The American Journal of Human Genetics, Volume *104*

Supplemental Data

Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct

Manuel A.R. Ferreira, Riddhima Mathur, Judith M. Vonk, Agnieszka Szwajda, Ben Brumpton, Raquel Granell, Bronwyn K. Brew, Vilhelmina Ullemar, Yi Lu, Yunxuan Jiang, 23andMe Research Team, eQTLGen Consortium, BIOS Consortium, Patrik K.E. Magnusson, Robert Karlsson, David A. Hinds, Lavinia Paternoster, Gerard H. Koppelman, and Catarina Almqvist

Genetic architectures of childhood- and adult-onset asthma are partly distinct

SUPPLEMENTAL DATA

SUPPLEMENTAL FIGURES

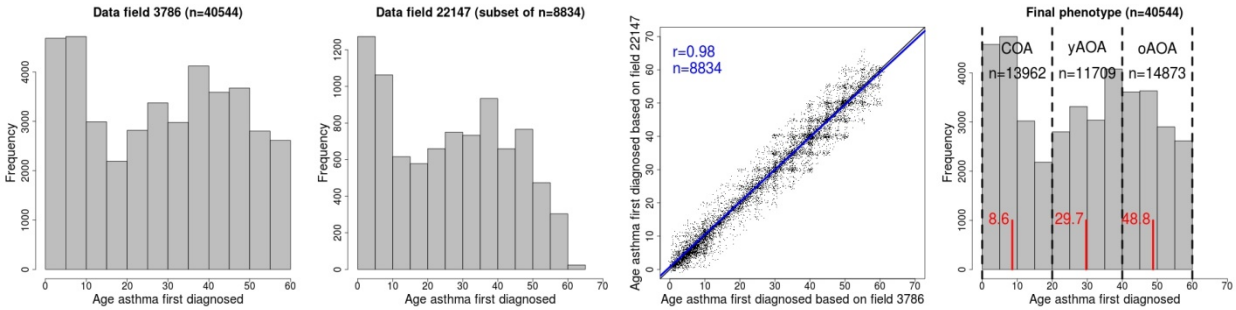

Figure S1. UK Biobank data fields used to classify individuals with COA or AOA.

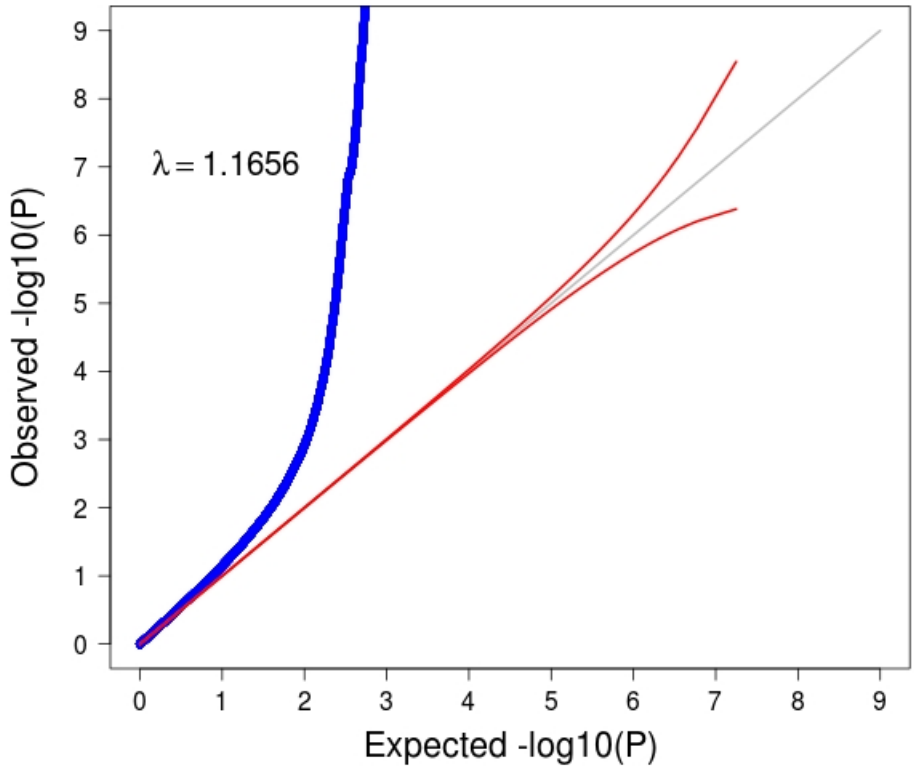

Figure S2. Distribution of the observed and expected association P-values for the GWAS of COA in the UK Biobank study. The genomic inflation factor (lambda, estimated as the median chi-square divided by 0.4549) is also shown. The intercept of LD score regression was 1.039.

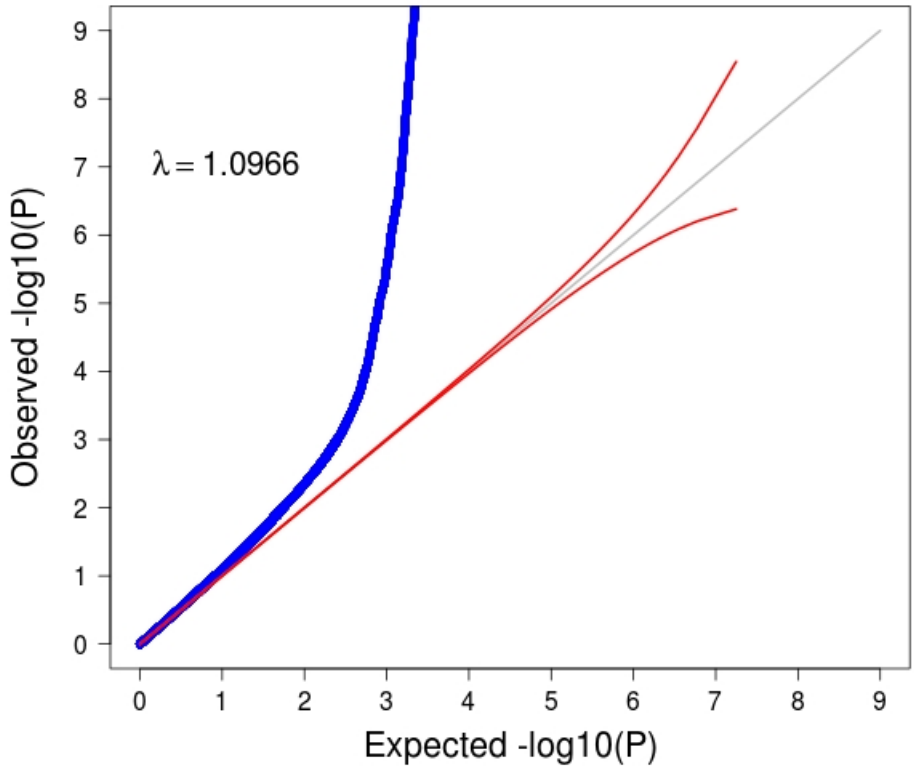

Figure S3. Distribution of the observed and expected association P-values for the GWAS of AOA in the UK Biobank study. The genomic inflation factor (lambda, estimated as the median chi-square divided by 0.4549) is also shown. The intercept of LD score regression was 1.018.

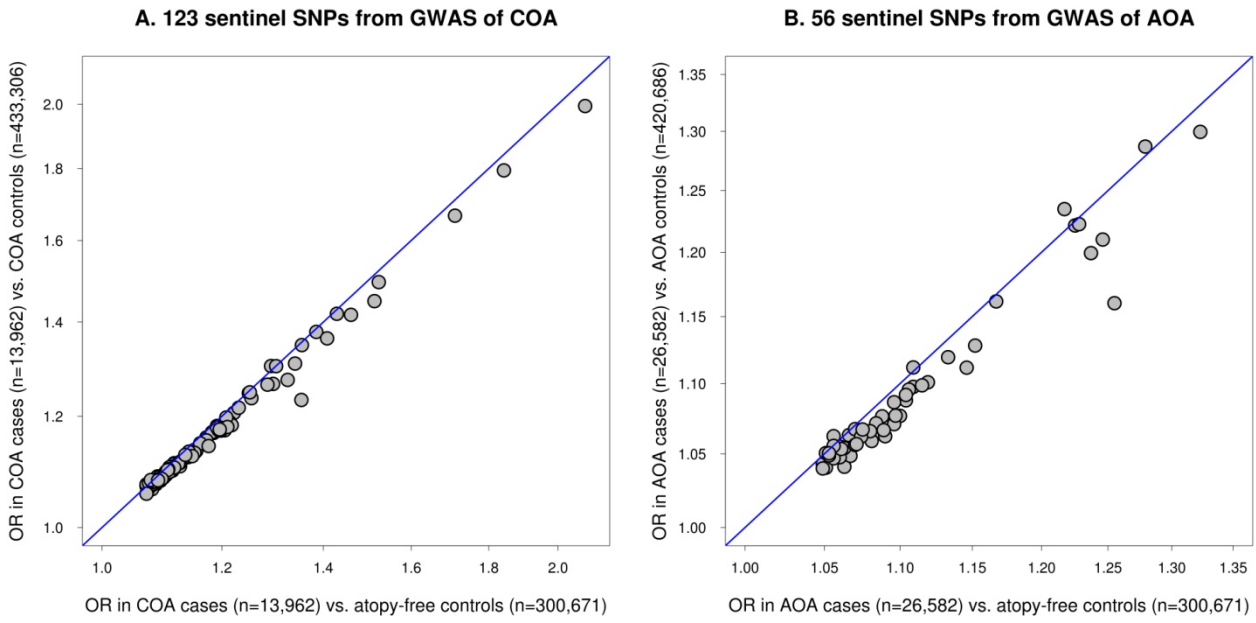

Figure S4. SNP associations when individuals with other allergies were not excluded from the control group used in the GWAS. The association (odds ratio) between the sentinel variants and asthma risk was consistent when comparing results from the analysis described in the main manuscript (*i.e.* asthma cases vs. atopy-free controls; *x*-axis) and an analysis that did not exclude individuals with other allergies (*e.g.* hay fever or eczema; *y*-axis) from the control group.

SUPPLEMENTAL METHODS

The Avalon Longitudinal Study of Parent and Children (ALSPAC)

Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled is 14,541 (for these at least one questionnaire has been returned or a "Children in Focus" clinic had been attended by 19/07/99). Of these initial pregnancies, there was a total of 14,676 foetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above. The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented on the built files and reflecting enrolment status at the age of 24 is 904 (452, 254 and 198 recruited during Phases II, III and IV respectively), resulting in an additional 811 children being enrolled. The phases of enrolment are described in more detail in the cohort profile paper (see footnote 4 below). Please note that phase 4 enrolment (age 18-24) is not currently included in the cohort profile. The total sample size for analyses using any data collected after the age of seven is therefore 15,247 pregnancies, resulting in 15,458 foetuses. Of this total sample of 15,656 foetuses, 14,973 were live births and 14,899 were alive at 1 year of age. A 10% sample of the ALSPAC cohort, known as the Children in Focus (CiF) group, attended clinics at the University of Bristol at various time intervals between 4 to 61 months of age. The CiF group were chosen at random from the last 6 months of ALSPAC births (1432 families attended at least one clinic). Excluded were those mothers who had moved out of the area or were lost to follow-up, and those partaking in another study of infant development in Avon. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/).

Child and Adolescent Twin Study in Sweden (CATSS)

The CATSS is an ongoing longitudinal twin study targeting all twins born from 1992 and living in Sweden. Since 2004, twins are invited to participate in CATSS following their ninth birthday. During the first three years of data collection, twelve-year-old twins were also invited. Participation in CATSS starts with a parental telephone interview on the children's health, perinatal factors, living situation. A module including questions regarding the twin pair's physical similarities is the basis for an algorithm-based assessment of zygosity. Since 2008, twins have also been offered a DNA-based zygosity test using the saliva samples collected by mail in connection with invitation to the study. DNA from saliva is then stored in the biobank of Karolinska Institutet. To date, approximately n=29,100 twins have participated in CATSS-9/12. DNA was extracted from saliva samples using either the Chemagic STAR instrument from Hamilton Robotics, with magnet bead purification kits from Chemagen, or the Puregene extraction kit (Gentra systems, Minneapolis, USA). Genotyping was performed in 18 batches at the SNP&SEQ Technology Platform in Uppsala, Sweden, using the Illumina PsychArray bead chip. Genotype calls from the zCall algorithm for rare variants were combined with those from the Illumina GenCall algorithm to increase sensitivity at low minor allele frequencies. After initial intensity-level quality control, 18,193 samples remained. Additional QC filtering was applied as follows: SNPs with missingness > 2%, SNPs with more than 10% discordant genotypes across replicates or MZ pairs, SNPs out of Hardy-Weinberg equilibrium (exact test P-value < 10^{-6}), SNPs with clear batch effects or absolute MAF difference from 1000 Genomes European samples > 10%, Y-chromosome and mitochondrial SNPs, and SNPs with minor allele count <= 1 were all excluded. Further, individuals with missingness > 2%, individuals with deviant autosomal heterozygosity (autosomal inbreeding coefficient F outside [-0.02, 0.02]), individuals showing excessive mean relatedness to the rest of the sample (mean relatedness > 6 s.d. above the sample mean), individuals where genotype-based sex did not match phenotype information, and individuals identified as non-European ancestral outliers were excluded. Non-genotyped monozygous twins in the study were imputed from their genotyped twin, resulting in a total of 21,752 individuals with genotypes. Genotypes were imputed to the 1000 genomes phase 3

version 5 reference panel using Shapeit v2.r790 and Minimac version 1.0.13.

23andMe

All research participants included in the analyses provided informed consent and answered surveys online according to our human subjects protocol, which was reviewed and approved by Ethical & Independent Review Services, a private institutional review board (http://www.eandireview.com). We restricted participants to a set of individuals who have >97% European ancestry, as determined through an analysis of local ancestry. Briefly, our algorithm first partitions phased genomic data into short windows of about 100 SNPs. Within each window, we use a support vector machine (SVM) to classify individual haplotypes into one of 31 reference populations. The SVM classifications are then fed into a hidden Markov model (HMM) that accounts for switch errors and incorrect assignments, and gives probabilities for each reference population in each window. Finally, we used simulated admixed individuals to recalibrate the HMM probabilities so that the reported assignments are consistent with the simulated admixture proportions. The reference population data is derived from public datasets (the Human Genome Diversity Project, HapMap, and 1000 Genomes), as well as 23andMe customers who have reported having four grandparents from the same country. A maximal set of unrelated individuals was chosen for each analysis using a segmental identity-by-descent (IBD) estimation algorithm. Individuals were defined as related if they shared more than 700 cM IBD, including regions where the two individuals share either one or both genomic segments identical-by-descent. This level of relatedness (roughly 20% of the genome) corresponds approximately to the minimal expected sharing between first cousins in an outbred population. DNA extraction and genotyping were performed on saliva samples by CLIA-certified and CAP-accredited clinical laboratories of Laboratory Corporation of America. Samples have been genotyped on one of four genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550+ BeadChip, including about 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The V3 platform was based on the Illumina OmniExpress+ BeadChip, with custom content to improve the overlap with our V2 array, with a total of about

950,000 SNPs. The V4 platform in current use is a fully custom array, including a lower redundancy subset of V2 and V3 SNPs with additional coverage of lower-frequency coding variation, and about 570,000 SNPs. Samples that failed to reach 98.5% call rate were re-analyzed. Individuals whose analyses failed repeatedly were re-contacted by 23andMe customer service to provide additional samples, as is done for all 23andMe customers.

Participant genotype data were imputed against the March 2012 "v3" release of 1000 Genomes reference haplotypes. We phased and imputed data for each genotyping platform separately. First, we used Beagle (version 3.3.1) to phase batches of 8000-9000 individuals across chromosomal segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs. We excluded SNPs with Hardy-Weinberg equilibrium $P<10^{-20}$, call rate < 95%, or with large allele frequency discrepancies compared to European 1000 Genomes reference data. Frequency discrepancies were identified by computing a 2x2 table of allele counts for European 1000 Genomes samples and 2000 randomly sampled 23andMe customers with European ancestry, and identifying SNPs with a chi squared $P<10^{-15}$. We imputed each phased segment against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac2, using 5 rounds and 200 states for parameter estimation. For the non-pseudoautosomal region of the X chromosome, males and females were phased together in segments, treating the males as already phased; the pseudoautosomal regions were phased separately. We then imputed males and females together using minimac, as with the autosomes, treating males as homozygous pseudo-diploids for the non-pseudoautosomal region.

eQTLGen Consortium – Author information

Mawussé Agbessi[1], Habibul Ahsan[2], Isabel Alves[1], Anand Andiappan[3], Wibowo Arindrarto[4], Philip Awadalla[1], Alexis Battle[5,6], Frank Beutner[7], Marc Jan Bonder[8,9,10], Dorret Boomsma[11], Mark Christiansen[12], Annique Claringbould[8], Patrick Deelen[8,13], Tõnu Esko[14], Marie-Julie Favé[1], Lude Franke[8], Timothy Frayling[15], Sina A. Gharib[16,12], Gregory Gibson[17], Bastiaan T. Heijmans[4], Gibran Hemani[18], Rick Jansen[19], Mika Kähönen[20,21], Anette Kalnapenkis[14], Silva Kasela[14], Johannes Kettunen[22], Yungil Kim[6,23], Holger Kirsten[24], Peter Kovacs[25], Knut Krohn[26], Jaanika Kronberg-Guzman[14], Viktorija Kukushkina[14], Zoltan Kutalik[27], Bernett Lee[3], Terho Lehtimäki[28,29], Markus Loeffler[24], Urko M. Marigorta[17], Hailang Mei[4], Lili Milani[14], Grant W. Montgomery[30], Martina Müller-Nurasyid[31,32,33], Matthias Nauck[34], Michel Nivard[11], Brenda Penninx[19], Markus Perola[35], Natalia Pervjakova[14], Brandon L. Pierce[2], Joseph Powell[36], Holger Prokisch[37,38], Bruce M. Psaty[12,39,40], Olli T. Raitakari[41,42], Samuli Ripatti[43], Olaf Rotzschke[3], Sina Rüeger[27], Ashis Saha[6], Markus Scholz[24], Katharina Schramm[31,32], Ilkka Seppälä[28,29], Eline P. Slagboom[4], Coen D.A. Stehouwer[44], Michael Stumvoll[45], Patrick Sullivan[46], Peter-Bram 't Hoen[47], Alexander Teumer[48], Joachim Thiery[49], Lin Tong[2], Anke Tönjes[45], Jenny van Dongen[11], Maarten van Iterson[4], Joyce van Meurs[50], Jan H. Veldink[51], Joost Verlouw[50], Peter M. Visscher[30], Uwe Völker[52], Urmo Võsa[8,14], Harm-Jan Westra[8], Cisca Wijmenga[8], Hanieh Yaghootkar[15], Jian Yang[30,53], Biao Zeng[17], Futao Zhang[30]

Author list is ordered alphabetically.

1. Computational Biology, Ontario Institute for Cancer Research, Toronto, Canada
2. Department of Public Health Sciences, University of Chicago, Chicago, United States of America
3. Singapore Immunology Network, Agency for Science, Technology and Research, Singapore, Singapore

4. Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
5. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, United States of America
6. Department of Computer Science, Johns Hopkins University, Baltimore, United States of America
7. Heart Center Leipzig, Universität Leipzig, Leipzig, Germany
8. Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands
9. European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, United Kingdom
10. Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany
11. Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
12. Cardiovascular Health Research Unit, University of Washington, Seattle, United States of America
13. Genomics Coordination Center, University Medical Centre Groningen, Groningen, The Netherlands
14. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia
15. Exeter Medical School, University of Exeter, Exeter, United Kingdom
16. Department of Medicine, University of Washington, Seattle, United States of America
17. School of Biological Sciences, Georgia Tech, Atlanta, United States of America
18. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
19. Department of Psychiatry and Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands
20. Department of Clinical Physiology, Tampere University, Tampere, Finland
21. Department of Clinical Physiology, Finnish Cardiovascular Research Center – Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland

22. Centre for Life Course Health Research, University of Oulu, Oulu, Finland
23. Genetics and Genomic Science Department, Icahn School of Medicine at Mount Sinai, New York, United States of America
24. Institut für Medizinische Informatik, Statistik und Epidemiologie, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
25. IFB Adiposity Diseases, Universität Leipzig, Leipzig, Germany
26. Interdisciplinary Center for Clinical Research, Faculty of Medicine, Universität Leipzig, Leipzig, Germany
27. Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
28. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
29. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
30. Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
31. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
32. Department of Medicine I, University Hospital Munich, Ludwig Maximilian's University, München, Germany
33. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
34. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
35. National Institute for Health and Welfare, University of Helsinki, Helsinki, Finland
36. Garvan Institute of Medical Research, Garvan-Weizmann Centre for Cellular Genomics, Sydney, Australia
37. Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany

38. Institute of Human Genetics, Technical University Munich, Munich, Germany
39. Departments of Epidemiology, Medicine, and Health Services, University of Washington, Seattle, United States of America
40. Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States of America
41. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
42. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
43. Statistical and Translational Genetics, University of Helsinki, Helsinki, Finland
44. Department of Internal Medicine, Maastricht University Medical Centre, Maastricht, The Netherlands
45. Department of Medicine, Universität Leipzig, Leipzig, Germany
46. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
47. Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands
48. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
49. Institute for Laboratory Medicine, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
50. Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands
51. Department of Neurology, University Medical Center Utrecht, Utrecht, The Netherlands
52. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
53. Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China

BIOS Consortium (Biobank-based Integrative Omics Study)

Management Team Bastiaan T. Heijmans (chair)[1], Peter A.C. 't Hoen[2], Joyce van Meurs[3], Aaron Isaacs[4], Rick Jansen[5], Lude Franke[6].

Cohort collection Dorret I. Boomsma7, René Pool7, Jenny van Dongen7, Jouke J. Hottenga7 (Netherlands Twin Register); Marleen MJ van Greevenbroek[8], Coen D.A. Stehouwer[8], Carla J.H. van der Kallen[8], Casper G. Schalkwijk[8] (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga[6], Lude Franke[6], Sasha Zhernakova[6], Ettje F. Tigchelaar[6] (LifeLines Deep); P. Eline Slagboom[1], Marian Beekman[1], Joris Deelen[1], Diana van Heemst[9] (Leiden Longevity Study); Jan H. Veldink[10], Leonard H. van den Berg[10] (Prospective ALS Study Netherlands); Cornelia M. van Duijn[4], Bert A. Hofman[11], Aaron Isaacs[4], André G. Uitterlinden[3] (Rotterdam Study).

Data Generation Joyce van Meurs (Chair)[3], P. Mila Jhamai[3], Michael Verbiest[3], H. Eka D. Suchiman[1], Marijn Verkerk[3], Ruud van der Breggen[1], Jeroen van Rooij[3], Nico Lakenberg[1].

Data management and computational infrastructure Hailiang Mei (Chair)[12], Maarten van Iterson[1], Michiel van Galen[2], Jan Bot[13], Dasha V. Zhernakova[6], Rick Jansen[5], Peter van 't Hof[12], Patrick Deelen[6], Irene Nooren[13], Peter A.C. 't Hoen[2], Bastiaan T. Heijmans[1], Matthijs Moed[1].

Data Analysis Group Lude Franke (Co-Chair)[6], Martijn Vermaat[2], Dasha V. Zhernakova[6], René Luijk[1], Marc Jan Bonder[6], Maarten van Iterson[1], Patrick Deelen[6], Freerk van Dijk[14], Michiel van Galen[2], Wibowo Arindrarto[12], Szymon M. Kielbasa[15], Morris A. Swertz[14], Erik. W van Zwet[15], Rick Jansen[5], Peter-Bram 't Hoen (Co-Chair)[2], Bastiaan T. Heijmans (Co-Chair)[1].

1. Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden

University Medical Center, Leiden, The Netherlands

2. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

3. Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands

4. Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands

5. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

6. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

7. Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands

8. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands

9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

11. Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands

12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands

13. SURFsara, Amsterdam, the Netherlands

14. Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

15. Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

ACKNOWLEDGMENTS

23andMe: We would like to thank the research participants and employees of 23andMe for making this work possible. We particularly thank the following members of the 23andMe Research Team: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, Barry Hicks, Karen E. Huber, Ethan M. Jewett, Aaron Kleinman, Keng-Han Lin, Nadia K. Litterman, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Xin Wang, Catherine H. Wilson.

CATSS: We acknowledge The Swedish Twin Registry for access to data. The Swedish Twin Registry is managed by Karolinska Institutet and receives funding through the Swedish Research Council under the grant no 2017-00641. Financial support was provided from the Swedish Research Council through the Swedish Initiative for Research on Microdata in the Social And Medical Sciences (SIMSAM) framework grant no 340-2013-5867, grants provided by the Stockholm County Council (ALF-projects), the Swedish Heart-Lung Foundation and the Swedish Asthma and Allergy Association's Research Foundation.

ALSPAC: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.