Genetic studies of birth weight give biological insights into links with adult disease

Momoko Horikoshi[1,2,*], Robin N Beaumont[3,*], Felix R Day[4,*], Nicole M Warrington[5,6,*], Marjolein N Kooijman[7,8,9,*], Juan Fernandez-Tajes[1,*], Bjarke Feenstra[10], Natalie R van Zuydam[1,2], Kyle J Gaulton[1,11], Niels Grarup[12], Jonathan P Bradfield[13], David P Strachan[14], Ruifang Li-Gao[15], Tarunveer S Ahluwalia[12,16,17], Eskil Kreiner-Møller[16], Rico Rueedi[18,19], Leo-Pekka Lyytikäinen[20,21], Diana L Cousminer[22,23,24], Ying Wu[25], Elisabeth Thiering[26,27], Carol A Wang[6], Christian T Have[12], Jouke-Jan Hottenga[28], Natalia Vilor-Tejedor[29,30,31], Peter K Joshi[32], Eileen Tai Hui Boh[33], Ioanna Ntalla[34,35], Niina Pitkänen[36], Anubha Mahajan[1], Elisabeth M van Leeuwen[8], Raimo Joro[37], Vasiliki Lagou[1,38,39], Michael Nodzenski[40], Louise A Diver[41], Krina T Zondervan[1,42], Mariona Bustamante[29,30,31,43], Pedro Marques-Vidal[44], Josep M Mercader[45], Amanda J Bennett[2], Nilufer Rahmioglu[1], Dale R Nyholt[46], Ronald Ching Wan Ma[47,48,49], Claudia Ha Ting Tam[47], Wing Hung Tam[50], CHARGE Consortium Hematology Working Group, Santhi K Ganesh[51], Frank JA van Rooij[8], Samuel E Jones[3], Po-Ru Loh[52,53], Katherine S Ruth[3], Marcus A Tuke[3], Jessica Tyrrell[3,54], Andrew R Wood[3], Hanieh Yaghootkar[3], Denise M Scholtens[40], Lavinia Paternoster[55,56], Inga Prokopenko[1,57], Peter Kovacs[58], Mustafa Atalay[37], Sara M Willems[8], Kalliope Panoutsopoulou[59], Xu Wang[33], Lisbeth Carstensen[10], Frank Geller[10], Katharina E Schraut[32], Mario Murcia[31,60], Catharina EM van Beijsterveldt[28], Gonneke Willemsen[28], Emil V R Appel[12], Cilius E Fonvig[12,61], Caecilie Trier[12,61], Carla MT Tiesler[26,27], Marie Standl[26], Zoltán Kutalik[19,62], Sílvia Bonas-Guarch[45], David M Hougaard[63,64], Friman Sánchez[45,65], David Torrents[45,66], Johannes Waage[16], Mads V Hollegaard[63,64,‡], Hugoline G de Haan[15], Frits R Rosendaal[15], Carolina Medina-Gomez[7,8,67], Susan M Ring[55,56], Gibran Hemani[55,56], George McMahon[56], Neil R Robertson[1,2], Christopher J Groves[2], Claudia Langenberg[4], Jian'an Luan[4], Robert A Scott[4], Jing Hua Zhao[4], Frank D Mentch[13], Scott M MacKenzie[41], Rebecca M Reynolds[68], William L Lowe Jr[69], Anke Tönjes[70], Michael Stumvoll[58,70], Virpi Lindi[37], Timo A Lakka[37,71,72], Cornelia M van Duijn[8], Wieland Kiess[73], Antje Körner[58,73], Thorkild IA Sørensen[55,56,74,75], Harri Niinikoski[76,77], Katja Pahkala[36,78], Olli T Raitakari[36,79], Eleftheria Zeggini[59], George V Dedoussis[35], Yik-Ying Teo[33,80,81], Seang-Mei Saw[33,82], Mads Melbye[10,83,84], Harry Campbell[32], James F Wilson[32,85], Martine Vrijheid[29,30,31], Eco JCN de Geus[28,86], Dorret I Boomsma[28], Haja N Kadarmideen[87], Jens-Christian Holm[12,61], Torben Hansen[12], Sylvain Sebert[88,89], Andrew T Hattersley[3], Lawrence J Beilin[90], John P Newnham[6], Craig E Pennell[6], Joachim Heinrich[26,91], Linda S Adair[92], Judith B Borja[93,94], Karen L Mohlke[25], Johan G Eriksson[95,96,97], Elisabeth E Widén[22], Mika Kähönen[98,99], Jorma S Viikari[100,101], Terho Lehtimäki[20,21], Peter Vollenweider[44], Klaus Bønnelykke[16], Hans Bisgaard[16], Dennis O Mook-Kanamori[15,102,103], Albert Hofman[7,8], Fernando Rivadeneira[7,8,67], André G Uitterlinden[7,8,67], Charlotta Pisinger[104], Oluf Pedersen[12], Christine Power[105], Elina Hyppönen[105,106,107], Nicholas J Wareham[4], Hakon Hakonarson[13,23,108], Eleanor Davies[41], Brian R Walker[68], Vincent WV Jaddoe[7,8,9], Marjo-Riitta Jarvelin[88,89,109,110], Struan FA Grant[13,23,108,111], Allan A Vaag[83,112], Debbie A Lawlor[55,56], Timothy M Frayling[3], George Davey Smith[55,56], Andrew P Morris[1,113,114,§], Ken K Ong[4,115,§], Janine F Felix[7,8,9,§], Nicholas J Timpson[55,56,§], John RB Perry[4,§], David M Evans[5,55,56,§], Mark I McCarthy[1,2,116,§], Rachel M Freathy[3,55,§], on behalf of the Early Growth Genetics (EGG) Consortium

1. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
2. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK.
3. Institute of Biomedical and Clinical Science, University of Exeter Medical School, Royal Devon and Exeter Hospital, Exeter, UK.
4. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK.
5. The University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia.
6. School of Women's and Infants' Health, The University of Western Australia, Perth, Australia.

7. The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, the Netherlands.
8. Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, the Netherlands.
9. Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, the Netherlands.
10. Department of Epidemiology Research, Statens Serum Institute, Copenhagen, Denmark.
11. Department of Pediatrics, University of California San Diego, La Jolla, California, USA.
12. The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
13. Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.
14. Population Health Research Institute, St George's University of London, London, Cranmer Terrace, UK.
15. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.
16. COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.
17. Steno Diabetes Center, Gentofte, Denmark.
18. Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland.
19. Swiss Institute of Bioinformatics, Lausanne, Switzerland.
20. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland.
21. Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland.
22. Institute for Molecular Medicine, Finland (FIMM), University of Helsinki, Helsinki, Finland.
23. Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.
24. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
25. Department of Genetics, University of North Carolina, Chapel Hill, NC, USA.
26. Institute of Epidemiology I, Helmholtz Zentrum München- German Research Center for Environmental Health, Neuherberg, Germany.
27. Division of Metabolic and Nutritional Medicine, Dr. von Hauner Children's Hospital, University of Munich Medical Center, Munich, Germany.
28. Netherlands Twin Register, Department of Biological Psychology, VU University, Amsterdam, the Netherlands.
29. ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain.
30. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
31. CIBER de Epidemiología y Salud Pública (CIBERESP), Spain.
32. Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, UK.
33. Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore, Singapore.
34. William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
35. Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece.
36. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
37. Institute of Biomedicine, Physiology, University of Eastern Finland, Kuopio, Finland.
38. KUL – University of Leuven, Department of Neurosciences, Leuven, Belgium.
39. Translational Immunology Laboratory, VIB, Leuven, Belgium.

40. Department of Preventive Medicine, Division of Biostatistics, Feinberg School of Medicine, Northwestern University, Chicago, USA.
41. Institute of Cardiovascular & Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.
42. Endometriosis CaRe Centre, Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford, UK.
43. Center for Genomic Regulation (CRG), Barcelona, Spain.
44. Department of Internal Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland.
45. Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain.
46. Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland, Australia.
47. Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, Hong Kong, China.
48. Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, Hong Kong, China.
49. Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China.
50. Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, Hong Kong, China.
51. Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA.
52. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.
53. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.
54. European Centre for Environment and Human Health, University of Exeter, Truro, UK.
55. Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK.
56. School of Social and Community Medicine, University of Bristol, Bristol, UK.
57. Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK.
58. IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany.
59. Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK.
60. FISABIO–Universitat Jaume I–Universitat de València, Joint Research Unit of Epidemiology and Environmental Health, Valencia, Spain.
61. The Children's Obesity Clinic, Department of Pediatrics, Copenhagen University Hospital Holbæk, Holbæk, Denmark.
62. Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland.
63. Danish Center for Neonatal Screening, Statens Serum Institute, Copenhagen, Denmark.
64. Department for Congenital Disorders, Statens Serum Institute, Copenhagen, Denmark.
65. Computer Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain.
66. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
67. Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, the Netherlands.
68. BHF Centre for Cardiovascular Science, University of Edinburgh, Queen's Medical Research Institute, Edinburgh, Scotland, UK.
69. Department of Medicine, Division of Endocrinology, Metabolism, and Molecular Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA.
70. Medical Department, University of Leipzig, Leipzig, Germany.

71. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland.
72. Kuopio Research Institute of Exercise Medicine, Kuopio, Finland.
73. Pediatric Research Center, Department of Women´s & Child Health, University of Leipzig, Leipzig, Germany.
74. Novo Nordisk Foundation Center for Basic Metabolic Research and Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
75. Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark.
76. Department of Pediatrics, Turku University Hospital, Turku, Finland.
77. Department of Physiology, University of Turku, Turku, Finland.
78. Paavo Nurmi Centre, Sports and Exercise Medicine Unit, Department of Physical Activity and Health, Turku, Finland.
79. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.
80. Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore.
81. Life Sciences Institute, National University of Singapore, Singapore, Singapore.
82. Singapore Eye Research Institute, Singapore, Singapore.
83. Department of Clinical Medicine, Copenhagen University, Copenhagen, Denmark.
84. Department of Medicine, Stanford School of Medicine, Stanford, California, USA.
85. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland, UK.
86. EMGO Institute for Health and Care Research, VU University and VU University Medical Center, Amsterdam, the Netherlands.
87. Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
88. Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland.
89. Biocenter Oulu, University of Oulu, Finland.
90. School of Medicine and Pharmacology, Royal Perth Hospital Unit, The University of Western Australia, Perth, Australia.
91. Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital Munich, Ludwig Maximilian University of Munich, Munich, Germany.
92. Department of Nutrition, University of North Carolina, Chapel Hill, NC, USA.
93. USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines.
94. Department of Nutrition and Dietetics, University of San Carlos, Cebu City, Philippines.
95. National Institute for Health and Welfare, Helsinki, Finland.
96. Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.
97. Folkhälsan Research Center, Helsinki, Finland.
98. Department of Clinical Physiology, Tampere University Hopital, Tampere, Finland.
99. Department of Clinical Physiology, University of Tampere School of Medicine, Tampere, Finland.
100. Division of Medicine, Turku University Hospital, Turku, Finland.
101. Department of Medicine, University of Turku, Turku, Finland.
102. Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands.
103. Epidemiology Section, BESC Department, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia.

104. Research Center for Prevention and Health Capital Region, Center for Sundhed, Rigshospitalet – Glostrup, Copenhagen University, Glostrup, Denmark.
105. Population, Policy and Practice, UCL Institute of Child Health, University College London, London, UK.
106. Centre for Population Health Research, School of Health Sciences, and Sansom Institute, University of South Australia, Adelaide, Australia.
107. South Australian Health and Medical Research Institute, Adelaide, Australia.
108. Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
109. Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK.
110. Unit of Primary Care, Oulu University Hospital, Oulu, Finland.
111. Division of Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.
112. Department of Endocrinology, Rigshospitalet, Copenhagen, Denmark.
113. Department of Biostatistics, University of Liverpool, Liverpool, UK.
114. Estonian Genome Center, University of Tartu, Tartu, Estonia.
115. Department of Paediatrics, University of Cambridge, Cambridge, UK.
116. Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Oxford, UK.

[*]These authors contributed equally to this work.
[§]These authors jointly directed this work.
[‡]Deceased.

Birth weight (BW) is influenced by both foetal and maternal factors and in observational studies is reproducibly associated with future risk of adult metabolic diseases including type 2 diabetes (T2D) and cardiovascular disease[1]. These lifecourse associations have often been attributed to the impact of an adverse early life environment. We performed a multi-ancestry genome-wide association study (GWAS) meta-analysis of BW in 153,781 individuals, identifying 60 loci where foetal genotype was associated with BW ($P<5x10^{-8}$). Overall, ~15% of variance in BW could be captured by assays of foetal genetic variation. Using genetic association alone, we found a strong inverse genetic correlation between BW and systolic blood pressure (r_g=-0.22, $P=5.5x10^{-13}$), T2D (r_g=-0.27, $P=1.1x10^{-6}$) and coronary artery disease (r_g=-0.30, $P=6.5x10^{-9}$) and, in large cohort data sets, that genetic factors were the major contributor to the negative covariance between BW and future cardiometabolic risk. Pathway analyses indicated that the protein products of genes within BW-associated regions were enriched for diverse processes including insulin signalling, glucose homeostasis, glycogen biosynthesis and chromatin remodelling. There was also enrichment of associations with BW in known imprinted regions ($P=1.9x10^{-4}$). We have demonstrated that lifecourse associations between early growth phenotypes and adult cardiometabolic disease are in part the result of shared genetic effects and have highlighted some of the pathways through which these causal genetic effects are mediated.

We combined GWAS data for BW in 153,781 individuals representing multiple ancestries from 37 studies across three components (**Extended Data 1, Supplementary Table 1**): (i) 75,891 individuals of European ancestry from 30 studies; (ii) 67,786 individuals of European ancestry from the UK Biobank; and (iii) 10,104 individuals of diverse ancestries (African American, Chinese, Filipino, Surinamese, Turkish and Moroccan) from six studies. Within each study, BW was z-score transformed separately in males and females after excluding non-singletons and premature births and adjusting for gestational age where available. Genotypes were imputed using reference panels from the 1000 Genomes (1000G)[2] or combined 1000G and UK10K Project[3] (**Supplementary Table 2**). We performed quality control assessments to confirm that the distribution of BW was consistent across studies, irrespective of the data collection protocol, and confirmed that self-reported BW in UK Biobank showed genetic and phenotypic associations consistent with those seen for measured BW in other studies[4] (**Methods**).

We identified 60 loci (59 autosomal) associated with BW at genome-wide significance ($P<5x10^{-8}$) in either the European ancestry or trans-ancestry meta-analyses (**Figure 1a, Extended Data 2, Supplementary Data; Methods**). At lead SNPs, we observed no heterogeneity in allelic effects between the three study components (Cochran's *Q* statistic *P*>0.00083) (**Supplementary Table 3**). Fifty-three of these loci were novel in that the lead SNP mapped >2Mb away from, and was statistically independent (EUR r^2<0.05) of, the seven previously-reported BW signals[5], all of which were confirmed in this larger analysis (**Supplementary Table 4**). Approximate conditional analysis in the European ancestry data indicated that three of these novel loci (near *ZBTB7B, HMGA1* and *PTCH1*) harboured multiple distinct association signals attaining genome-wide significance (**Methods**; **Supplementary Table 5, Extended Data 3**).

The lead variants for most signals mapped to non-coding sequence, and at only two loci, *ADRB1* (rs7076938; r^2=0.99 with *ADRB1* G389R) and *NRIP1* (rs2229742, R448G) did the association data point to likely causal non-synonymous coding variants (**Supplementary Table 6; Methods**). Lead SNPs for all but two loci (those mapping near *YKT6-GCK* and *SUZ12P1-CRLF3*) were common (minor allele frequency (MAF)≥5%) with individually modest effects on BW (β=0.020-0.053 per allele, equivalent to 10 to 26g). This was despite much improved coverage of low-frequency variants in this study (compared to previous HapMap 2 imputed meta-analyses[5]) reflecting imputation from larger, and more complete, reference panels (**Extended Data 4**). Indeed, all but five of the common variant association signals were tagged by variants (EUR r^2>0.6) in the HapMap 2 reference panel

(**Supplementary Tables 4 and 5**), indicating that most of the novel discovery in the present study was driven by increased sample size[5]. Fine-mapping analysis yielded 14 regions within which fewer than ten variants contributed to the locus-specific "credible set" accounting for >99% of the posterior probability of association (**Methods; Supplementary Table 7**). Collectively, these credible set variants showed enrichment for overlap with DNaseI hypersensitivity sites, particularly those generated by ENCODE, from foetal (4.2-fold, 95% CI [1.8-10.7]) and neonatal tissues (4.9 [1.8-11.0]) (**Supplementary Table 8; Methods**). The greatest refinement was at *YKT6-GCK*, where the credible set included only the low frequency variant rs138715366, which maps intronic to *YKT6.*

In combination, the 62 distinct genome-wide significant signals at the 59 autosomal loci, explained 2.0% (standard error (SE) 1.1%) of variance in BW (**Supplementary Table 9; Methods**), similar in magnitude to that attributable to sex or maternal body mass index (BMI)[5]. However, the variance in BW captured collectively by all autosomal genotyped variants on the array was considerably larger, estimated at 15.1% (SE=0.9) in UK Biobank (**Methods**). These figures are consistent with a long tail of genetic variants of smaller effects contributing to variation in BW.

Associations between foetal genotype and BW could result from indirect effects of the maternal genotype (*r* with foetal genotype ≈ 0.5) influencing BW via the intrauterine environment. However, two lines of evidence indicated that variation in the foetal genome was the predominant driver of the BW associations. First, an analysis of the global contribution of maternal vs. foetal genetic variation, using a maternal-GCTA model[6] in 4,382 mother-child pairs, estimated that the child's genotype (σ_C^2=0.24, SE=0.11) makes a larger contribution to BW variance than either the mother's genotype (σ_M^2=0.04, SE=0.10), or the covariance between the two (σ_{CM}=0.04, SE=0.08). Second, when we compared, at each of the 60 loci, the point estimates of the BW effect size dependent on maternal genotype (as measured in up to 68,254 women[7]) with those dependent on foetal genotype (using European ancestry data from 143,677 individuals in the present study), foetal variation had greater impact than maternal at 93% of loci (55/60; binomial P=1x10^{-11}) (**Supplementary Table 10, Extended Data 5 and 6**; **Methods**). Notably, the association at *MTNR1B,* one of the five loci at which the maternal effect size exceeded the foetal (maternal: β=0.048, P=5.1x10^{-15}; foetal: β=0.023, P=2.9x10^{-8}), is consistent with prior evidence pointing towards a maternal effect: *MTNR1B* variation is associated with hyperglycaemia and impaired glucose tolerance[8,9] and these in turn have been shown to be causally related to offspring BW[10,11]. Power to disentangle maternal and foetal contributions using analyses of foetal genotype conditional on maternal genotype was constrained by the limited sample size available (n=12,909 mother-child pairs) (**Supplementary Table 11**).

These analyses collectively provide compelling evidence that foetal genotype has a substantial impact on early growth, as measured by BW. We sought to use these genetic associations to understand the causal relationships between the observed BW-disease associations and to characterise the underlying processes responsible.

To quantify the shared genetic contribution to BW and other health-related traits, we estimated their genetic correlations using LD Score regression[12] (**Methods**). BW (in European ancestry samples) showed strong positive genetic correlations with anthropometric and obesity-related traits including birth length (r_g=0.81, P=2.0x10^{-44}), and in adults, height (r_g=0.41, P=4.8x10^{-52}), waist circumference (r_g=0.18, P=3.9x10^{-10}) and BMI (r_g=0.11, P=7.3x10^{-6}). In contrast, BW showed inverse genetic correlations with indicators of adverse metabolic and cardiovascular health including coronary artery disease (CAD r_g=-0.30, P=6.5x10^{-9}), systolic blood pressure (SBP r_g=-0.22, P=5.5x10^{-13}) and T2D (r_g=-0.27, P=1.1x10^{-6}) (**Figure 2, Supplementary Table 12**). These correlations between BW and adult cardiometabolic phenotypes are of similar magnitude, although directionally-opposite, to the reported genetic correlations between adult BMI and those same cardiometabolic outcomes[12]. These findings support observational associations between paternal T2D and lower BW[4], and

establish more generally that the observed lifecourse associations between early growth and adult disease, at least in part, reflect the impact of shared genetic variants that influence both directly. In an effort to estimate the extent of genetic contribution to these lifecourse associations, we first focused on data from UK Biobank (n=57,715). For many of the traits for which data were available, genetic variation significantly contributed to the lifecourse relationship between BW and adult phenotypes, and in some cases appeared to be the major source of covariance between the traits. For example, we estimated that 85% (95% CI=70%-99%) of the negative covariance between BW and SBP was explained by shared genetic associations captured by directly genotyped SNPs (**Supplementary Table 13**). For continuous cardiometabolic measures, including lipids and fasting glycaemia, for which measures are not currently available in UK Biobank, we turned to the Northern Finland Birth Cohort (n=5,009), and obtained similar results (**Supplementary Table 13**). However, these estimates are limited, not only by wide confidence intervals, but also by the assumption of a linear relationship between BW and each of the phenotypes and inability to explicitly model maternal genotypic effects. In other words, the negative genetic correlations between BW and cardiometabolic traits may not exclusively reflect genetic effects mediated directly through the offspring, but also effects mediated by maternal genotype acting indirectly via perturbation of the *in utero* environment. Nevertheless, these estimates indicate that a substantial proportion of the variance in cardiometabolic risk that covaries with BW can be attributed to the effects of common genetic variation.

To elucidate the biological pathways and processes underlying regulation of foetal growth, we first performed gene set enrichment analysis of our BW GWAS analysis using MAGENTA[13] (**Methods**). Twelve pathways reached study-wide significance (FDR<0.05), including pathways involved in metabolism (insulin signalling, glycogen biosynthesis, cholesterol biosynthesis), growth (IGF-signalling, growth hormone pathway) and development (chromatin remodelling) (**Extended Data 7a**). Similar pathways were detected in a complementary analysis where we interrogated empirical protein-protein interaction (PPI) data (collated within Inweb[14]) identifying 13 PPI network modules with marked (z-score>5) enrichment for BW-association scores (**Extended Data 7b and 8a,b**; **Methods**). The proteins within these modules were themselves enriched for diverse processes related to metabolism, growth and development (**Extended Data 8a,b**).

We also observed enrichment of BW association signals across the set of 77 imprinted genes defined by the Genotype-Tissue Expression (GTEx) project[15] (P=1.9 x 10^{-4}; **Extended Data 7a, Supplementary Table 14**). Such enrichment is consistent with the "parental conflict" hypothesis regarding the allocation of maternal resources to the foetus[16]. Although the role of imprinted genes in foetal growth is described in animal models and rare human disorders[17], our result is the first large-scale, systematic demonstration of their contribution to normal variation in BW. Of the 60 genome-wide significant loci, two (*INS-IGF2*, *RB1*) fall within (or near) imprinted regions (**Figure 1b**), with a noteworthy third signal at *DLK1* (previously foetal antigen-1; P=5.6x10^{-8}). Parent-of-origin specific analyses to further investigate these individual loci (comparing heterozygote vs. homozygote BW variance in 57,715 unrelated individuals and testing BW associations with paternal vs. maternal alleles in 4,908 mother-child pairs; see **Methods**) proved, despite these sample sizes, to be underpowered (**Extended Data 9, Supplementary Tables 15 and 16**).

Many of the genome-wide signals for BW detected here are also established genome-wide association signals for a wide variety of cardiometabolic traits (**Figure 3**). These include the BW signals near *CDKAL1, ADCY5, HHEX/IDE* and *ANK1* (also genome-wide significant for T2D), *NT5C2* and *ADRB1* (BP/CAD/BMI). We used two approaches to understand whether this pattern of adult trait association represented a generic property of BW-associated loci, or reflected heterogeneous mechanisms linking BW to adult disease.

First, we applied unsupervised hierarchical clustering (**Methods**) to the non-BW trait association statistics for the 60 significant BW loci**.** The resultant heatmap indicates the heterogeneity of between-locus effect-size across the range of adult traits (**Figure 3, Supplementary Table 17**). For example, it shows that the associations between BW-raising alleles and increased adult height concentrated amongst a subset of loci including *HHIP* and *GNA12*, and highlights particularly strong associations with lipid traits for variants at the *TRIB1* and *MAFB* loci.

Second, we constructed trait-specific "point-of-contact" (PoC) PPI networks from proteins represented in both the global BW PPI network and equivalent PPI networks generated for each of the adult traits (**Methods; Extended Data 8c-e**). We reasoned that these PoC PPI networks would be enriched for the specific proteins mediating the observed links between BW and adult traits. To highlight processes implicated in specific BW-trait associations, we overlaid these PoC PPI with the 50 pathways over-represented in the global BW PPI network. These analyses revealed, for example, that proteins in the Wnt canonical signalling pathway were only detected in the PoC PPI network for BP traits. We can use these PPI overlaps to highlight the specific transcripts within BW GWAS loci that are likely to mediate the mechanistic links. For example, the overlap between the Wnt signalling pathway and the PoC PPI network for the intersection of BW and BP-related traits implicates *FZD9* as the likely effector gene at the *MLXIPL* BW locus (**Extended Data 8d, Supplementary Table 6**).

We focused our more detailed investigation of the mechanistic links between early growth and adult traits on two phenotypic areas: arterial BP and T2D/glycaemia.

Across both the overall GWAS and specifically among the 60 significant BW loci, most BW-raising alleles were associated with reduced BP (**Figures 2, 3**): the strongest inverse associations were seen for the loci near *NT5C2, FES, NRIP1, EBF1* and *PTH1R*. However, we also observed locus-specific heterogeneity in the genetic relationships between BP and BW: the SBP-raising allele at *ADRB1*[18] is associated with higher, rather than lower, BW (**Extended Data 10a**). When we considered the reciprocal relationship, i.e. the effects on BW of BP-raising alleles at 30 reported loci for SBP[18,19], there was an excess of associations (5/30 with lower BW at $P<0.05$; $P=0.0026$; **Extended Data 10a**). To dissect maternal and foetal genotype effects at these loci, we tested the impact on BW of a risk score generated from 30 SBP SNPs, restricted to the untransmitted maternal haplotype score[20] in a set of 5,201 mother-child pairs. Analysis of these loci indicated that maternal genotype effects on intrauterine environment are likely to contribute to the inverse genetic correlation between SBP and BW (**Methods; Supplementary Table 18**), and was consistent with the results of a wider study of >30,000 women which demonstrated associations between a maternal genetic score for SBP (conditional on foetal genotype) and lower offspring BW[10].

The BP-raising allele with the largest BW-lowering effect maps to the *NT5C2* locus (index variant for BW, rs74233809, $r^2=0.98$ with index variant for BP, rs11191548[19]) and is also associated with lower adult BMI ($r^2=0.99$ with rs11191560[21]). The BW-lowering allele at rs74233809 is a proxy for a recently-described functional variant in the nearby *CYP17A1* gene ($r^2=0.92$ with rs138009835)[22]. The *CYP17A1* gene encodes the cytochrome P450c17α enzyme, CYP17[23], which catalyses key steps in steroidogenesis that determine the balance between mineralocorticoid, glucocorticoid and androgen synthesis. This variant is known to alter transcriptional efficiency *in vitro* and is associated with higher urinary tetrahydroaldosterone excretion[22]. *CYP17A1* is expressed in foetal adrenal glands and testes from early gestation[24] as well as in the placenta[25]. These data implicate variation in *CYP17A1* expression as a contributor to the observational association between low BW and adult hypertension[26].

When we examined 45 loci associated with CAD[27], the inverse genetic correlation between CAD and BW was concentrated amongst the five CAD loci with primary BP associations. This suggests that

genetic determinants of BP play a leading role in mediating the lifecourse associations between BW and CAD (**Extended Data 10b,e**).

The LD score regression analyses demonstrated overall inverse genetic correlation between lower BW and elevated risk of T2D (**Figure 2**)**.** However**,** the locus specific heatmap indicates a heterogeneous pattern across individual loci (**Figure 3**). To explore this further, we tested the 84 reported T2D loci[28] for association with BW. Some T2D risk alleles (such as at *ADCY5*, *CDKAL1* and *HHEX-IDE*) were strongly associated with lower BW, whilst others (e.g. *ANK1 and MTNR1B*) were associated with higher (**Extended Data 10c**). This was in contrast with the BW effects of 422 known height loci[29] (**Extended Data 10d**), which showed a strong positive correlation consistent with the overall genetic correlation between height and BW, indicating that the growth effects of many height loci start prenatally and persist into adulthood.

The contrasting associations of T2D risk-alleles with both higher and lower BW are likely to reflect the differential impacts across loci of variation in the maternal and foetal genomes. Observational data link paternal diabetes with lower offspring BW[4] indicating that the inheritance of T2D risk alleles by the foetus tends, in line with the LD score regression analysis, to reduce growth. These relationships are consistent with the precepts of the "foetal insulin hypothesis"[30] and reflect the potential for reduced insulin secretion and/or signalling to lead to both reduced foetal growth and, many decades later, enhanced predisposition to T2D. In line with this, the paternal transmitted haplotype score generated from the 84 T2D risk variants was associated with lower BW (P=0.045) in 5,201 mother-child pairs (**Methods**; **Supplementary Table 18**). In contrast, maternal diabetes is observationally associated with higher offspring BW[4], reflecting the impact of maternal hyperglycaemia to stimulate foetal insulin secretion. The contribution of genotype-dependent maternal hyperglycaemia to BW is in line with the observation that some T2D risk alleles (most notably at *MTNR1B*) are associated with higher BW[7], and with evidence, from a recent study, that a maternal genetic score for T2D (conditional on foetal genotype) was associated with higher offspring BW[10]. Thus, both maternal and foetal genetic effects connect BW to later T2D risk, albeit acting in opposing directions. When we categorised T2D loci using a classification of physiological function derived from their effects on related glycaemic and anthropometric traits[31], we found that T2D-risk alleles associated with lower BW were those typically characterised by reduced insulin processing and secretion without detectable changes in fasting glucose (the "Beta Cell" cluster in **Extended Data 10f**).

The *YTK6* signal at rs138715366 is notable, not only because the genetic data indicates that a single low-frequency non-coding variant is driving the association signal (see above) but because of the proximity of this signal to *GCK*. Rare coding variants in glucokinase are causal for a form of monogenic hyperglycaemia and lead to large reductions in BW when parental alleles are passed to their offspring[32]. In addition, common non-coding variants nearby are implicated in T2D-risk and fasting hyperglycaemia[33,34]. However, the latter variants are conditionally independent of rs138715366 (**Supplementary Table 19**) and show no comparable association with lower BW. Either rs138715366 acts through effector transcripts other than *GCK*, or the impact of the low-frequency SNP near *YKT6* on *GCK* expression involves tissue- and/or temporal-specific variation in regulatory impact.

In conclusion, we have identified 60 genetic loci associated with BW and used these to gain insights into the aetiology of foetal growth and into well-established, but until now poorly understood, lifecourse disease associations. The evidence that the relationship between early growth and later metabolic disease has an appreciable genetic component contrasts with, but is not necessarily incompatible with, the emphasis on adverse early environmental events highlighted by the Foetal Origins Hypothesis[1]. As we have shown, these genetic effects reflect variation in both the foetal and

the maternal genome: the impact of the latter on the offspring's predisposition to adult disease could be mediated, at least in part, through perturbation of the antenatal and early life environment. Future mechanistic and genetic studies should support reconciliation between these alternative, but complementary, explanations for the powerful lifecourse associations that exist between events in early life and predisposition to cardiometabolic disease several decades later.

REFERENCES

1. Barker DJ. The developmental origins of chronic adult disease. *Acta Paediatr Suppl* **93**, 26-33 (2004).
2. The 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, *et al*. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
3. The UK10K Project Consortium, Walter K, Min JL, Huang J, Crooks L, *et al*. The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82-90 (2015).
4. Tyrrell JS, Yaghootkar H, Freathy RM, Hattersley AT, Frayling TM. Parental diabetes and birthweight in 236 030 individuals in the UK biobank study. *Int J Epidemiol* **42**, 1714-1723 (2013).
5. Horikoshi M, Yaghootkar H, Mook-Kanamori DO, Sovio U, Taal HR, *et al*. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet* **45**, 76-82 (2013).
6. Eaves LJ, Pourcain BS, Smith GD, York TP, Evans DM. Resolving the effects of maternal and offspring genotype on dyadic outcomes in genome wide complex trait analysis ("M-GCTA"). *Behav Genet* **44**, 445-455 (2014).
7. Feenstra B, Beaumont RN, Cavadino A, Tyrrell J, McMahon G, *et al*. Maternal genome-wide association study identifies a fasting glucose variant associated with offspring birth weight. (in preparation) doi: http://dx.doi.org/10.1101/034207
8. Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, et al. Variants in MTNR1B influence fasting glucose levels. *Nat Genet* **41**, 77-81 (2009).
9. Kwak SH, Kim SH, Cho YM, Go MJ, Cho YS, et al. A genome-wide association study of gestational diabetes mellitus in Korean women. *Diabetes* **61**, 531-541 (2012).
10. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, *et al*. Genetic evidence for causal relationships between maternal obesity-related traits and birth weight. *JAMA* **315**, 1129-1140 (2016).
11. Landon MB, Spong CY, Thom E, Carpenter MW, Ramin SM, *et al*. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med* **361**, 1339-1348 (2009).
12. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015).
13. Segrè AV, DIAGRAM Consortium, MAGIC investigators, Groop L, Mootha VK, *et al*. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* **6**, e1001058 (2010).
14. Lage K, Karlberg EO, Størling ZM, Olason PI, Pedersen AG, *et al*. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* **25**, 309–316 (2007).
15. Baran Y, Subramaniam M, Biton A, Tukiainen T, Tsang EK, *et al*. The landscape of genomic imprinting across diverse adult human tissues. *Genome Res* **25**, 927-936 (2015).
16. Haig D and Westoby M. Parent-Specific Gene Expression and the Triploid Endosperm. *The American Naturalist* **134**, 147-155 (1989).
17. Peters J. The role of genomic imprinting in biology and disease: an expanding view. *Nat Rev Genet* **15**, 517-530 (2014).
18. Johnson T, Gaunt TR, Newhouse SJ, Padmanabhan S, Tomaszewski M, *et al*. Blood pressure loci identified with a gene-centric array. *Am J Hum Genet* **89**, 688-700 (2011).
19. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, *et al*. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-109 (2011).

20. Zhang G, Bacelis J, Lengyel C, Teramo K, Hallman M, *et al*. Assessing the Causal Relationship of Maternal Height on Birth Size and Gestational Age at Birth: A Mendelian Randomization Analysis. *PLoS Med* **12**, e1001865 (2015).
21. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
22. Diver LA, MacKenzie SM, Fraser R, McManus F, Freel EM, *et al*. Common Polymorphisms at the *CYP17A1* locus associate with steroid phenotype: support for blood pressure GWAS signals at this locus. *Hypertension* **67**, 724-732 (2016).
23. Picado-Leonard J, Miller WL. Cloning and sequence of the human gene for P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase): similarity with the gene for P450c21. *DNA* **6**, 439-448 (1987).
24. Pezzi V, Mathis JM, Rainey WE, Carr BR. Profiling transcript levels for steroidogenic enzymes in fetal tissues. *J Steroid Biochem Mol Biol* **87**, 181-189 (2003).
25. Escobar JC, Patel SS, Beshay VE, Suzuki T, Carr BR. The human placenta expresses CYP17 and generates androgens *de novo*. *J Clin Endocrinol Metab* **96**, 1385-1392 (2011).
26. Reynolds RM, Walker BR, Phillips DI, Dennison EM, Fraser R, *et al*. Programming of hypertension: associations of plasma aldosterone in adult men and women with birth weight, cortisol, and blood pressure. *Hypertension* **53**, 932-936 (2009).
27. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, *et al*. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25-33 (2013).
28. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, *et al*. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-244 (2014).
29. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, *et al*. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-1186 (2014).
30. Hattersley AT, Tooke JE. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* **353**, 1789-1792 (1999).
31. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, *et al*. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158-2171 (2014).
32. Hattersley AT, Beards F, Ballantyne E, Appleton M, Harvey R, *et al*. Mutations in the glucokinase gene of the fetus result in reduced birth weight. *Nat Genet* **19**, 268-270 (1998).
33. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, *et al*. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-990 (2012).
34. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, *et al*. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-116 (2010).

METHODS

Ethics statement. All human research was approved by the relevant institutional review boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

Study-level analyses. Within each study, BW was collected from a variety of sources, including measurements at birth by medical practitioners, obstetric records, medical registers, interviews with the mother and self-report as adults (**Supplementary Table 1**). BW was *z*-score transformed, separately in males and females. Individuals with extreme BW (>5 standard deviations (SD) from the sex-specific study mean), monozygotic or polyzygotic siblings, or preterm births (gestational age <37 weeks, where this information was available) were excluded from downstream association analyses (**Supplementary Table 1**).

Within each study, stringent quality control of the GWAS genotype scaffold was undertaken, prior to imputation (**Supplementary Table 2**). Each scaffold was then pre-phased and imputed[35,36] up to reference panels from the 1000 Genomes Project Consortium[2] or 1000G and UK10K Project Consortium[3] (**Supplementary Table 2**). Association of BW with each variant passing established GWAS quality control filters[37] was tested in a linear regression framework, under an additive model for the allelic effect, after adjustment for study-specific covariates, including gestational age, where available (**Supplementary Table 2**). Where necessary, population structure was accounted for by adjustment for axes of genetic variation from principal components analysis[38] and subsequent genomic control correction[39], or inclusion of a genetic relationship matrix in a mixed model[40] (**Supplementary Table 2**). We calculated the genomic control inflation factor (λ) in each study to confirm that study-level population structure was accounted for prior to meta-analysis.

Preparation, quality control and genetic analysis in UK Biobank samples. UK Biobank phenotype data were available for 502,655 participants[41]. All participants in the UK Biobank were asked to recall their BW, of which 279,971 did so at either the baseline or follow-up assessment visit. Of these, 7,686 participants reported being part of multiple births and were excluded from downstream analyses. Ancestry checks, based on self-reported ancestry, resulted in the exclusion of 8,998 additional participants reported not to be white European. Of those individuals reporting BW at baseline and follow-up assessments, 393 were excluded because the two reported values differed by more than 0.5 kg. For those reporting different values (≤0.5 kg) between baseline and follow-up, we took the baseline measure forward for downstream analyses. We then excluded 36,716 individuals reporting values <2.5 kg or >4.5 kg as implausible for live term births before 1970. In total 226,178 participants had data relating to BW that matched these inclusion criteria.

Genotype data from the May 2015 release were available for a subset of 152,249 participants from UK Biobank. In addition to the quality control metrics performed centrally by UK Biobank, we defined a subset of "white European" ancestry samples using a K-means (K=4) clustering approach based on the first four genetically determined principal components. A maximum of 67,786 individuals (40,425 females and 27,361 males) with genotype and valid BW measures were available for downstream analyses. We tested for association with BW, assuming an additive allelic effect, in a linear mixed model implemented in BOLT-LMM[42] to account for cryptic population structure and relatedness. Genotyping array was included as a binary covariate in all models. Total chip heritability (i.e. the variance explained by all autosomal polymorphic genotyped SNPs passing quality control) was calculated using Restricted Maximum Likelihood (REML) implemented in BOLT-LMM[42]. We additionally analysed the association between BW and directly genotyped SNPs on the X chromosome: for this analysis, we used 57,715 unrelated individuals with BW available and identified by UK Biobank as white British. We excluded SNPs with evidence of deviation from Hardy-Weinberg Equilibrium ($P<1\times10^{-6}$), MAF<0.01 or overall missing rate >0.015,

resulting in 19,423 SNPs for analysis in Plink v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/)[43], with the first five ancestry principal components as covariates.

In both the full UK Biobank sample and our refined sample, we observed that BW was associated with sex, year of birth and maternal smoking (P<0.0015, all in the expected directions), confirming more comprehensive previous validation of self-reported BW[4]. We additionally verified that BW associations with lead SNPs at seven established loci[5] based on self-report in UK Biobank were consistent with those previously published.

European ancestry meta-analysis. The European ancestry meta-analysis consisted of two components: (i) 75,891 individuals from 30 GWAS from Europe, USA and Australia; and (ii) 67,786 individuals of white European origin from UK Biobank. In the first component, we combined sex-specific BW association summary statistics across studies in a fixed-effects meta-analysis, implemented in GWAMA[44] and applied a second round of genomic control[39] (λ_{GC}= 1.001). Subsequently, we combined association summary statistics from this component with UK Biobank in a European ancestry fixed-effects meta-analysis, implemented in GWAMA[44]. Variants failing GWAS quality control filters in UK Biobank, reported in less than 50% of the total sample size in the first component, or with MAF<0.1%, were excluded from the European ancestry meta-analysis. We meta-analysed the X-Chromosome UK Biobank results (19,423 SNPs) with corresponding results from the European ancestry meta-analysis using fixed effects P-value based meta-analysis in METAL[45] (max N=99,152).

We were concerned that self-reported BW as adults in UK Biobank would not be comparable with that obtained from more stringent collection methods used in other European ancestry GWAS. In addition, UK BioBank lacked information on gestational age for adjustment, which could have an impact on difference in strength of association compared to the results obtained from other European ancestry GWAS. However, we observed no evidence of heterogeneity in BW allelic effects at lead SNPs between the two components of European ancestry meta-analysis, using Cochran's Q statistic[46], implemented in GWAMA[44], after Bonferroni correction (P>0.00083) (**Supplementary Table 3**). We tested for heterogeneity in allelic effects between studies within the European component using Cochran's Q. At loci demonstrating evidence of heterogeneity, we confirmed that association signals were not being driven by outlying studies by visual inspection of forest plots. We performed sensitivity analyses to assess the impact of covariate adjustment (gestational age and population structure) on heterogeneity.

We were also concerned that overlap of individuals (duplicated or related) between the two components of the European ancestry meta-analysis might lead to false positive association signals. We performed bivariate LD Score regression[12] using the two components of the European ancestry meta-analysis and observed a genetic covariance intercept of 0.0156 (SE 0.0058), indicating a maximum of 1,119 duplicate individuals. Univariate LD Score regression[12] of the European ancestry meta-analysis estimated the intercept as 1.0426, which may indicate population structure or relatedness that is not adequately accounted for in the analysis. To assess the impact of this inflation on the European ancestry meta-analysis, we expanded the standard errors of BW allelic effect size estimates and re-calculated association P-values. On the basis of this adjusted analysis, the lead SNP only at *MTNR1B* dropped below genome-wide significance (rs10830963, P=5.5x10^{-8}).

Trans-ancestry meta-analysis. The trans-ancestry meta-analysis combined the two European ancestry components with an additional 10,104 individuals from six GWAS from diverse ancestry groups: African American, Chinese, Filipino, Surinamese, Turkish and Moroccan. Within each GWAS, we first combined sex-specific BW association summary statistics in a fixed-effects meta-analysis, implemented in GWAMA[44] and applied a second round of genomic control[39]. Subsequently, we combined association summary statistics from the six non-European GWAS and the two European ancestry components in a trans-ancestry fixed-effects meta-analysis, implemented in GWAMA[44]. Variants failing GWAS quality control filters in UK Biobank, reported in less than 50% of the total

sample size in the first component, or with MAF <0.1%, were excluded from the trans-ancestry meta-analysis. We tested for heterogeneity in allelic effects between ancestries using Cochran's *Q*[46].

Prioritising candidate genes in each BW locus. We combined a number of approaches to prioritise the most likely candidate gene(s) in each BW locus. Expression quantitative trait loci (eQTLs) were obtained from the Genotype Tissue Expression (GTEx) Project[47], the GEUVADIS Project[48] and eleven other studies[49-59] using HaploReg v4[60]. We interrogated coding variants for each BW lead SNP and its proxies (EUR r^2>0.8) using Ensembl[61] and HaploReg. Their likely functional consequences were predicted by SIFT[62] and PolyPhen2[63]. Biological candidacy was assessed by presence in significantly enriched gene set pathways from MAGENTA analyses (see below for details). We extracted all genes within 300 kb of all lead BW SNPs and searched for connectivity between any genes using STRING[64]. If two or more genes between two separate BW loci were connected, they were given an increased prior for both being plausible candidates. We also applied protein-protein interaction (PPI) analysis (see below for details) to all genes within 300 kb of each lead BW SNPs and ranked the genes based on the score for connectivity with the surrounding genes.

Approximate conditional analysis. We searched for multiple distinct BW association signals in each of the established and novel loci, defined as 1Mb up- and down-stream of the lead SNP from the trans-ancestry meta-analysis, through approximate conditional analysis. We applied GCTA[65] to identify "index SNPs" for distinct association signals attaining genome-wide significance (P<5x10^{-8}) in the European ancestry meta-analysis using a reference sample of 5,000 individuals of white British origin, randomly selected from UK Biobank, to approximate patterns of linkage disequilibrium (LD) between variants in these regions. Note that we performed approximate conditioning on the basis of only the European ancestry meta-analysis because GCTA cannot accommodate LD variation between diverse populations.

Fine-mapping analyses. We sought to leverage LD differences between populations contributing to the trans-ancestry meta-analysis and to take advantage of the improved coverage of common and low-frequency variation offered by 1000G or 1000G and UK10K combined imputation to localise variants driving each distinct association signal achieving locus-wide significance. For each distinct signal, we used MANTRA[66] to construct 99% credible sets of variants[67] that together account for 99% of the posterior probability of driving the association. MANTRA incorporates a prior model of relatedness between studies, based on mean pair-wise allele frequency differences across loci, to account for heterogeneity in allelic effects (**Supplementary Table 3**). MANTRA has been demonstrated, by simulation, to improve localisation of causal variants compared with either a fixed- or random-effects trans-ancestry meta-analysis[66,68].

For loci with only one signal of association, we used MANTRA to combine summary statistics from the six non-European GWAS and the two European ancestry components. However, for loci with multiple distinct association signals, we used MANTRA to combine summary statistics from approximate conditioning for the two European components, separately for each signal.

For each distinct signal, we calculated the posterior probability that the jth variant, π_{Cj}, is driving the association, given by

$$\pi_{Cj} = \frac{\Lambda_j}{\sum_k \Lambda_k},$$

where the summation is over all variants mapping within the (conditional) meta-analysis across the locus. In this expression, Λ_j is the Bayes' factor (BF) in favour of association from the MANTRA analysis. A 99% credible set[67] was then constructed by: (i) ranking all variants according to their BF, Λ_j; and (ii) including ranked variants until their cumulative posterior probability exceeds 0.99.

Genomic annotation. We used genomic annotations of DNaseI hypersensitive sites (DHS) from the ENCODE[69] project and protein coding genes from GENCODE[70]. We filtered cell types that are cancer cell lines (karyotype 'cancer' from https://genome.ucsc.edu/ENCODE/cellTypes.html), and merged data from multiple samples from the same cell type. This resulted in 128 DHS cell-type annotations, as well as 4 gene-based annotations (coding exon, 5UTR, 3UTR and 1kb upstream of TSS). First, we tested for the effect of each cell type DHS and gene annotation individually using the Bayes Factors for all variants in the 62 credible sets using fgwas[71]. Second, we categorised the annotations into 'genic', 'foetal DHS', 'embryonic DHS', 'stem cell DHS', 'neonatal DHS' and 'adult DHS' based on the description fields from ENCODE, and tested for the effect of each category individually as described above using fgwas. Third, we then tested the effect of each category by including all categories in a joint model using fgwas. For each of the three analyses, we obtained the estimated effects and 95% confidence intervals (CI) for each annotation, and considered an annotation enriched if the 95% CI did not overlap zero.

Evaluation of imputation quality of low-frequency variant at the *YKT6-GCK* locus. At the *YKT6-GCK* locus, the lead SNP (rs138715366) is of low-frequency in European ancestry populations (MAF=0.92%) and even rarer in other ancestry groups (MAF=0.23% in African Americans, otherwise monomorphic) and is not present in the HapMap reference panel[72]. To assess the accuracy of imputation for this low-frequent variant, we genotyped rs138715366 in the Northern Finland Birth Cohort (NFBC) 1966 (**Supplementary Table 1**). Of the 5,009 samples in the study, 4,704 were successfully imputed and genotyped (or sequenced) for rs138715366. The overall concordance rate between imputed and directly assayed genotypes was 99.8% and for directly assayed heterozygote calls was 75.0%.

Estimation of genetic variance explained. Variance explained was calculated using the REML method implemented in GCTA[73]. We considered the variance explained by two sets of SNPs: (i) lead SNPs of all 62 distinct association signals at the 59 established and novel autosomal BW loci identified in the European-specific or trans-ancestry meta-analyses; (ii) lead SNPs of 55 distinct association signals at the 52 novel autosomal BW loci (**Extended Data 2** and **Supplementary Table 7**). Variance explained was calculated in samples of European ancestry in the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study[74] (independent of the meta-analysis) and two studies that were part of the European ancestry meta-analysis: NFBC1966 and Generation R (**Supplementary Table 1**). In each study, the genetic relationship matrix was estimated for each set of SNPs and was tested individually against BW (males and females combined) with study specific covariates. These analyses provided an estimate and standard error for the variance explained by each of the given sets of SNPs.

Examining the relative effects on BW of maternal and foetal genotype at the 60 identified loci. We performed four sets of analyses:

(i) We used GWAS data from 4,382 mother-child pairs in the Avon Longitudinal Study of Parents and Children (ALSPAC) study to fit a "maternal-GCTA model"[6] to estimate the extent to which the maternal genome might influence offspring BW independent of the foetal genome. The m-GCTA model uses genome-wide genetic similarity between mothers and offspring to partition the phenotypic variance in BW into components due to the maternal genotype, the child's genotype, the covariance between the two and environmental sources of variation;

(ii) We compared associations with BW of the foetal versus maternal genotype at each of the 60 BW loci. The maternal allelic effect on offspring BW was obtained from a maternal GWAS meta-analysis of 68,254 European mothers from the EGG Consortium (n=19,626)[7] and the UK Biobank (n=48,628). In the UK Biobank, mothers were asked to report the BW of their first child. Women of European ancestry with genotype data

available in the May 2015 data release were included, and those with reported BW equivalent to < 2.5 kg or > 4.5kg were excluded. No information on gestational age or gender of child was available. BW of first child was associated with maternal factors such as smoking status, BMI and height in the expected directions. Of the 68,254 women included in the maternal GWAS, 13% were mothers of individuals included in the current foetal European ancestry GWAS, and a further approximately 45% were themselves (with their own BW) included in the foetal GWAS;

(iii) We additionally conducted analyses in 12,909 mother-child pairs from nine contributing studies: at each of the 60 loci, we compared the effect of the foetal genotype on BW adjusted for sex and gestational age, with and without adjustment for maternal genotype. We reciprocally compared the association between the maternal genotype and BW with and without adjustment for foetal genotype;

(iv) We used the method of Zhang et al[20] to test associations between BW and the maternal untransmitted, maternal transmitted and paternal transmitted haplotype score of 422 height SNPs[29], 30 SBP SNPs[18,19] and 84 T2D SNPs[28] in 5,201 mother-child pairs from the ALSPAC study.

LD Score Regression. The use of LD Score regression to estimate the genetic correlation between two traits/diseases has been described in detail elsewhere[75]. Briefly, "LD Score" is a measure of how much genetic variation each variant tags; if a variant has a high LD Score then it is in high LD with many nearby polymorphisms. Variants with high LD Scores are more likely to contain more true signals and hence provide more chance of overlap with genuine signals between GWAS. The LD score regression method uses summary statistics from the GWAS meta-analysis of BW and the other traits of interest, calculates the cross-product of test statistics at each SNP, and then regresses the cross-product on the LD Score. Bulik-Sullivan et al[75] show that the slope of the regression is a function of the genetic covariance between traits:

$$E\left(z_{1j}z_{2j}\right) = \frac{\sqrt{N_1N_2}\rho_g}{M} l_j + \frac{\rho N_s}{\sqrt{N_1N_2}}$$

where N_i is the sample size for study i, ρ_g is the genetic covariance, M is the number of SNPs in the reference panel with MAF between 5% and 50%, l_j is the LD score for SNP j, N_s quantifies the number of individuals that overlap both studies, and ρ is the phenotypic correlation amongst the N_s overlapping samples. Thus, if there is sample overlap (or cryptic relatedness between samples), it will only affect the intercept from the regression (i.e. the term $\frac{\rho N_s}{\sqrt{N_1N_2}}$) and not the slope, and hence estimates of the genetic covariance will not be biased by sample overlap. Likewise, population stratification will affect the intercept but will have minimal impact on the slope (i.e. intuitively since population stratification does not correlate with linkage disequilibrium between nearby markers).

Summary statistics from the GWAS meta-analysis for traits and diseases of interest were downloaded from the relevant consortium website. The summary statistics files were reformatted for LD Score regression analysis using the munge_sumstats.py python script provided on the developer's website (https://github.com/bulik/ldsc). For each trait, we filtered the summary statistics to the subset of HapMap 3 SNPs[76], as advised by the developers, to ensure that no bias was introduced due to poor imputation quality. Summary statistics from the European-specific BW meta-analysis were used because of the variable LD structure between ancestry groups. Where the sample size for each SNP was included in the results file this was flagged using --N-col; if no sample size was available then the maximum sample size reported in the reference for the GWAS meta-analysis was used. SNPs were excluded for the following reasons: MAF<0.01; ambiguous strand; duplicate rsID; non-autosomal SNPs; reported sample size less than 60% of the total available. Once all files were

reformatted, we used the ldsc.py python script, also on the developers' website, to calculate the genetic correlation between BW and each of the traits and diseases. The European LD Score files that were calculated from the 1000G reference panel and provided by the developers were used for the analysis. Where multiple GWAS meta-analyses had been conducted on the same phenotype (i.e. over a period of years), the genetic correlation with BW was estimated using each set of summary statistics and presented in **Supplementary Table 12**. The phenotypes with multiple GWAS included height, BMI, waist-hip ratio (adjusted for BMI), total cholesterol, triglycerides, HDL and LDL. The estimate of the genetic correlation between the multiple GWAS meta-analyses on the same phenotype were comparable and the later GWAS had a smaller standard error due to the increased sample size, so only the genetic correlation between BW and the most recent meta-analyses were presented in **Figure 2**.

In the published GWAS for BP[19] the phenotype was adjusted for BMI. Caution is needed when interpreting the genetic correlation between BW and BMI-adjusted SBP due to the potential for collider bias[77]. Since BMI is associated with both BP and BW, it is possible that the use of a BP genetic score adjusted for BMI might bias the genetic correlation estimate towards a more negative value. To verify that the inverse genetic correlation with BW (r_g=-0.26, SE=0.05, P=6.5x10^{-9}) was not due to collider bias caused by the BMI adjustment of the phenotype, we obtained an alternative estimate using UK Biobank GWAS data for SBP that was unadjusted for BMI and obtained a similar result (r_g=-0.22, SE=0.03, P=5.5x10^{-13}). The SBP phenotype in UK Biobank was prepared as follows. Two BP readings were taken at assessment, approximately 5 minutes apart. We included all individuals with an automated BP reading (taken using an automated Omron BP monitor). Two valid measurements were available for most participants (averaged to create a BP variable, or alternatively a single reading was used if only one was available). Individuals were excluded if the two readings differed by more than 4.56 SD. BP measurements more than 4.56 SD away from the mean were excluded. We accounted for BP medication use by adding 15 mmHg to the SBP measure. BP was adjusted for age, sex and centre location and then inverse rank normalised. We performed the GWAS on 127,698 individuals of British descent using BOLT-LMM[42], with genotyping array as covariate.

Estimating the proportion of the BW-adult traits covariance attributable to genotyped SNPs. We estimated the phenotypic, genetic and residual correlations as well as the genetic and residual covariance between BW and several quantitative traits/disease outcomes in UK Biobank using directly genotyped SNPs and the REML method implemented in BOLT-LMM[42]. The traits examined included T2D, SBP, DBP, CAD, height, BMI, weight, waist-hip ratio, hip circumference, waist circumference, obesity, overweight, age at menarche, asthma, and smoking. Where phenotypes were not available (e.g. serum blood measures are not currently available in UK Biobank), we obtained estimates using the NFBC1966 study (for correlations/covariance between BW and triglycerides, Total-, HDL- and LDL-cholesterol, fasting glucose and fasting insulin). In the UK Biobank analysis, we used 57,715 unrelated individuals with BW available and identified by UK Biobank as white British. SNPs with evidence of deviation from Hardy-Weinberg Equilibrium (P<1x10^{-6}), MAF<0.05 or overall missing rate >0.015 were excluded, resulting in 328,928 SNPs for analysis. We included the first five ancestry principal components as covariates. In the NFBC1966 analysis, 5,009 individuals with BW were enrolled. Genotyped SNPs that passed quality control (**Supplementary Table 2**) were included, resulting in 324,895 SNPs for analysis. The first three ancestry principal components and sex were included as covariates.

Gene set enrichment analysis. Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) was used to explore pathway-based associations using summary statistics from the trans-ancestry meta-analysis. MAGENTA implements a gene set enrichment analysis (GSEA) based approach, as previously described[13]. Briefly, each gene in the genome is mapped to a single index SNP with the lowest P-value within a 110 kb upstream and 40 kb downstream window. This P-value,

representing a gene score, is then corrected for confounding factors such as gene size, SNP density and LD-related properties in a regression model. Genes within the HLA-region were excluded from analysis due to difficulties in accounting for gene density and LD patterns. Each mapped gene in the genome is then ranked by its adjusted gene score. At a given significance threshold (95th and 75th percentiles of all gene scores), the observed number of gene scores in a given pathway, with a ranked score above the specified threshold percentile, is calculated. This observed statistic is then compared to 1,000,000 randomly permuted pathways of identical size. This generates an empirical GSEA *P*-value for each pathway. Significance was attained when an individual pathway reached a false discovery rate (FDR) <0.05 in either analysis. In total, 3,216 pre-defined biological pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with BW. The MAGENTA software was also used for enrichment testing of custom gene sets.

Protein-Protein interaction network analyses. We used the integrative Protein-Interaction-Network-Based Pathway Analysis (iPINBPA) method[78]. Briefly, we generated gene-wise *P*-values from the trans-ancestry meta-analysis using VEGAS2[79], which map the SNPs to genes and account for possible cofounders, such as LD between markers. The empirical gene-wise *P*-values are calculated using simulations from the multivariate normal distribution. Those that were nominally significant ($P \leq 0.01$) were selected as "seed genes", and were collated within high confidence version of inweb3[14], to weight the nodes in the network following a guilt-by-association approach. In a second step, a network score was defined by the combination of the *z*-scores derived from the gene-wise *P*-values with node weights using the Liptak-Stouffer method[80]. A heuristic algorithm was then applied to extensively search for modules enriched in genes with low *P*-values. The modules were further normalised using a null distribution of 10,000 random networks. Only those modules with *z*-score >5 were selected. Finally, the union of all modules constructed a BW-overall PPI network. Both the proteins on the individual modules and on the overall BW-PPI were interrogated for enrichment in Gene Ontology Terms (Biological Processes) using a Hypergeometric test. Terms were considered as significant when adjusted *P*-value, following Benjamini-Hochberg procedure, was below 0.05.

Point of contact analyses (PoC). The same methodology described above was applied to 16 different adult traits resulting in a number of enriched modules per trait. Different modules for each trait were combined in a single component and the intersection between these trait-specific components and the BW component was calculated. This intersection is defined as the PoC network. We used the resulting PoC networks in downstream analyses to interrogate which set of proteins connects BW variation and adult trait variation via pathways enriched in the overall BW analysis.

Parent-of-origin specific associations. We first searched for evidence of parent of origin effects in the UK Biobank samples by comparing variance between heterozygotes and homozygotes using Quicktest[81]. In this analysis, we used only unrelated individuals identified genetically as of white British origin (n=57,715). Principal components were generated using these individuals and the first five were used to adjust for population structure as covariates in the analysis, in addition to a binary indicator for genotyping array.

We also examined 4,908 mother-child pairs in ALSPAC and determined the parental origin of the alleles where possible[82]. Briefly, the method uses mother-child pairs to determine the parent of origin of each allele. For example, if the mother/child genotypes are AA/Aa, the child's maternal/paternal allele combination is A/a. For the situation where both mother and child are heterozygous, the child's maternal/paternal alleles cannot be directly specified. However, the parental origin of the alleles can be determined by phasing the genotype data and comparing maternal and child haplotypes. We then tested these alleles for association with BW adjusting for sex and gestational age.

Statistical power in these currently available sample sizes is insufficient to rule out widespread parent-of-origin effects across the regions tested. Using the mean beta (0.034SD) and MAF (0.28) of the identified loci, we estimate that we would need at least 200,000 unrelated individuals or 70,000 mother-child pairs for 80% power to detect parent-of-origin effects at P<0.00085.

Hierarchical clustering of BW loci. To explore the different patterns of association between BW and other anthropometric/metabolic/endocrine traits and diseases, we performed hierarchical clustering analysis. The lead SNP (or proxy, EUR r^2>0.6) at the 60 BW loci was queried in publicly available GWAS meta-analysis datasets or in GWAS result obtained through collaboration[83]. Results were available for 53 of those loci and the extracted *z*-score (allelic effect/SE, **Supplementary Table 17**) was aligned to the BW-raising allele. We performed 2D clustering by trait and by locus. We computed the Euclidean distance amongst *z*-scores of the extracted traits/loci and performed complete hierarchical clustering implemented in the pvclust package (http://www.sigmath.es.osaka-u.ac.jp/shimo-lab/prog/pvclust/) in R v3.2.0 (http://www.R-project.org/). Clustering uncertainty was measured by multiscale bootstrap resampling estimated from 1,000 replicates. We used α=0.05 to define distinct clusters and, based on the bootstrap analysis, calculated the Calinski index to identify the number of well-supported clusters (cascadeKM function, Vegan package, http://CRAN.R-project.org/package=vegan). Clustering was visualised by constructing dendrograms and a heatmap.

Separately from the hierarchical clustering analysis, we queried the lead SNP at *EPAS1* in a GWAS of haematological traits[84] because variation at that locus has previously been implicated in BW and adaptation to hypoxia at high altitudes in Tibetans[85,86] (**Supplementary Table 17**).

ADDITIONAL REFERENCES FOR METHODS

35. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**, 499-511 (2010).
36. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-959 (2012).
37. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, *et al*. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-1212 (2014).
38. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).
39. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
40. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, *et al*. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42**, 348-354 (2010).
41. Allen NE, Sudlow C, Peakman T, Collins R. UK Biobank. UK biobank data: come and get it. *Sci Transl Med* **6**, 224ed4 (2014).
42. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290 (2015).
43. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR *et al*. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* **81**, 559-575 (2007).
44. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
45. Willer CJ, Li Y, and Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

46. Ioannidis J, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* **2**, e0000841 (2007).
47. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene. *Science* **348**, 648-660 (2015).
48. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, *et al*. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-511 (2013).
49. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, *et al*. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777 (2010).
50. Schadt EE, Molony C, Chudin E, Hao K, Yang X, *et al*. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* **6**, e107 (2008).
51. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, *et al*. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* **6**, e1000952 (2010).
52. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, *et al*. Population genomics of human gene expression. *Nat Genet* **9**, 1217-1224 (2007).
53. Li Q, Stram A, Chen C, Kar S, Gayther S, *et al*. Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. *Hum Mol Genet* **23**, 5294-5302 (2014).
54. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).
55. Zou F, Chai HS, Younkin CS, Allen M, Crook J, *et al*. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet* **8**, e1002707 (2012).
56. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, *et al*. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
57. Koopmann TT, Adriaens ME, Moerland PD, Marsman RF, Westerveld ML, *et al*. Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS One* **9**, e97380 (2014).
58. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, *et al*. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
59. Grundberg E, Adoue V, Kwan T, Ge B, Duan QL, *et al*. Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet* **7**, e1001279 (2011).
60. Ward LD and Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-934 (2012).
61. Flicek P, Amode MR, Barrell D, Beal K, Billis K, *et al*. Ensembl 2014. *Nucleic Acids Res* **42**, D749-755 (2014).
62. Kumar P, Henikoff S, Ng P. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081 (2009).
63. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-249 (2010).
64. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, *et al*. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-452 (2015).
65. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, *et al*. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375 (2012).
66. Morris AP. Transethnic meta-analysis of genome-wide association studies. *Genet Epidemiol* **35**, 809-822 (2011).

67. Wellcome Trust Case Control Consortium, Maller JB, McVean G, Byrnes J, Vukcevic D, *et al*. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* **44**, 1294-1301 (2012).
68. Wang X, Chua HX, Chen P, Ong RT, Sim X, *et al*. Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum Mol Genet* **22**, 2303-2311 (2013).
69. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
70. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, *et al*. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774 (2012).
71. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet* **94**, 559–573 (2014).
72. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58 (2010).
73. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, *et al*. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569 (2010).
74. Urbanek M, Hayes MG, Armstrong LL, Morrison J, Lowe LP *et al*. The chromosome 3q25 genomic region is associated with measures of adiposity in newborns in a multi-ethnic genome-wide association study. *Hum Mol Genet* **22**, 3583-3596 (2013).
75. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, *et al*. An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241 (2015).
76. The International HapMap Consortium. A second generation haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
77. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet* **96**, 329-339 (2015).
78. Wang L, Mousavi P, Baranzini SE. iPINBPA: an integrative network-based functional module discovery tool for genome-wide association studies. *Pac Symp Biocomput* 255-266 (2015).
79. Mishra A and Macgregor S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet* **18**, 86-91 (2015).
80. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* **18**, 1368-1373 (2005).
81. Hoggart, CJ, Venturini G, Mangino M, Gomez F, Ascari G, *et al*. Novel Approach Identifies SNPs in SLC2A10 and KCNK9 with Evidence for Parent-of-Origin Effect on Body Mass Index. *PLoS Genet* **10**, 1–12 (2014).
82. Wang S, Yu Z, Miller RL, Tang D & Perera FP. Methods for detecting interactions between imprinted genes and environmental exposures using birth cohort designs with mother-offspring pairs. *Hum Hered* **71**, 196–208 (2011).
83. Painter JN, Anderson CA, Nyholt DR, Macgregor S, Lin J, *et al*. Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet* **43**, 51-54 (2011).
84. Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, *et al*. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet* **41**, 1191-1198 (2009).
85. Xu XH, Huang XW, Qun L, Li YN, Wang Y, *et al*. Two functional loci in the promoter of EPAS1 gene involved in high-altitude adaptation of Tibetans. *Sci Rep* **4**, 7465 (2014).
86. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, *et al*. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194-197 (2014).
87. Kato N, Loh M, Takeuchi F, Verweij N, Wang X, *et al*. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet* **47**, 1282-1293 (2015).

ENDNOTES

Supplementary Information is available in the online version of the paper.

Acknowledgements Full grant supports and acknowledgements can be found in the Supplementary Information.

Author Contributions

Core analyses and writing: M.H., R.N.B., F.R.D., N.M.W., M.N.K., J.F-T., N.R.v.Z., K.J.G., A.P.M., K.K.O., J.F.F., N.J.T., J.R.P., D.M.E., M.I.M., R.M.F. Statistical analysis in individual studies (lead analysts in italics): *M.H., R.N.B.,* F.R.D., N.M.W., *M.N.K., B.F., N.G., J.P.B., D.P.S., R.L-G., T.S.A., E.K-M., R.R., L-P.L., D.L.C., Y.W., E.T., C.A.W., C.T.H., J-J.H., N.V-T., P.K.J., E.T.H.B., I.N., N.P., A.M., E.M.v.L., R.J., V. Lagou, M.N.,* J.M.M., S.E.J., P-R.L., K.S.R., M.A.T., J.T., A.R.W., H.Y., D.M.S., I.P., K. Panoutsopoulou, X.W., L.C., F.G., K.E.S., M. Murcia, E.V.R.A., Z.K., S.B.-G., F.S., D.T., J.W., C.M-G., N.R.R., E.Z., G.V.D., Y-Y.T., H.N.K., A.P.M., J.F.F., *N.J.T., J.R.P.*, D.M.E., R.M.F. GWAS look-up in unpublished datasets: K.T.Z., N.R., D.R.N., R.C.W.M., C.H.T.T., W.H.T., S.K.G., F.J.v.R. Sample collection and data generation in individual studies: F.R.D., M.N.K., B.F., N.G., J.P.B., D.P.S., R.L-G., R.R., L-P.L., J-J.H., I.N., E.M.v.L., M.B., P.M-V., A.J.B., L.P., P.K., M.A., S.M.W., F.G., C.E.v.B., G.W., E.V.R.A., C.E.F., C.T., C.M.T., M. Standl, Z.K., M.V.H., H.G.d.H., F.R.R., C.M-G., S.M.R., G.H., G.M., N.R.R., C.J.G., C.L., J.L., R.A.S., J.H.Z., F.D.M., W.L.L.Jr, A.T., M. Stumvoll, V. Lindi, T.A.L., C.M.v.D., A.K., T.I.S., H.N., K. Pahkala, O.T.R., E.Z., G.V.D., S-M.S., M. Melbye, H.C., J.F.W., M.V., J-C.H., T.H., L.J.B., J.P.N., C.E.P., L.S.A., J.B.B., K.L.M., J.G.E., E.E.W., M.K., J.S.V., T.L., P.V., K.B., H.B., D.O.M-K., F.R., A.G.U., C. Pisinger, O.P., N.J.W., H.H., V.W.J., S.F.G., A.A.V., D.A.L., G.D.S., K.K.O., J.F.F., N.J.T., J.R.P., M.I.M. Functional follow-up experiment: L.A.D., S.M.M., R.M.R., E.D., B.R.W. Individual study design and principal investigators: J.P.B., I.N., M.A., F.D.M., W.L.L.Jr, A.T., M. Stumvoll, V. Lindi, T.A.L., C.M.v.D., W.K., A.K., T.I.S., H.N., K. Pahkala, O.T.R., G.V.D., Y-Y.T., S-M.S., M. Melbye, H.C., J.F.W., M.V., E.J.d.G., D.I.B., H.N.K., J-C.H., T.H., A.T.H., L.J.B., J.P.N., C.E.P., J.H., L.S.A., J.B.B., K.L.M., J.G.E., E.E.W., M.K., J.S.V., T.L., P.V., K.B., H.B., D.O.M-K., A.H., F.R., A.G.U., C. Pisinger, O.P., C. Power, E.H., N.J.W., H.H., V.W.J., M-R.J., S.F.G., A.A.V., T.M.F., A.P.M., K.K.O., N.J.T., J.R.P., M.I.M., R.M.F.

Author Information

Correspondence and requests for materials should be addressed to mark.mccarthy@drl.ox.ac.uk and r.freathy@ex.ac.uk. Reprints and permissions information is available at www.nature.com/reprints.

Disclosures

Krina Zondervan has a scientific collaboration with Bayer HealthCare Ltd. and Population Diagnostics Inc.

Grants and funding supports for studies

The Academy of Finland [41071, 1114194, 117787, 120315, 121584, 124282, 126925, 129287, 129378, 134309, 206374, 218029, 24300796, 251360, 267561, 276861, 286284 and EGEA-project]; Althingi (the Icelandic Parliament); the American Diabetes Association; the Arthritis Research UK; the Augustinus Foundation; Baylor Medical College [N01- HC-55016]; the Becket Foundation; Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL); Biocenter Oulu, University of Oulu, Finland; Biomedical Research Council, Singapore (BMRC 06/1/21/19/466); the British Heart Foundation [SP/13/2/30111]; the C.G. Sundell Foundation; Cambridge Institute for Medical Research (CIMR); the Canadian Institutes of Health Research [MOP-82893]; the Cancer Research UK [SP2024-0201 and SP2024-0204]; the Capital Region Research Foundation; Center for Medical Systems

Biology (CMSB); the Chief Scientist Office of the Scottish Government [CZB/4/276 and CZB/4/710]; the Children's Hospital of Philadelphia [Institute Development Award]; Chinese University of Hong Kong (CUHK) [Faculty of Medicine Outstanding Fellowship]; the city of Kuopio; Conselleria de Sanitat Generalitat Valenciana; the Cotswold Foundation [Research Development Award]; Daniel B. Burke Endowed Chair for Diabetes Research; Danish Centre for Health Technology Assessment; Danish Council for Independent Research; Danish Innovation Foundation [0603-00484B and 0603-00457B]; Danish Diabetes Association; Danish Heart Foundation; Danish Innovation Foundation; the Danish National Research Foundation; Danish Pharmaceutical Association; Danish Pharmacists' Fund; Danish Research Council; DHFD (Diabetes Hilfs- und Forschungsfonds Deutschland); Diabetes and Inflammation Laboratory; Diabetes Research Foundation of Finland; EFSD/Lilly research fellowship; the Egmont Foundation; Emil Aaltonen Foundation; Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam; European Commission [ENGAGE (HEALTH-F4-2007-201413), Framework Programme 5 (QLG2-CT-2002-01254), Framework Programme 6 (018996, 018947 (LSHG-CT-2006-01947) and LSHG-CT-2006-018947), Framework Programme 7 (FP7/2007-2013), H2020-633595 DynaHEALTH action, Beta-JUDO n° 279153 and DG XII]; European Research Council (ERC Advanced, 230374); European Science Foundation (ESF, EU/QLRT-2001-01254); European Union (European Social Fund - ESF); Faculty of Biology and Medicine of Lausanne; Finnish Cardiac Research Foundation; Finnish Cultural Foundation; Finnish Foundation of Cardiovascular Research, Finnish Innovation Fund Sitra; Finnish Ministry of Education and Culture; FIS-FEDER [03/1615, 04/1112, 04/1509, 04/1931, 05/1052, 05/1079, 06/1213, 07/0314, 09/02647, 11/00178, 11/01007, 11/02591, 13/02032, 13/1944, 14/00891, 14/01687, 97/0588, 00/0021-2, PI041436, PI061756, PI081151 and PS0901958], Foundation for Paediatric Research of Finland; French Ministry of Research; Fundació La Marató de TV3; Gene-diet Interactions in Obesity project (GENDINOB); Generalitat de Catalunya [CIRIT 1999SGR 00241], Genetic Association Information Network (GAIN); the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC; German Diabetes Association; German Research Council [DFG - SFB 1052 "Obesity mechanisms"; A01, B01, B03, C01 and SPP 1629 TO 718/2-1]; German Research Foundation Collaborative Research Center [CRC1085]; GlaxoSmithKline; Greek national funds: Heracleitus II; Health Fund of the Danish Health Insurance Societies; Hjartavernd (the Icelandic Heart Association); the Ib Henriksen Foundation; the Impact of our Genomes on Individual Treatment Response in Obese Children (TARGET); the Indo-Danish bi-lateral project, Genetics and Systems Biology of Childhood Obesity in India and Denmark (BioChild); Instituto de Salud Carlos III [Red INMA G03/176 and CB06/02/0041]; Integrated Research and Treatment Center (IFB) Adiposity Diseases [01EO1001]; the Italian Ministry of Health [ICS 110.1RS97.71]; Johns Hopkins University [N01-HC-55020]; the Juho Vainio Foundation; the Juvenile Diabetes Research Foundation International (JDRF); Kuopio University Hospital [5031343]; Kuopio, Tampere and Turku University Hospital Medical Funds [X51001]; Leiden University Medical Center; Leiden University, Research Profile Area Vascular and Regenerative Medicine; the Lundbeck Foundation; the March of Dimes Birth Defects Foundation; Medical Research Council [G0000934, G0500539, G0600705, G0601261, G9502233, MC_U106179471, PrevMetSyn and MRC Doctoral Training Grant Scholarship]; Ministry of Education and Culture of Finland; Ministry of Education, Culture and Science of the Netherlands; Ministry of Health of Denmark; Ministry of Health, Welfare and Sport of the Netherlands; Ministry of Internal Affairs and Health of Denmark; Ministry of Social Affairs and Health of Finland; Ministry of Youth and Families of the Netherlands; MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_12013/1-9); MRC Human Genetics Unit; Municipality of Rotterdam; National Center for Advancing Translational Sciences [CTSI grant UL1TR000124]; National Heart, Lung and Blood Institute (NHLBI) [5R01HL087679, STAMPEED program (1RL1MH083268-01), HHSN268201100005C, HHSN268201100006C, HHSN2682 01100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C and HHSN268 201100012C, HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC8508, U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, N01-HC-25195 and N02-HL-6-4278]; National Health

and Medical Research Council of Australia [003209, 403981 and 572613]; National Human Genome Research Institute (NHGRI) [U01HG 004402]; National Institute of Aging (NIA) [Intramural Research Program, R01AG023629, N01-AG-916413, N01-AG-821336, 263 MD 9164 13 and 263 MD 821336]; NIA/NIH [AG000932-2]; National Institute of Allergy and Infectious Diseases (NIAID); National Institute of Child Health and Human Development (NICHD); National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) [DK063491]; National Institutes of Health (NIH) [1RC2MH089995-01; DK056350, DK078150, DK099820, ES10126, HD34242, HD34243, HG004415, HL085144, RR20649, R01D0042157-01A, R01DK092127-01; R01HD056465, TW05596, U01DK062418, U01HG004423, U01HG004438, U01HG004446, R01HL087641, R01HL59367, R01HL086694, UL1RR025005 and N01-AG-12100]; National Institute for Health Research Cambridge Biomedical Research Centre; National Institute of Mental Health (NIMH) [MH081802, U24 MH068457-06]; National Institute of Neurological Disorders and Stroke (NINDS); National Medical Research Foundation, Singapore (NMRC/0975/2005); NBIC/BioAssist/RK(2008.024); the Netherlands Consortium for Healthy Aging (NCHA); the Netherlands Genomics Initiative (NGI); the Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research [NWO-RFBR 047.017.043]; the Netherlands Organisation for Scientific Research (NWO) [NWO/ZonMw; NWO Genomics; NWO: MagW/ZonMW grants 400-05-717, 480-04-004, 481-08-011, 451- 04-034, 463-06-001, 904-61-090, 904-61-193, 912-10-020, 916-76-125, 985-10-002, Addiction-31160008 Middelgroot-911-09-032 and Spinozapremie 56-464-14192, 175.010.2005.011, 911-03-012]; Neuroscience Campus Amsterdam (NCA); Novo Nordisk Foundation; Novo Nordisk Inc.; the Paavo Nurmi Foundation; Paulo Foundation; the Region Zealand Health and Medical Research Foundation; Research Committee of the Kuopio University Hospital Catchment Area; Research Foundation of Copenhagen County; Research Grant Council General Research Fund [CU473408, CU471713]; Research Institute for Diseases in the Elderly (RIDE) [014-93-015]; Robert Dawson Evans Endowment; the Royal Society; Rutgers University Cell and DNA Repository; the Sigrid Juselius Foundation; Social Insurance Institution of Finland; Spanish Government [SEV-2011-00067]; Spanish Ministry of Science and Innovation [SAF2008-00357]; Spanish National Genotyping Centre (CEGEN-Barcelona); Special Governmental Grants for Health Sciences Research, Turku University Hospital; Swiss National Science Foundation [33CSCO-122661, 33CS30-139468 and 33CS30-148401]; Tampere Tuberculosis Foundation; Turku University Foundation; University of Bristol; University of Cambridge; University of Minnesota [N01-HC-55019]; University of Mississippi Medical Center [N01-HC- 55021]; University of North Carolina [N01-HC-55018]; University of North Carolina at Chapel Hill [N01-HC-55015]; University of Texas Houston [N01-HC-55017]; the VU University's Institute for Health and Care Research (EMGO+); Wellcome Trust [068545/Z/02, 076113/B/04/Z, 079895, 090532, 098017, 098051, 098381, 102215/2/13/2, GR069224]; Yrjö Jahnsson Foundation.

FIGURE LEGENDS

Figure 1 | Manhattan and quantile-quantile (QQ) plots of the trans-ancestry meta-analysis for birth weight. a, Manhattan (main panel) and QQ (top right) plots of genome-wide association results for birth weight from trans-ancestry meta-analysis of up to 153,781 individuals. The association *P*-value (on -$\log_{10}$ scale) for each of up to 22,185,636 SNPs (*y* axis) is plotted against the genomic position (NCBI Build 37; *x* axis). Association signals that reached genome-wide significance ($P<5x10^{-8}$) are shown in green if novel and pink if previously reported. In the QQ plot, the black dots represent observed *P*-values and the grey line represents expected *P*-values under the null distribution. The red dots represent observed *P*-values after excluding the previously identified signals described in Extended Data 2. **b,** Manhattan (main panel) and QQ (top right) plots of trans-ethnic GWAS meta-analysis for BW highlighting the reported imprinted regions described in Supplementary Table 14. Novel association signals that reached genome-wide significance ($P<5x10^{-8}$) and mapped to imprinted regions are shown in green. Genomic regions outside imprinted regions are shaded in grey. SNPs in the imprinted regions are shown in light blue or dark blue, if they are on chromosome with odd number or even number, respectively. In the QQ plot, the black dots represent observed *P* values and the grey lines represent expected *P*-values and their 95% confidence intervals under the null distribution for the SNPs within the imprinted regions.

Figure 2 | Genome-wide genetic correlation between birth weight and a range of traits and diseases in later life. Genetic correlation (r_g) and corresponding standard error between birth weight and the traits displayed on the *x* axis are estimated using LD Score regression[12]. The genetic correlation estimates (r_g) are colour coded according to their intensity and direction (red for positive and blue for negative correlation). HC=head circumference, WHR=waist-hip ratio, WHR (adj BMI)=waist-hip ratio adjusted for body mass index, BMI=body mass index, Pubertal growth=standardized difference in height between age 8 and adult height, Total growth=standardized difference in height at age 14 and adult height, CAD=coronary artery disease, DBP=diastolic blood pressure, SBP=systolic blood pressure*, Chol=total cholesterol, TG=triglycerides, LDL=low-density lipoprotein, HDL=high-density lipoprotein, T2D=type 2 diabetes, HOMA-B=homeostatic model assessment of beta-cell function, HOMA-IR=homeostatic model assessment of insulin resistance, HbA1C=Hemoglobin A1c, LSBMD=lumbar spine bone mineral density, FNBMD=femoral neck bone mineral density, Edu Att=educational attainment, ADHD=attention deficit hyperactivity disorder, Cigarettes/Day=number of cigarettes per day. See Supplementary Table 12 for references for each of the traits and diseases displayed here.

Figure 3 | Hierarchical clustering of birth weight loci based on similarity of overlap with adult diseases, metabolic and anthropometric traits. For the lead SNP at each of the birth weight loci (listed on *y* axis), the z-score aligned to the birth weight-raising allele is obtained from publicly available GWAS results for various diseases and traits (listed on *x* axis) and are displayed in the heatmap. Of the 60 birth weight loci, z-score for lead SNP (or proxy) at 7 loci were not available (see Supplementary Table 17). Positive z-scores in red indicate the BW-raising allele increases the value of the trait concerned, and negative z-scores in blue show that it is associated with decrease of the trait value of interest. Birth weight loci and traits are clustered according to the Euclidean distance amongst z-score of the loci/traits (see Methods). White cross with a black circle in the middle signifies that BW locus is significantly ($P<5x10^{-8}$) associated with the trait of interest in the publicly available GWAS look-up (Supplementary Table 17). If the BW locus does not attain $P<5x10^{-8}$ in the GWAS look-up, but is reported significant elsewhere, it is marked with plain white cross. For example, *ADRB1* (lead BW variant rs7076938) was not significantly associated with various measurements of blood pressure in Ehret *et al.*[19] from which the z-scores were extracted, but was reported as blood pressure locus by Johnson *et al.*[18] and Kato *et al.*[87]. BMI=body mass index, CAD=coronary artery disease, Chol=total cholesterol, DBP=diastolic blood pressure, HC=head

circumference, HDL=high-density lipoprotein, LDL=low-density lipoprotein, SBP=systolic blood pressure, TG=triglycerides, T2D=type 2 diabetes, WHRadjBMI=waist-hip ratio adjusted for BMI.

EXTENDED DATA LEGENDS

Extended Data 1 | Flow chart of the study design.

Extended Data 2 | Sixty loci associated with birth weight (P<5x10^{-8}) in European ancestry meta-analysis of up to 143,677 individuals and/or trans-ancestry meta-analysis of up to 153,781 individuals.

Extended Data 3 | Regional plots for multiple distinct signals at three birth weight loci, *ZBTB7B* (a), *HMGA1* (b) and *PTCH1* (c). Regional plots for each locus are displayed from: the unconditional European-specific meta-analysis of up to 143,677 individuals (left); the approximate conditional meta-analysis for the primary signal after adjustment for the index variant for the secondary signal (middle); and the approximate conditional meta-analysis for the secondary signal after adjustment for the index variant for the primary signal (right). Directly genotyped or imputed SNPs are plotted with their association P-values (on a -$\log_{10}$ scale) as a function of genomic position (NCBI Build 37). Estimated recombination rates (blue lines) are plotted to reflect the local LD structure around the index SNPs and their correlated proxies. SNPs are coloured in reference to LD with the particular index SNP according to a blue to red scale from r^2 = 0 to 1, based on pairwise r^2 values estimated from a reference of 5,000 individuals of white British origin, randomly selected from the UK Biobank.

Extended Data 4 | Absolute effect size and minor allele frequency of lead SNPs at 60 known and novel birth weight loci from the trans-ancestry meta-analysis of up to 153,781 individuals. The effect of the lead SNP (absolute value of beta, *y* axis) is given as a function of minor allele frequency (*x* axis) for known birth weight loci in pink and novel loci in green. Error bars are proportional to the standard error of the effect size. The dashed line indicates 80% power to detect association at genome-wide significance level for the sample size in trans-ancestry meta-analysis. All birth weight loci were common except for two: at the *YKT6-GCK* locus, the index SNP (rs138715366) was of low-frequency in European ancestry populations (MAF=0.92%), and even rarer in other ancestry groups (MAF=0.23% in African Americans, otherwise monomorphic). Similarly, at the *SUZ12P1-CRLF3* locus, the index SNP (rs144843919) was of low-frequency in European and African American ancestry studies (MAF of 3.5% and 4.0%, respectively), and absent from other ancestry groups. None of the variants attaining genome-wide significance in European ancestry or trans-ancestry meta-analyses at these two loci was present in HapMap, suggesting that these two association signals would not have been identified without imputation up to the denser reference panels from 1000G/UK10K, irrespective of the available sample size.

Extended Data 5 | Comparison of foetal effect sizes and maternal effect sizes at 60 known and novel birth weight loci (continues to Extended Data 6). These plots illustrate that in large GWAS of BW, foetal effect size estimates are larger than maternal effect size estimates at 55/60 identified loci (binomial P=1x10^{-11}), suggesting that most of the associations are likely to be driven by the foetal genotype. In conditional analyses that modelled the effects of both maternal and foetal genotypes (n=12,909 mother-child pairs), confidence intervals around the estimates were wide, precluding inference about the likely contribution of maternal vs. foetal genotype at individual loci. For each birth weight locus, the following six effect sizes (with 95% CI) are shown, all aligned to the birth weight-raising allele in the overall trans-ancestry meta-analysis: **foetal_GWAS** = foetal allelic effect on birth weight (from European ancestry meta-analysis of up to n=143,677 individuals); **foetal_unadjusted** = foetal allelic effect on birth weight (unconditioned in n=12,909 mother-child pairs); **foetal_adjusted** = foetal effect conditioned on maternal genotype in n=12,909 mother-child pairs); **maternal_GWAS** = maternal allelic effect on offspring birth weight (from meta-analysis of up to n=68,254 European mothers)[7]; **maternal_unadjusted** = maternal allelic effect on offspring birth weight (unconditioned in n=12,909 mother-child pairs); **maternal_adjusted** = maternal effect

conditioned on foetal genotype (in n=12,909 mother-child pairs). The 60 birth weight loci are ordered by chromosome and position (Supplementary Tables 10 and 11).

Extended Data 6 | Comparison of foetal effect sizes and maternal effect sizes at 60 known and novel birth weight loci (continued from Extended Data 5). The scatterplot illustrates the difference between the foetal (*x* axis) and maternal (*y* axis) effect sizes in the overall maternal vs. foetal GWAS results.

Extended Data 7 | Gene set enrichment analysis and protein-protein interaction (PPI) analysis. Two complementary analyses of the overall GWAS summary data identified enrichment of BW associations in biological pathways related to metabolism, growth and development: Table (a) shows the top results (FDR<0.05 at the 95th percentile enrichment threshold) from a total of 3216 biological pathways tested for enrichment of multiple modest associations with BW. Additionally, results are presented for custom sets of imprinted genes. In Table (b) are the results of a complementary analysis of empirical PPI data, displaying the top 10 most significant pathways enriched for BW-association scores.

Extended Data 8 | Protein-Protein Interaction (PPI) Network analysis. a, illustrates the global largest component birth weight (BW) PPI network containing 13 modules. **b,** the histogram shows the null distribution of z-scores of BW PPI networks based on 10,000 random networks and where the z-scores for the 13 birth weight modules (M1-13) lie. For each module, the two most significant GO terms are depicted. **c,** heatmap showing the enrichment for the top 50 biological processes over-represented in the global BW PPI network of the trait-specific "point of contact" (PoC) PPI networks. **d-e,** trait-specific PoC PPI networks composed of proteins that are shared in both the global birth weight PPI network and networks generated using the same pipeline for each of the adult traits: **d,** canonical Wnt signalling pathway enriched for PoC PPI between BW and blood pressure (BP)-related phenotypes; and **e,** regulation of insulin secretion pathway enriched for PoC between BW and type 2 diabetes (T2D)/fasting glucose (FG). Red nodes are those that are present in PoC for BW and traits of interest; blue nodes correspond to the pathway nodes; purple nodes are those present in both the pathway and PoC. Large nodes correspond to genes in BW loci (within 300kb from the lead SNP), and are orange when they overlap the enriched pathway and the PoC, and have black border if they, amongst all BW loci, have a stronger (top 5) association with at least one of the pairing adult traits.

Extended Data 9 | Quantile-Quantile (QQ) plots of (a) variance comparison between heterozygotes and homozygotes analysis in 57,715 UK Biobank samples and (b) parent-of-origin specific analysis in 4,908 ALSPAC mother-child pairs at 59 autosomal birth weight loci plus *DLK1*. Despite the large sample size, especially of unrelated individuals, these analyses were underpowered to detect parent-of-origin-specific effects at the identified loci. **a,** QQ plot from the Quicktest[81] analysis comparing the BW variance of heterozygotes with homozygotes in 57,715 UK Biobank samples. **b,** QQ plot from the parent-of-origin specific analysis testing the association between birth weight and maternally transmitted vs. paternally transmitted alleles in 4,908 mother-child pairs from the ALSPAC study (Methods, Supplementary Tables 15 and 16). In both panels, the black dots represent lead SNPs at 59 identified autosomal birth weight loci and a further sub-genome-wide significant signal for birth weight near *DLK1* (rs6575803; P=5.6x10^{-8}). The grey lines represent expected *P* values and their 95% confidence intervals under the null distribution for the 60 SNPs.

Extended Data 10 | Summary of previously reported loci for systolic blood pressure (SBP, a), coronary artery disease (CAD, b, e), type 2 diabetes (T2D, c, f) and adult height (d) and their effect on birth weight. a-d, Effect sizes (left *y* axis) of previously reported 30 SBP loci[18,19], 45 CAD loci[27], 84 T2D loci[28] and 422 adult height loci[29] are plotted against changes in birth weight z-score (*x* axis). Effect sizes are aligned to the adult trait-raising allele. The colour of each dot indicates birth weight

association *P* value: red, $P<5\times10^{-8}$; orange, $5\times10^{-8}\leq P<0.001$; yellow, $0.001\leq P<0.01$; white, $P\geq0.01$. The superimposed grey frequency histogram shows the number of SNPs (right *y* axis) in each category of birth weight effect size. **e,** Effect sizes (with 95% CI) on BW of 45 known CAD loci are plotted arranged in the order of CAD effect size from highest to lowest, separating out the known SBP loci and **f,** those of 32 known T2D loci are plotted, subdivided by previously reported categories derived from detailed adult physiological data[31]. Heterogeneity in BW effect sizes between five T2D loci groups with different mechanistic categories was substantial ($P_{het}=1.2x10^{-9}$). In pairwise comparisons, the "beta cell" group of variants differed from the other four groups: fasting hyperglycaemia ($P_{het}=3x10^{-11}$), insulin resistance ($P_{het}=0.002$), proinsulin ($P_{het}=0.78$) and unclassified ($P_{het}=0.02$) groups. All of the birth weight effect sizes plotted in the forest plots are aligned to the trait (or risk)-raising allele.

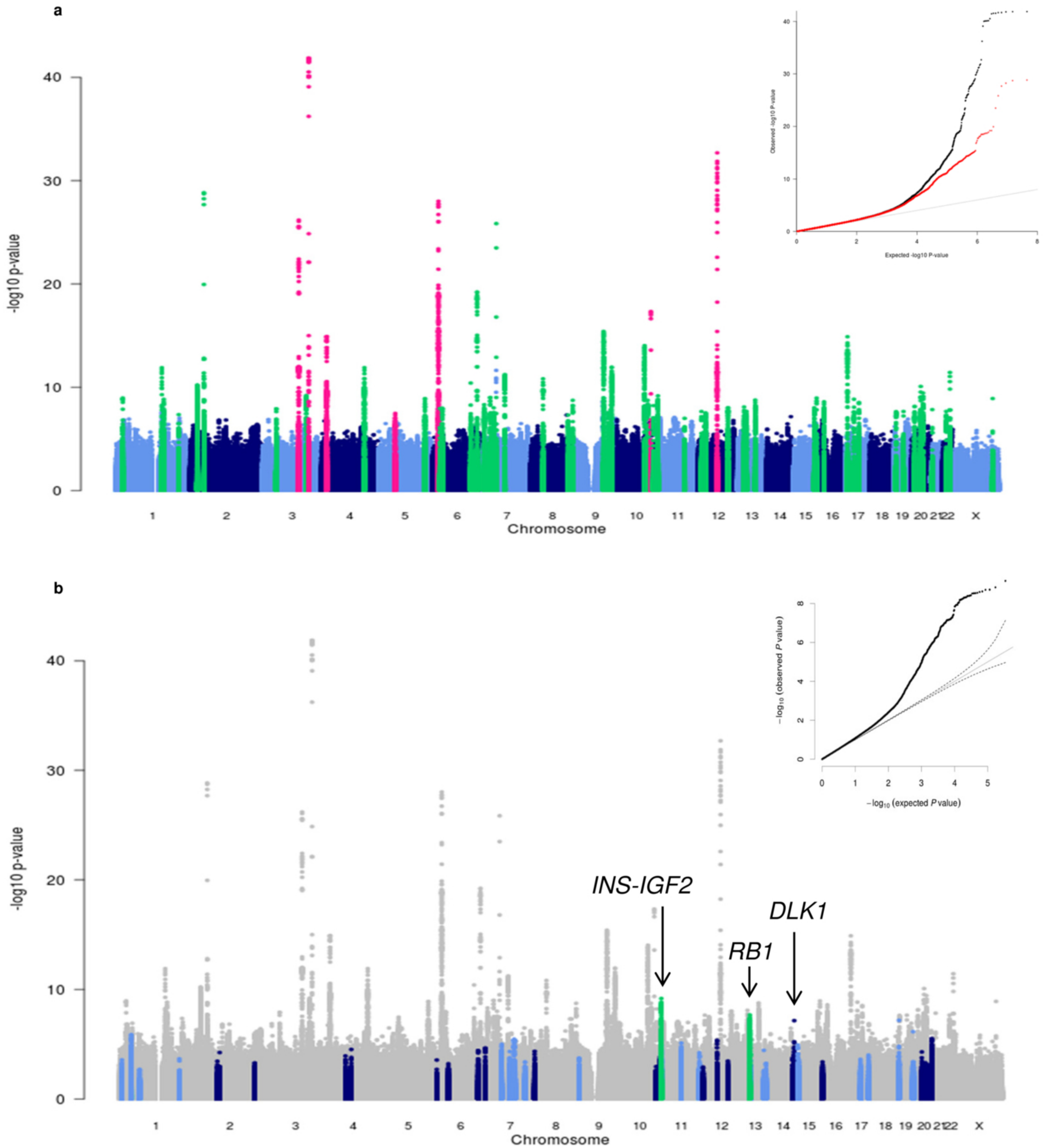
a
-log10 p-value
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Chromosome
Observed -log10 P-value
Expected -log10 P-value
b
-log10 p-value
INS-IGF2
RB1
DLK1
Chromosome
−log10 (observed P value)
−log10 (expected P value)

Figure 1

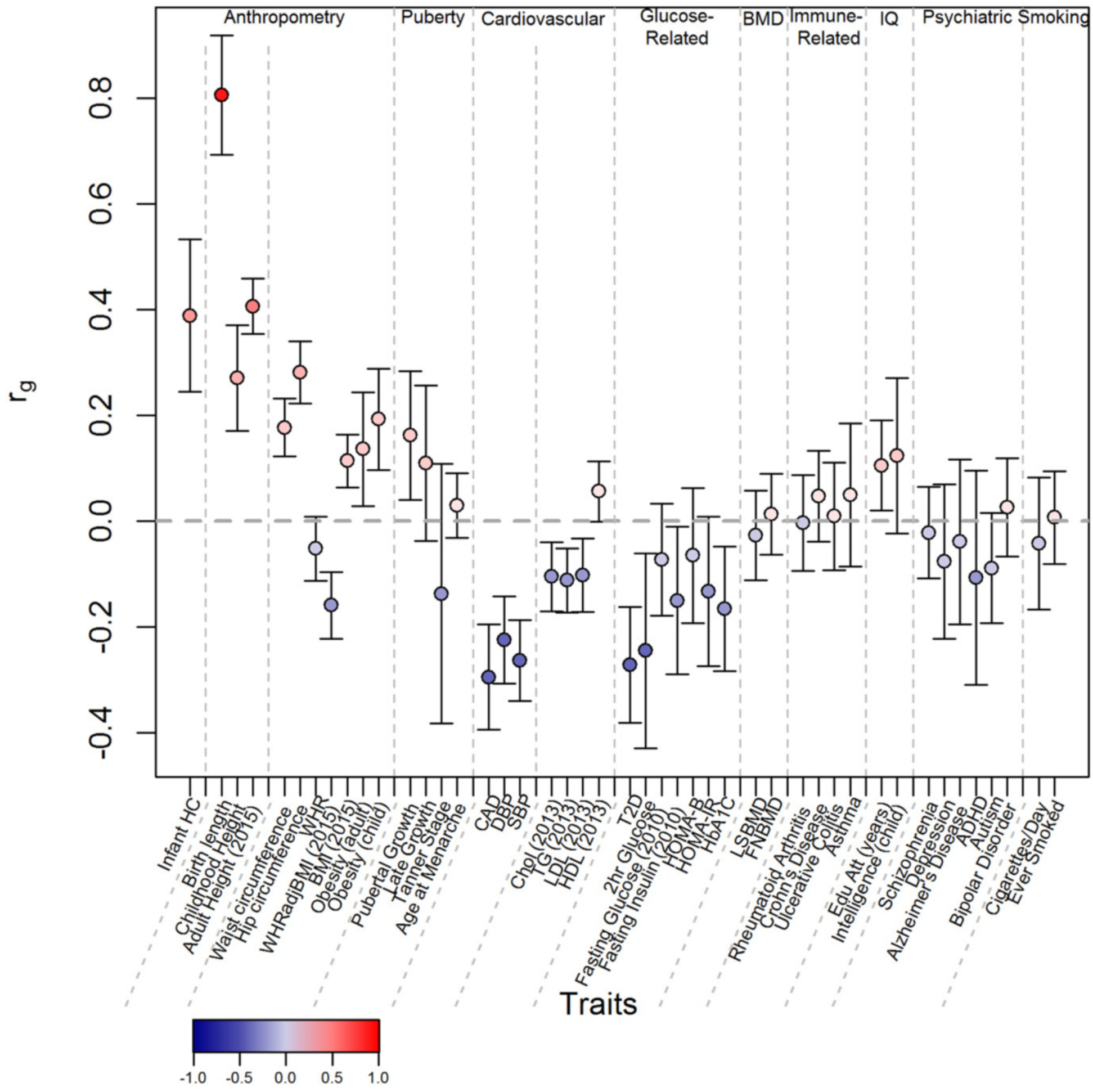
Anthropometry
Puberty
Cardiovascular
Glucose-Related
BMD
Immune-Related
IQ
Psychiatric
Smoking
r_g
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
Infant HC
Birth length
Childhood Height
Adult Height (2015)
Waist circumference
Hip circumference
WHRadjBMI (2015)
WHR
BMI (2015)
Obesity (adult)
Obesity (child)
Pubertal Growth
Late Growth
Tanner Stage
Age at Menarche
CAD
DBP
SBP
Chol (2013)
TG (2013)
LDL (2013)
HDL (2013)
T2D
2hr Glucose
Fasting Glucose (2010)
Fasting Insulin (2010)
HOMA-B
HOMA-IR
HbA1C
LSBMD
FNBMD
Rheumatoid Arthritis
Crohn's Disease
Ulcerative Colitis
Asthma
Edu Att (years)
Inteligence (child)
Schizophrenia
Depression
Alzheimer's Disease
ADHD
Autism
Bipolar Disorder
Cigarettes/Day
Ever Smoked
Traits
-1.0
-0.5
0.0
0.5
1.0

Figure 2

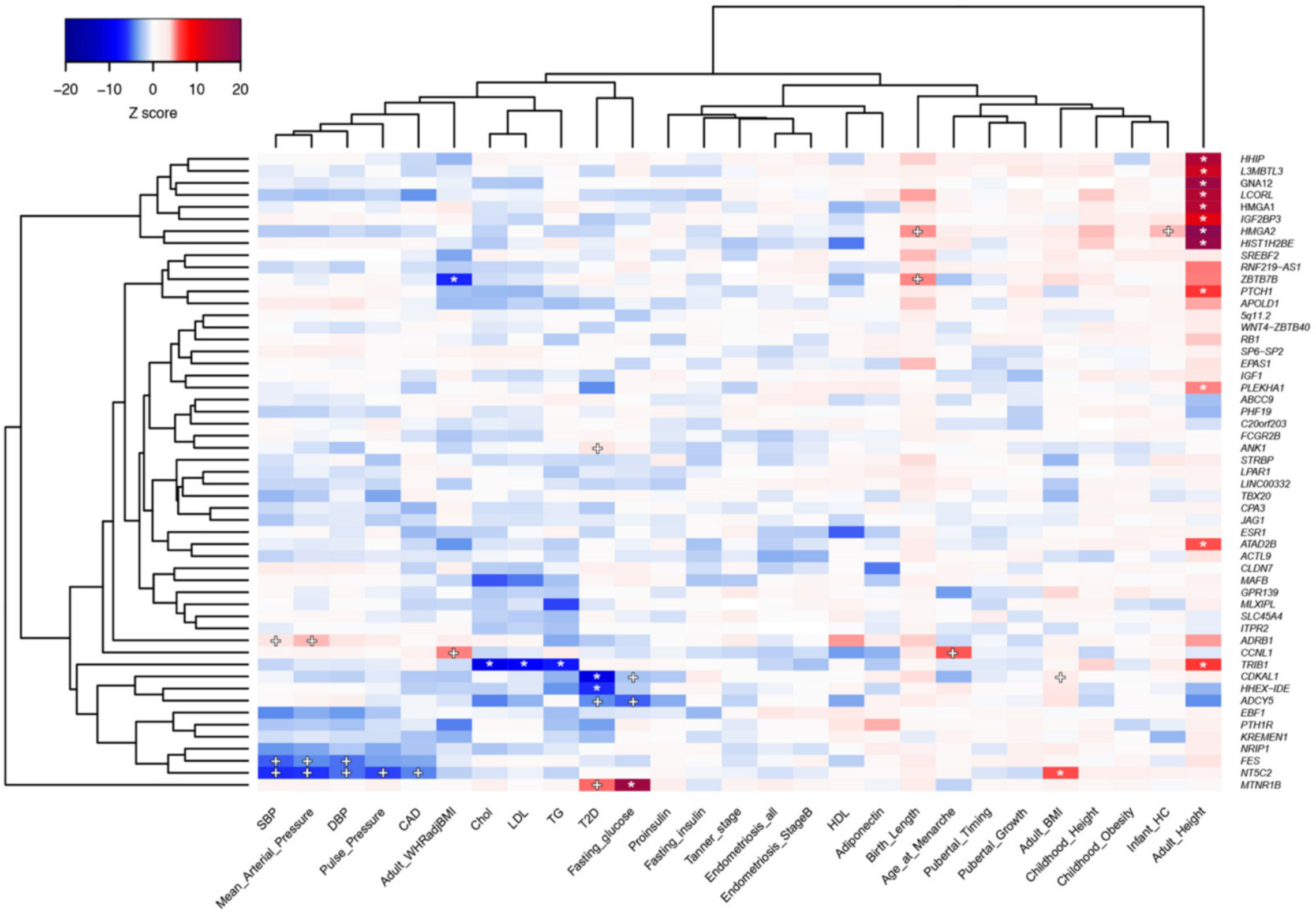
−20
−10
0
10
20
Z score
HHIP
L3MBTL3
GNA12
LCORL
HMGA1
IGF2BP3
HMGA2
HIST1H2BE
SREBF2
RNF219-AS1
ZBTB7B
PTCH1
APOLD1
5q11.2
WNT4-ZBTB40
RB1
SP6-SP2
EPAS1
IGF1
PLEKHA1
ABCC9
PHF19
C20orf203
FCGR2B
ANK1
STRBP
LPAR1
LINC00332
TBX20
CPA3
JAG1
ESR1
ATAD2B
ACTL9
CLDN7
MAFB
GPR139
MLXIPL
SLC45A4
ITPR2
ADRB1
CCNL1
TRIB1
CDKAL1
HHEX-IDE
ADCY5
EBF1
PTH1R
KREMEN1
NRIP1
FES
NT5C2
MTNR1B
SBP
Mean_Arterial_Pressure
DBP
Pulse_Pressure
CAD
Adult_WHRadjBMI
Chol
LDL
TG
T2D
Fasting_glucose
Proinsulin
Fasting_insulin
Tanner_stage
Endometriosis_all
Endometriosis_StageB
HDL
Adiponectin
Birth_Length
Age_at_Menarche
Pubertal_Timing
Pubertal_Growth
Adult_BMI
Childhood_Height
Childhood_Obesity
Infant_HC
Adult_Height

Figure 3