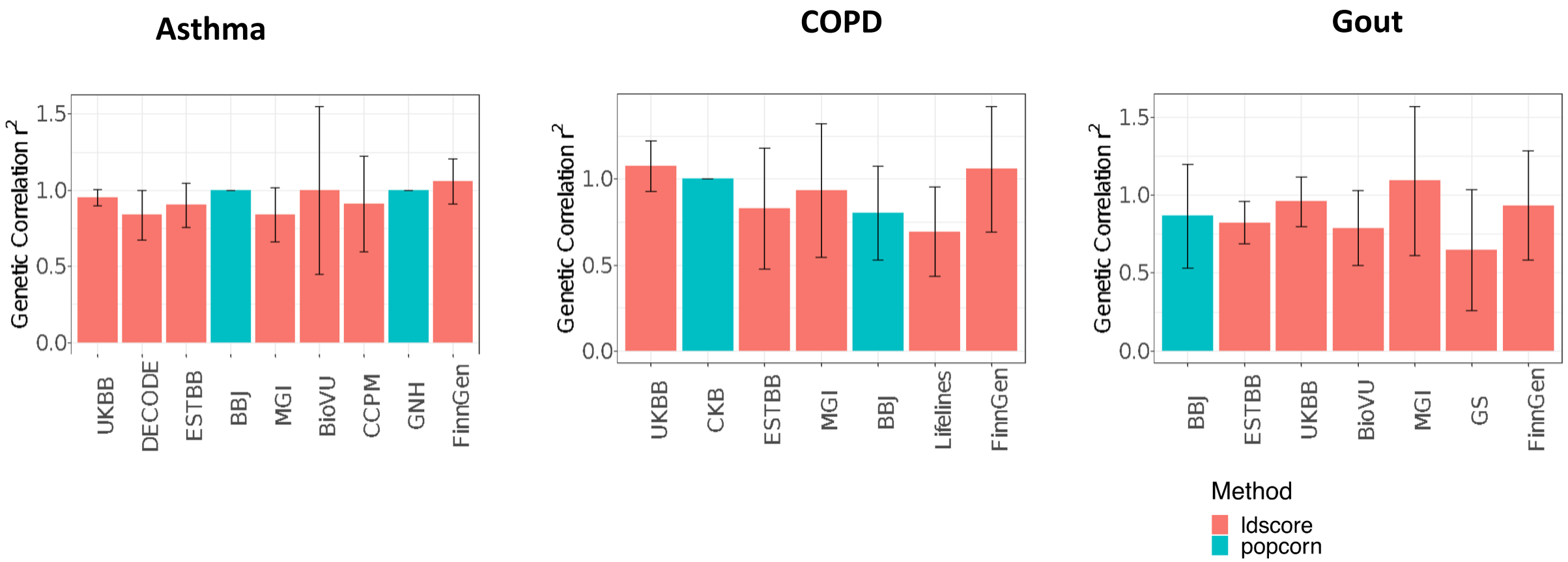
Supplementary Figure 8. The slopes of Deming regression for effect sizes for index variants in each biobank and leave-one-biobank-out meta-analysis (LOBO) pair are plotted against the effective sample sizes. Index variants with association p-values < $1x10^{-10}$ in the all-biobank meta-analysis were used for the regression. Biobanks, in which at least three index variants passed the cutoff, are plotted. Biobanks are annotated by phenotype source, sampling strategy and sample ancestry. The dotted line indicates y=1. A positive slope indicates that effect size estimates of the top hits are higher in the leave-one-biobank-out (LOBO) meta-analysis than in the individual biobank and a negative slope suggests lower effect size estimates in LOBO meta-analysis than in the individual biobank. The effective sample sizes is calculated as 4/(1/case number + 1/control number).



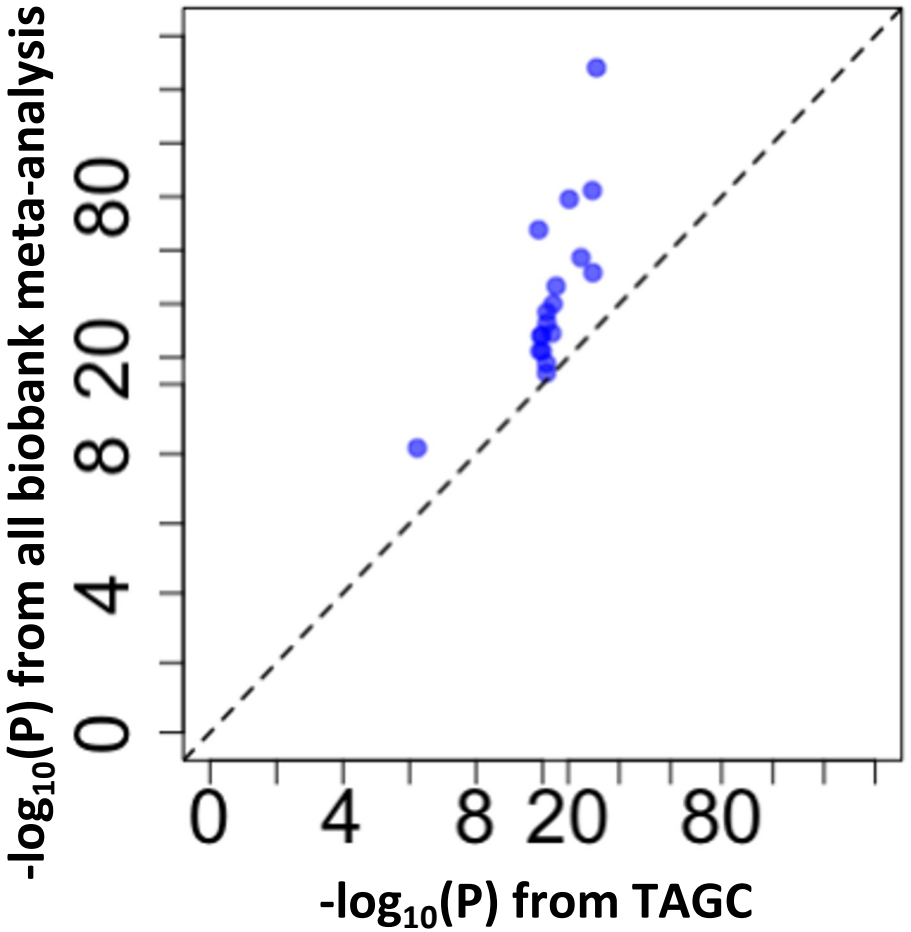
Supplementary Figure 9. Scatter plots of the effect size estimates in population-based biobanks and hospital-/healthcare-based biobanks with the Deming regression lines and the slope estimates (intercepts were fixed to 0). Loci that were genome-wide significant in all-biobank meta-analyses and have p-value < $1x10^{-6}$ in both meta-analyses of population-based biobanks and hospital-/healthcare-based biobanks, respectively, were included in the analyses. Endpoints that have more than 5 loci included in the analyses were plotted.



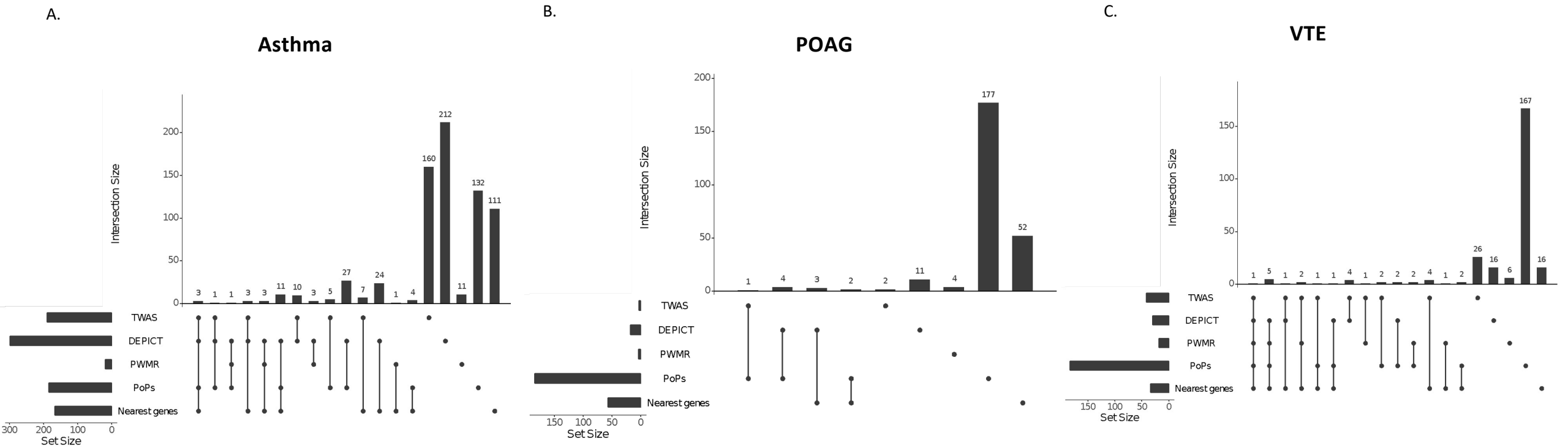
Supplementary Figure 10. Genetic correlation between each biobank and leave-on-biobank meta-analysis in GBMI



Supplementary Figure 11. All 18 loci identified by previous GWAS for Asthma (Demenais et al., 2018) have more significant p-values in all-biobank meta-analysis (**Supplementary Table 12**)

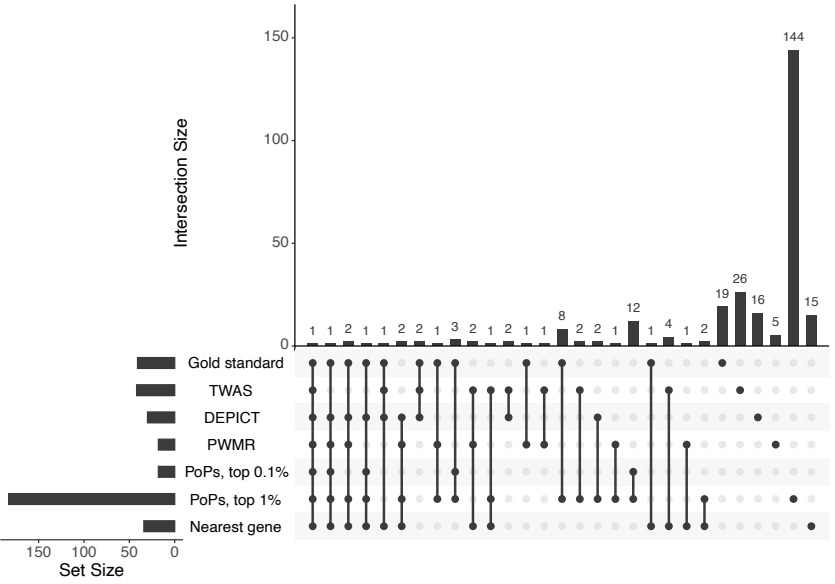


Supplementary Figure 12. Number of genes prioritized by different methods: PoPs (top 1%), DEPICT (FDR < 0.05), TWAS (P < 2.5 x 10 -6), PWMR (P < 0.001, Colocalization probability > 0.7), nearest genes around the top hits (Nearest gene, for intergenic variants, the nearest gene on each side will be included if both are located within 50kb from the top hit).

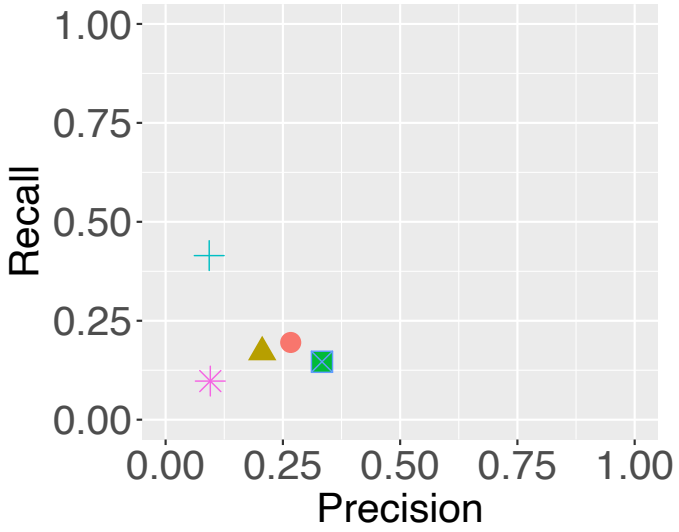


Supplementary Figure 13. For VTE, A. number of genes prioritized by different methods: PoPs (top 1% and top 0.1%), DEPICT (FDR < 0.05), TWAS (P < 2.5 x 10 -6), PWMR (P < 0.001, Colocalization probability > 0.7), nearest genes around the top hits (Nearest gene, for intergenic variants, the nearest gene on each side will be included if both are located within 50kb from the top hit). B. precision and recall of the different gene prioritization methods based on a gold standard set of 41 VTE genes that was curated prior to the meta-analysis by medical and molecular genetics experts in VTE (Wolford et al. 2021). C. . precision and recall of the different gene prioritization methods based on 13 genes (*ADAMTS13,F10,F2,F5,F7,FGA,FGB,FGG,PROC,PROS1,PROZ,THBD,VWF*) in the gold standard sets that fall within 1Mb around VTE top hits in GBMI meta-analysis. The prioritization gene lists are presented in **Supplementary Table 24.**

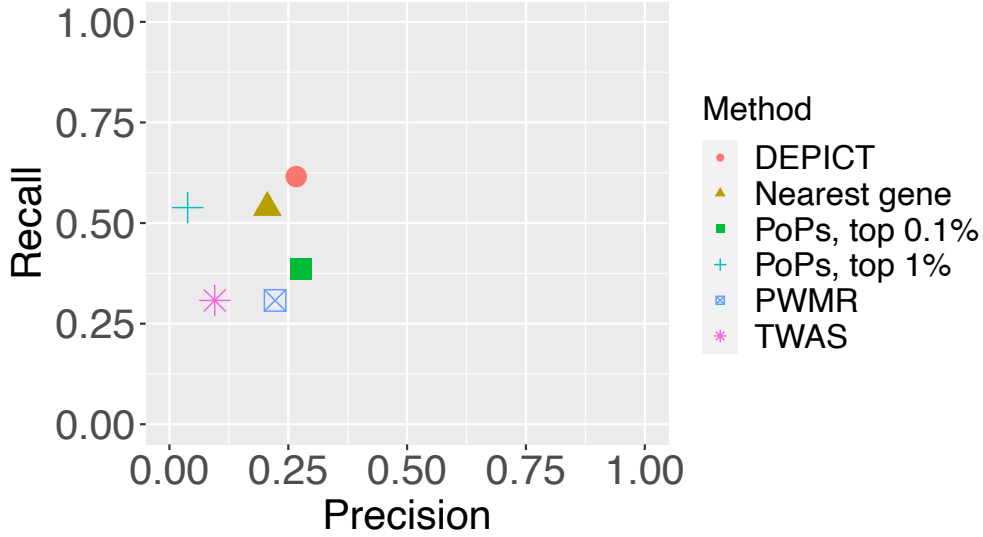
A.



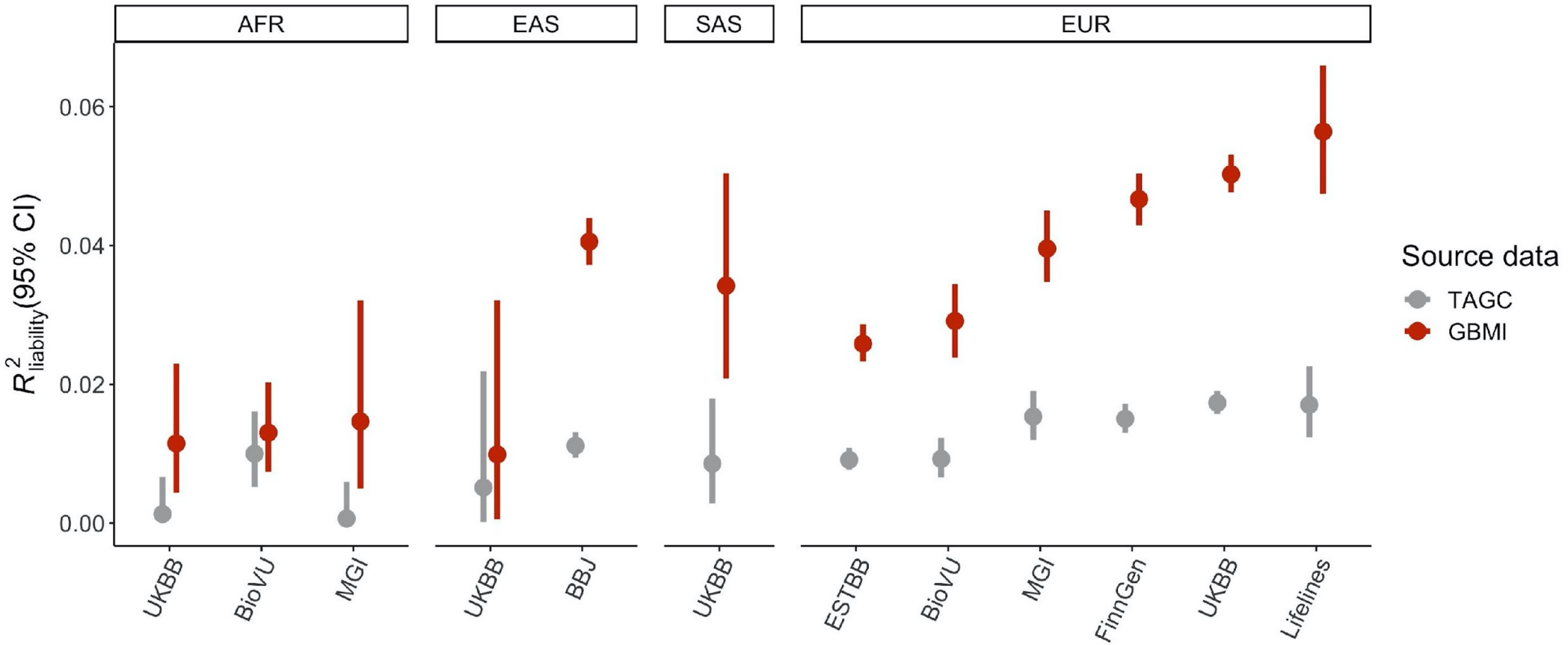
B.



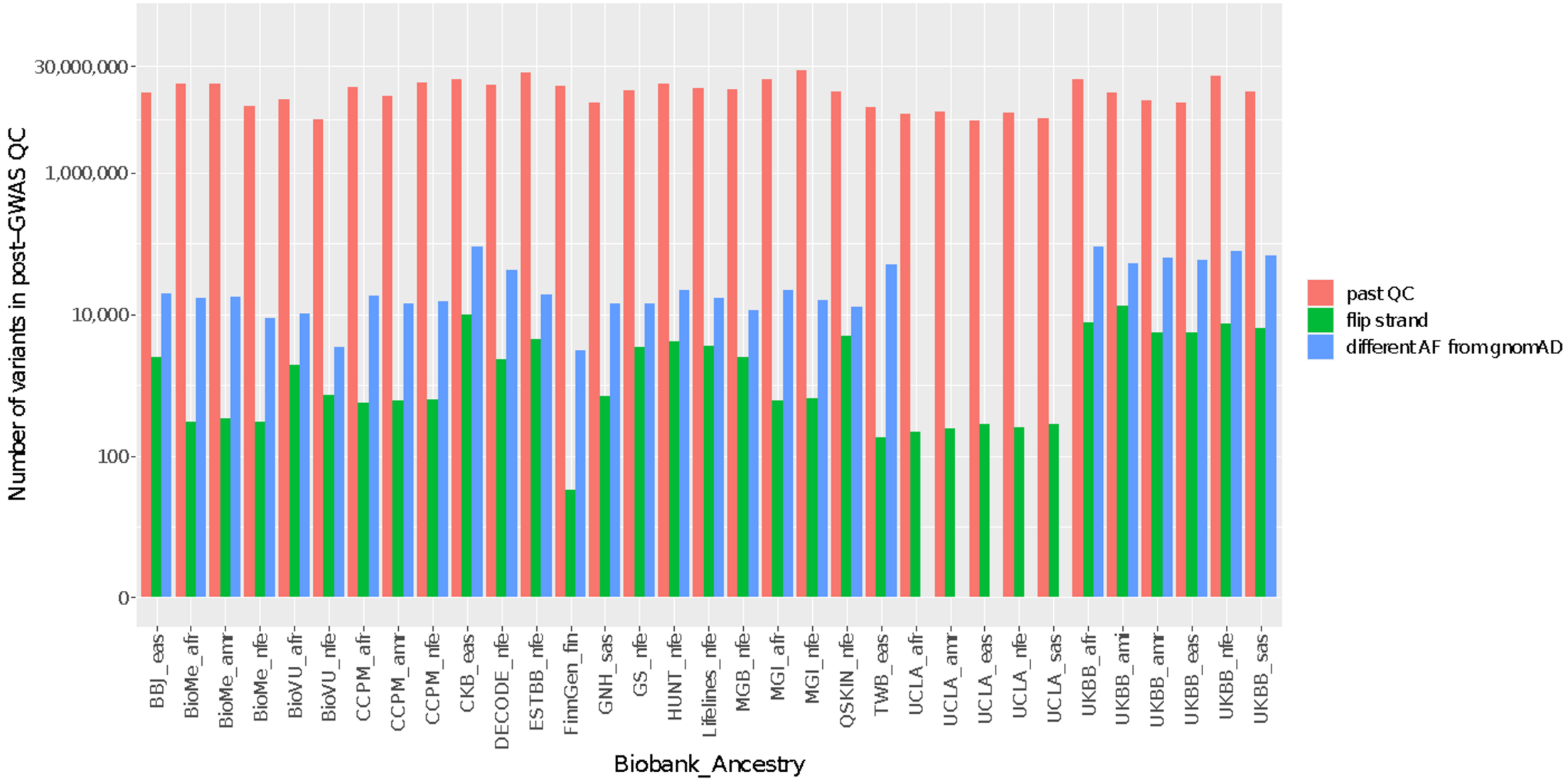
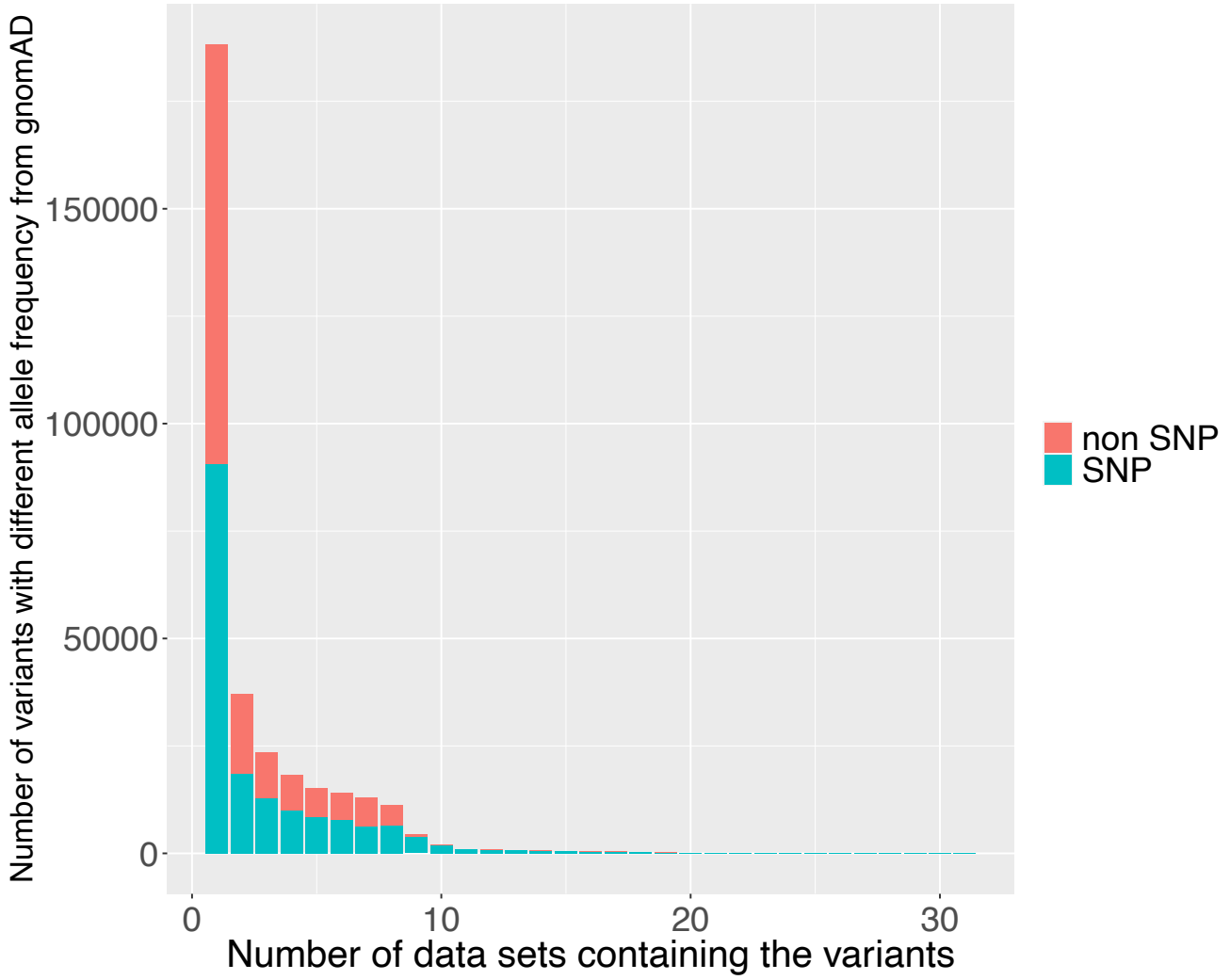
C.



Supplementary Figure 14. Improved polygenic risk scores (PRS) prediction accuracy using GBMI meta-analysis results compared to TAGC summary statistics.



Supplementary Figure 15. Palindromic SNPs with potential strand flip and genetic variants with different allele frequencies compared to gnomAD were flagged when included in the meta-analyses



Supplementary Figure 16. The distribution of number of biobanks that contain the genetic variants with different allele frequencies compared to gnomAD.