



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genome-wide associations for birth weight and correlations with adult disease

Citation for published version:

Horikoshi, M, Beaumont, RN, Day, FR, Warrington, NM, Kooijman, MN, Fernandez-tajes, J, Feenstra, B, Van Zuydam, NR, Gaulton, KJ, Grarup, N, Bradfield, JP, Strachan, DP, Li-gao, R, Ahluwalia, TS, Kreiner, E, Rueedi, R, Lyytikäinen, L, Cousminer, DL, Wu, Y, Thiering, E, Wang, CA, Have, CT, Hottenga, J-J, Vilor-Tejedor, N, Joshi, PK, Boh, ETH, Ntalla, I, Pitkänen, N, Mahajan, A, van Leeuwen, EM, Joro, R, Lagou, V, Nodzenski, M, Diver, LA, Zondervan, KT, Bustamante, M, Marques-Vidal, P, Mercader, JM, Bennett, AJ, Rahmioglu, N, Nyholt, DR, Ma, RCW, Tam, CHT, Tam, WH, CHARGE Consortium Hematology Working Group, Ganesh, SK, Van Rooij, FJA, Reynolds, RM, Early Growth Genetics (EGG) consortium, Campbell, H, Wilson, JF & Walker, BR 2016, 'Genome-wide associations for birth weight and correlations with adult disease', *Nature*, vol. 538, no. 7624, pp. 248-252. <https://doi.org/10.1038/nature19806>

Digital Object Identifier (DOI):

[10.1038/nature19806](https://doi.org/10.1038/nature19806)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature

Publisher Rights Statement:

Author's final peer-reviewed manuscript

General rights

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

Take down policy

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



Genome-wide associations for birth weight and correlations 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 Feenstra¹⁰, Natalie R van Zuydam^{1,2}, Kyle J Gaulton^{1,11}, Niels Grarup¹², Jonathan P Bradfield¹³, David P Strachan¹⁴, Ruifang Li-Gao¹⁵, Tarunveer S Ahluwalia^{12,16,17}, Eskil Kreiner¹⁶, Rico Rueedi^{18,19}, Leo-Pekka Lyytikäinen^{20,21}, Diana L Cousminer^{22,23,24}, Ying Wu²⁵, Elisabeth Thiering^{26,27}, Carol A Wang⁶, Christian T Have¹², Jouke-Jan Hottenga²⁸, Natalia Vilor-Tejedor^{29,30,31}, Peter K Joshi³², Eileen Tai Hui Boh³³, Ioanna Ntalla^{34,35}, Niina Pitkänen³⁶, Anubha Mahajan¹, Elisabeth M van Leeuwen⁸, Raimo Joro³⁷, Vasiliki Lagou^{1,38,39}, Michael Nodzenski⁴⁰, Louise A Diver⁴¹, Krina T Zondervan^{1,42}, Mariona Bustamante^{29,30,31,43}, Pedro Marques-Vidal⁴⁴, Josep M Mercader⁴⁵, Amanda J Bennett², Nilufer Rahmioglu¹, Dale R Nyholt⁴⁶, Ronald Ching Wan Ma^{47,48,49}, Claudia Ha Ting Tam⁴⁷, Wing Hung Tam⁵⁰, CHARGE Consortium Hematology Working Group, Santhi K Ganesh⁵¹, Frank JA van Rooij⁸, Samuel E Jones³, Po-Ru Loh^{52,53}, Katherine S Ruth³, Marcus A Tuke³, Jessica Tyrrell^{3,54}, Andrew R Wood³, Hanieh Yaghoobkar³, Denise M Scholtens⁴⁰, Lavinia Paternoster^{55,56}, Inga Prokopenko^{1,57}, Peter Kovacs⁵⁸, Mustafa Atalay³⁷, Sara M Willems⁸, Kalliope Panoutsopoulou⁵⁹, Xu Wang³³, Lisbeth Carstensen¹⁰, Frank Geller¹⁰, Katharina E Schraut³², Mario Murcia^{31,60}, Catharina EM van Beijsterveldt²⁸, Gonneke Willemsen²⁸, Emil V R Appel¹², Cilius E Fonvig^{12,61}, Caecilie Trier^{12,61}, Carla MT Tiesler^{26,27}, Marie Standl²⁶, Zoltán Kutalik^{19,62}, Sílvia Bonas-Guarch⁴⁵, David M Hougaard^{63,64}, Friman Sánchez^{45,65}, David Torrents^{45,66}, Johannes Waage¹⁶, Mads V Hollegaard^{63,64,†}, Hugoline G de Haan¹⁵, Frits R Rosendaal¹⁵, Carolina Medina-Gomez^{7,8,67}, Susan M Ring^{55,56}, Gibran Hemani^{55,56}, George McMahon⁵⁶, Neil R Robertson^{1,2}, Christopher J Groves², Claudia Langenberg⁴, Jian'an Luan⁴, Robert A Scott⁴, Jing Hua Zhao⁴, Frank D Mentch¹³, Scott M MacKenzie⁴¹, Rebecca M Reynolds⁶⁸, William L Lowe Jr⁶⁹, Anke Tönjes⁷⁰, Michael Stumvoll^{58,70}, Virpi Lindi³⁷, Timo A Lakka^{37,71,72}, Cornelia M van Duijn⁸, Wieland Kiess⁷³, 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 Zeggini⁵⁹, George V Dedoussis³⁵, Yik-Ying Teo^{33,80,81}, Seang-Mei Saw^{33,82}, Mads Melbye^{10,83,84}, Harry Campbell³², James F Wilson^{32,85}, Martine Vrijheid^{29,30,31}, Eco JCN de Geus^{28,86}, Dorret I Boomsma²⁸, Haja N Kadarmideen⁸⁷, Jens-Christian Holm^{12,61}, Torben Hansen¹², Sylvain Sebert^{88,89}, Andrew T Hattersley³, Lawrence J Beilin⁹⁰, John P Newnham⁶, Craig E Pennell⁶, Joachim Heinrich^{26,91}, Linda S Adair⁹², Judith B Borja^{93,94}, Karen L Mohlke²⁵, Johan G Eriksson^{95,96,97}, Elisabeth E Widén²², Mika Kähönen^{98,99}, Jorma S Viikari^{100,101}, Terho Lehtimäki^{20,21}, Peter Vollenweider⁴⁴, Klaus Bønnelykke¹⁶, Hans Bisgaard¹⁶, Dennis O Mook-Kanamori^{15,102,103}, Albert Hofman^{7,8}, Fernando Rivadeneira^{7,8,67}, André G Uitterlinden^{7,8,67}, Charlotta Pisinger¹⁰⁴, Oluf Pedersen¹², Christine Power¹⁰⁵, Elina Hyppönen^{105,106,107}, Nicholas J Wareham⁴, Hakon Hakonarson^{13,23,108}, Eleanor Davies⁴¹, Brian R Walker⁶⁸, 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 Frayling³, 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 Hospital, 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.

1 **Birth weight (BW) is influenced by both foetal and maternal factors and in observational studies is**
 2 **reproducibly associated with future risk of adult metabolic diseases including type 2 diabetes**
 3 **(T2D) and cardiovascular disease¹. These lifecourse associations have often been attributed to the**
 4 **impact of an adverse early life environment. We performed a multi-ancestry genome-wide**
 5 **association study (GWAS) meta-analysis of BW in 153,781 individuals, identifying 60 loci where**
 6 **foetal genotype was associated with BW ($P < 5 \times 10^{-8}$). Overall, ~15% of variance in BW could be**
 7 **captured by assays of foetal genetic variation. Using genetic association alone, we found strong**
 8 **inverse genetic correlations between BW and systolic blood pressure ($r_g = -0.22$, $P = 5.5 \times 10^{-13}$), T2D**
 9 **($r_g = -0.27$, $P = 1.1 \times 10^{-6}$) and coronary artery disease ($r_g = -0.30$, $P = 6.5 \times 10^{-9}$) and, in large cohort data**
 10 **sets, demonstrated that genetic factors were the major contributor to the negative covariance**
 11 **between BW and future cardiometabolic risk. Pathway analyses indicated that the protein**
 12 **products of genes within BW-associated regions were enriched for diverse processes including**
 13 **insulin signalling, glucose homeostasis, glycogen biosynthesis and chromatin remodelling. There**
 14 **was also enrichment of associations with BW in known imprinted regions ($P = 1.9 \times 10^{-4}$). We have**
 15 **demonstrated that lifecourse associations between early growth phenotypes and adult**
 16 **cardiometabolic disease are in part the result of shared genetic effects and have highlighted some**
 17 **of the pathways through which these causal genetic effects are mediated.**

18
 19 We combined GWAS data for BW in 153,781 individuals representing multiple ancestries from 37
 20 studies across three components (**Extended Data Fig. 1 and Supplementary Table 1**): (i) 75,891
 21 individuals of European ancestry from 30 studies; (ii) 67,786 individuals of European ancestry from
 22 the UK Biobank; and (iii) 10,104 individuals of diverse ancestries (African American, Chinese, Filipino,
 23 Surinamese, Turkish and Moroccan) from six studies. Within each study, BW was z-score
 24 transformed separately in males and females after excluding non-singletons and premature births
 25 and adjusting for gestational age where available. Genotypes were imputed using reference panels
 26 from the 1000 Genomes (1000G)² or combined 1000G and UK10K Project³ (**Supplementary Table 2**).
 27 We performed quality control assessments to confirm that the distribution of BW was consistent
 28 across studies, irrespective of the data collection protocol, and confirmed that self-reported BW in
 29 UK Biobank showed genetic and phenotypic associations consistent with those seen for measured
 30 BW in other studies⁴ (**Methods**).

31
 32 We identified 60 loci (59 autosomal) associated with BW at genome-wide significance ($P < 5 \times 10^{-8}$) in
 33 either the European ancestry or trans-ancestry meta-analyses (**Extended Data Fig. 2a, Extended**
 34 **Data Table 1a and Supplementary Data; Methods**). At lead SNPs, we observed no heterogeneity in
 35 allelic effects between the three study components (Cochran's Q statistic $P > 0.00083$)
 36 (**Supplementary Table 3**). Fifty-three of these loci were novel in that the lead SNP mapped >2 Mb
 37 away from, and was statistically independent (EUR $r^2 < 0.05$) of, the seven previously-reported BW
 38 signals⁵, all of which were confirmed in this larger analysis (**Supplementary Table 4**). Approximate
 39 conditional analysis in the European ancestry data indicated that three of these novel loci (near
 40 *ZBTB7B*, *HMG1* and *PTCH1*) harboured multiple distinct association signals attaining genome-wide
 41 significance (**Methods; Supplementary Table 5 and Extended Data Fig. 3**).

42
 43 The lead variants for most signals mapped to non-coding sequence, and at only two loci, *ADRB1*
 44 (rs7076938; $r^2 = 0.99$ with *ADRB1* G389R) and *NR1P1* (rs2229742, R448G) did the association data
 45 point to likely causal non-synonymous coding variants (**Supplementary Table 6; Methods**). Lead
 46 SNPs for all but two loci (those mapping near *YKT6-GCK* and *SUZ12P1-CRLF3*) were common (minor
 47 allele frequency (MAF) $\geq 5\%$) with individually modest effects on BW ($\beta = 0.020$ - 0.053 standard
 48 deviations (SD) per allele, equivalent to 10-26g). This was despite much improved coverage of low-
 49 frequency variants in this study (compared to previous HapMap 2 imputed meta-analyses⁵)
 50 reflecting imputation from larger, and more complete, reference panels (**Extended Data Table 1b**).
 51 Indeed, all but five of the common variant association signals were tagged by variants (EUR $r^2 > 0.6$) in

52 the HapMap 2 reference panel (**Supplementary Tables 4, 5**), indicating that most of the novel
 53 discovery in the present study was driven by increased sample size⁵. Fine-mapping analysis yielded
 54 14 regions within which fewer than ten variants contributed to the locus-specific credible set that
 55 accounted for >99% of the posterior probability of association (**Methods; Supplementary Table 7**).
 56 The greatest refinement was at *YKT6-GCK*, where the credible set included only the low frequency
 57 variant rs138715366, which maps intronic to *YKT6*. These credible set variants collectively showed
 58 enrichment for overlap with DNaseI hypersensitivity sites, particularly those generated, by ENCODE,
 59 from foetal (4.2-fold, 95% CI [1.8-10.7]) and neonatal tissues (4.9 [1.8-11.0]) (**Supplementary Fig. 1**
 60 **and Supplementary Table 8; Methods**).

61
 62 In combination, the 62 distinct genome-wide significant signals at the 59 autosomal loci explained
 63 2.0% (standard error (SE) 1.1%) of variance in BW (**Supplementary Table 9; Methods**), similar in
 64 magnitude to that attributable to sex or maternal body mass index (BMI)⁵. However, the variance in
 65 BW captured collectively by all autosomal genotyped variants on the array was considerably larger,
 66 estimated at 15.1% (SE=0.9) in UK Biobank (**Methods**). These figures are consistent with a long tail of
 67 genetic variants of smaller effects contributing to variation in BW.

68
 69 Associations between foetal genotype and BW could result from indirect effects of the maternal
 70 genotype influencing BW via the intrauterine environment given the correlation ($r \approx 0.5$) between
 71 maternal and foetal genotype. However, two lines of evidence indicated that variation in the foetal
 72 genome was the predominant driver of the BW associations. First, an analysis of the global
 73 contribution of maternal vs. foetal genetic variation, using a maternal-GCTA model⁶ (**Methods**)
 74 applied to 4,382 mother-child pairs, estimated that the child's genotype ($\sigma_c^2=0.24$, SE=0.11) makes a
 75 larger contribution to BW variance than either the mother's genotype ($\sigma_M^2=0.04$, SE=0.10), or the
 76 covariance between the two ($\sigma_{CM}=0.04$, SE=0.08). Second, when we compared the point estimates of
 77 the BW effect size dependent on maternal genotype at each of the 60 loci (as measured in up to
 78 68,254 women⁷) with those dependent on foetal genotype (using European ancestry data from
 79 143,677 individuals in the present study), foetal variation had greater impact than maternal at 93%
 80 of loci (55/60; binomial $P=1 \times 10^{-11}$) (**Supplementary Table 10, Extended Data Figs 4, 5; Methods**).
 81 Power to further disentangle maternal and foetal contributions using analyses of foetal genotype
 82 conditional on maternal genotype was constrained by the limited sample size available (n=12,909
 83 mother-child pairs) (**Supplementary Table 11**).

84
 85 Collectively, these analyses provide compelling evidence that foetal genotype has a substantial
 86 impact on early growth, as measured by BW. We sought to use these genetic associations to
 87 understand the causal relationships underlying observed associations between BW and disease, and
 88 to characterise the processes responsible.

89
 90 To quantify the shared genetic contribution to BW and other health-related traits, we estimated
 91 their genetic correlations using LD Score regression⁸ (**Methods**). BW (in European ancestry samples)
 92 showed strong positive genetic correlations with anthropometric and obesity-related traits including
 93 birth length ($r_g=0.81$, $P=2.0 \times 10^{-44}$), and in adults, height ($r_g=0.41$, $P=4.8 \times 10^{-52}$), waist circumference
 94 ($r_g=0.18$, $P=3.9 \times 10^{-10}$) and BMI ($r_g=0.11$, $P=7.3 \times 10^{-6}$). In contrast, BW showed inverse genetic
 95 correlations with indicators of adverse metabolic and cardiovascular health including coronary artery
 96 disease (CAD, $r_g=-0.30$, $P=6.5 \times 10^{-9}$), systolic blood pressure (SBP, $r_g=-0.22$, $P=5.5 \times 10^{-13}$) and T2D ($r_g=-$
 97 0.27 , $P=1.1 \times 10^{-6}$) (**Fig. 1, Supplementary Table 12 and Supplementary Fig. 2**). These correlations
 98 between BW and adult cardiometabolic phenotypes are of similar magnitude, although directionally-
 99 opposite, to the reported genetic correlations between adult BMI and those same cardiometabolic
 100 outcomes⁸. These findings support observational associations between a history of paternal T2D and
 101 lower BW⁴, and establish more generally that the observed lifecourse associations between early
 102 growth and adult disease, at least in part, reflect the impact of shared genetic variants that influence

103 both sets of phenotypes. In an effort to estimate the extent of genetic contribution to these
 104 lifecourse associations, we first focused on data from UK Biobank (n=57,715). For many of the traits
 105 for which data were available, genetic variation contributed substantially to the lifecourse
 106 relationship between BW and adult phenotypes, and in some cases appeared to be the major source
 107 of covariance between the traits. For example, we estimated that 85% (95% CI=70%-99%) of the
 108 negative covariance between BW and SBP was explained by shared genetic associations captured by
 109 directly genotyped SNPs (**Supplementary Table 13; Methods**). For continuous cardiometabolic
 110 measures, including lipids and fasting glycaemia, for which measures are not currently available in
 111 UK Biobank, we turned to the Northern Finland Birth Cohort (n=5,009), and obtained similar results
 112 (**Supplementary Table 13**). However, these estimates are limited, not only by wide confidence
 113 intervals, but also by the assumption of a linear relationship between BW and each of the
 114 phenotypes and by the inability to explicitly model maternal genotypic effects. In other words, the
 115 inverse genetic correlations between BW and cardiometabolic traits may not exclusively reflect
 116 genetic effects mediated directly through the offspring, but also effects mediated by maternal
 117 genotype acting indirectly via perturbation of the *in utero* environment. Nevertheless, these
 118 estimates indicate that a substantial proportion of the variance in cardiometabolic risk that covaries
 119 with BW can be attributed to the effects of common genetic variation.

120

121 To elucidate the biological pathways and processes underlying regulation of foetal growth, we first
 122 performed gene set enrichment analysis of our BW GWAS analysis using MAGENTA⁹ (**Methods**).
 123 Twelve pathways reached study-wide significance (FDR<0.05), including pathways involved in
 124 metabolism (insulin signalling, glycogen biosynthesis, cholesterol biosynthesis), growth (IGF-
 125 signalling, growth hormone pathway) and development (chromatin remodelling) (**Extended Data**
 126 **Table 2a**). Similar pathways were detected in a complementary analysis where we interrogated
 127 empirical protein-protein interaction (PPI) data identifying 13 PPI network modules with marked (z-
 128 score >5) enrichment for BW-association scores (**Extended Data Table 2b and Extended Data Figs**
 129 **6a, b; Methods**). The proteins within these modules were themselves enriched for diverse processes
 130 related to metabolism, growth and development (**Extended Data Figs 6a, b**).

131

132 We also observed enrichment of BW association signals across the set of 77 imprinted genes defined
 133 by the Genotype-Tissue Expression (GTEx) project¹⁰ ($P=1.9 \times 10^{-4}$; **Extended Data Table 2a and**
 134 **Supplementary Table 14**). Such enrichment is consistent with the “parental conflict” hypothesis
 135 regarding the allocation of maternal resources to the foetus¹¹. Although the role of imprinted genes
 136 in foetal growth is described in animal models and rare human disorders¹², our result is the first
 137 large-scale, systematic demonstration of their contribution to normal variation in BW. Of the 60
 138 genome-wide significant loci, two (*INS-IGF2*, *RB1*) fall within (or near) imprinted regions (**Extended**
 139 **Data Fig. 2b**), with a noteworthy third signal at *DLK1* (previously foetal antigen-1; $P=5.6 \times 10^{-8}$).
 140 Parent-of-origin specific analyses to further investigate these individual loci (comparing
 141 heterozygote vs. homozygote BW variance in 57,715 unrelated individuals, and testing BW
 142 associations with paternal vs. maternal alleles in 4,908 mother-child pairs; see **Methods**) proved,
 143 despite these sample sizes, to be underpowered (**Extended Data Fig. 7 and Supplementary Tables**
 144 **15, 16**).

145

146 Many of the genome-wide signals for BW detected here are also established genome-wide
 147 association signals for a wide variety of cardiometabolic traits (**Fig. 2**). These include the BW signals
 148 near *CDKAL1*, *ADCY5*, *HHEX/IDE* and *ANK1* (also genome-wide significant for T2D), *NT5C2* (for blood
 149 pressure (BP), CAD and BMI) and *ADRB1* (for BP). We used two approaches to understand whether
 150 this pattern of adult trait association represented a generic property of BW-associated loci, or
 151 reflected heterogeneous mechanisms linking BW to adult disease.

152

153 First, we applied unsupervised hierarchical clustering (**Methods**) to the non-BW trait association
 154 statistics for the 60 significant BW loci. The resultant heatmap indicates the heterogeneity of locus-
 155 specific effect sizes across the range of adult traits (**Fig. 2 and Supplementary Table 17**). For
 156 example, it shows that the associations between BW-raising alleles and increased adult height are
 157 concentrated amongst a subset of loci including *HHIP* and *GNA12*, and highlights particularly strong
 158 associations with lipid traits for variants at the *TRIB1* and *MAFB* loci.

159
 160 Second, we constructed trait-specific “point-of-contact” (PoC) PPI networks from proteins
 161 represented in both the global BW PPI network and equivalent PPI networks generated for each of
 162 the adult traits (**Methods; Extended Data Figs 6c-e**). We reasoned that these PoC PPI networks
 163 would be enriched for the specific proteins mediating the observed links between BW and adult
 164 traits, generating hypotheses that are amenable to subsequent empirical validation. To highlight
 165 processes implicated in specific BW-trait associations, we overlaid these PoC PPI with the 50
 166 pathways over-represented in the global BW PPI network. These analyses revealed, for example,
 167 that proteins in the Wnt canonical signalling pathway were only detected in the PoC PPI network for
 168 BP traits. We can use these PPI overlaps to highlight the specific transcripts within BW GWAS loci
 169 that are likely to mediate the mechanistic links. For example, the overlap between the Wnt signalling
 170 pathway and the PoC PPI network for the intersection of BW and BP-related traits implicates *FZD9* as
 171 the likely effector gene at the *MLXIPL* BW locus (**Extended Data Fig. 6d and Supplementary Table 6**).

172
 173 We focused our more detailed investigation of the mechanistic links between early growth and adult
 174 traits on two phenotypic areas: arterial BP and T2D/glycaemia. Across both the overall GWAS and
 175 specifically among the 60 significant BW loci, most BW-raising alleles were associated with reduced
 176 BP (**Figs 1, 2**): the strongest inverse associations were seen for the loci near *NT5C2*, *FES*, *NRIP1*, *EBF1*
 177 and *PTH1R*. However, we also observed locus-specific heterogeneity in the genetic relationships
 178 between BP and BW: the SBP-raising allele at *ADRB1*¹³ is associated with higher, rather than lower,
 179 BW (**Extended Data Fig. 8a**). When we considered the reciprocal relationship, i.e. the effects on BW
 180 of BP-raising alleles at 30 reported loci for SBP^{13,14}, there was an excess of associations (5/30 with
 181 lower BW at $P < 0.05$; $P = 0.0026$; **Extended Data Fig. 8a**). To dissect maternal and foetal genotype
 182 effects at these loci, we tested the impact on BW of a risk score generated from the 30 SBP SNPs,
 183 restricted to the untransmitted maternal haplotype score¹⁵ in a set of 5,201 mother-child pairs.
 184 Analysis of these loci indicated that maternal genotype effects on the intrauterine environment are
 185 likely to contribute to the inverse genetic correlation between SBP and BW (**Methods**;
 186 **Supplementary Table 18**), and was consistent with the results of a wider study of >30,000 women
 187 which demonstrated associations between a maternal genetic score for SBP (conditional on foetal
 188 genotype) and lower offspring BW¹⁶.

189
 190 The BP-raising allele with the largest BW-lowering effect maps to the *NT5C2* locus (index variant for
 191 BW, rs74233809, $r^2 = 0.98$ with index variant for BP, rs11191548¹⁴) and is also associated with lower
 192 adult BMI ($r^2 = 0.99$ with rs11191560¹⁷). The BW-lowering allele at rs74233809 is a proxy for a
 193 recently-described functional variant in the nearby *CYP17A1* gene ($r^2 = 0.92$ with rs138009835)¹⁸. The
 194 *CYP17A1* gene encodes the cytochrome P450c17 α enzyme, CYP17¹⁹, which catalyses key steps in
 195 steroidogenesis that determine the balance between mineralocorticoid, glucocorticoid and
 196 androgen synthesis. This variant is known to alter transcriptional efficiency *in vitro* and is associated
 197 with higher urinary tetrahydroaldosterone excretion¹⁸. *CYP17A1* is expressed in foetal adrenal glands
 198 and testes from early gestation²⁰ as well as in the placenta²¹. These data implicate variation in
 199 *CYP17A1* expression as a contributor to the observational association between low BW and adult
 200 hypertension²².

201
 202 When we examined 45 loci associated with CAD²³, the inverse genetic correlation between CAD and
 203 BW was concentrated amongst the five CAD loci with primary BP associations. This suggests that

204 genetic determinants of BP play a leading role in mediating the lifecourse associations between BW
 205 and CAD (**Extended Data Figs 8b, e**).

206

207 LD score regression analyses demonstrated overall inverse genetic correlation between lower BW
 208 and elevated risk of T2D (**Fig. 1**). However, the locus specific heatmap indicates a heterogeneous
 209 pattern across individual loci (**Fig. 2**). To explore this further, we tested the 84 reported T2D loci²⁴ for
 210 association with BW. Some T2D risk alleles (such as at *ADCY5*, *CDKAL1* and *HHEX-IDE*) were strongly
 211 associated with lower BW, whilst others (e.g. *ANK1* and *MTNR1B*) were associated with higher BW
 212 (**Extended Data Fig. 8c**). This was in contrast with the BW effects of 422 known height loci²⁵
 213 (**Extended Data Fig. 8d**), which showed a strong positive correlation consistent with the overall
 214 genetic correlation between height and BW, indicating that the growth effects of many height loci
 215 start prenatally and persist into adulthood.

216

217 The contrasting associations of T2D risk-alleles with both higher and lower BW are likely to reflect
 218 the differential impacts across loci of variation in the maternal and foetal genomes. Observational
 219 data link paternal diabetes with lower offspring BW⁴ indicating that the inheritance of T2D risk
 220 alleles by the foetus tends, in line with the LD score regression analysis, to reduce growth. These
 221 relationships are consistent with the precepts of the “foetal insulin hypothesis”²⁶ and reflect the
 222 potential for reduced insulin secretion and/or signalling to lead to both reduced foetal growth and,
 223 many decades later, enhanced predisposition to T2D. In line with this, the inferred paternal
 224 transmitted haplotype score generated from the 84 T2D risk variants was associated with lower BW
 225 ($P=0.045$) in 5,201 mother-child pairs (**Methods; Supplementary Table 18**). In contrast, maternal
 226 diabetes is observationally associated with higher offspring BW⁴, reflecting the impact of maternal
 227 hyperglycaemia to stimulate foetal insulin secretion. The contribution of genotype-dependent
 228 maternal hyperglycaemia to BW is in line with the evidence, from a recent study, that maternal
 229 genotype scores for fasting glucose and T2D (conditional on foetal genotype) were causally
 230 associated with higher offspring BW¹⁶. It is also consistent with the observation that a subset of
 231 glucose-raising alleles is associated with higher BW⁷. For example, the T2D-risk variant at *MTNR1B*
 232 (which also has a particularly marked effect on fasting glucose levels in non-diabetic individuals^{27,28})
 233 was amongst the subset of BW loci (5/60) for which the BW effect attributable to maternal genotype
 234 exceeded that associated with the foetal genotype (maternal: $\beta=0.048$, $P=5.1 \times 10^{-15}$; foetal: $\beta=0.023$,
 235 $P=2.9 \times 10^{-8}$) (**Supplementary Table 10, Extended Data Figs 4, 5**). Thus, both maternal and foetal
 236 genetic effects connect BW to later T2D risk, albeit acting in opposing directions. When we
 237 categorised T2D loci using a classification of physiological function derived from their effects on
 238 related glycaemic and anthropometric traits²⁷, we found that T2D-risk alleles associated with lower
 239 BW were those typically characterised by reduced insulin processing and secretion without
 240 detectable changes in fasting glucose (the “Beta Cell” cluster in **Extended Data Fig. 8f**).

241

242 The *YTK6* signal at rs138715366 is notable, not only because the genetic data indicates that a single
 243 low-frequency non-coding variant is driving the association signal (see above) but because of the
 244 proximity of this signal to *GCK*. Rare coding variants in glucokinase are causal for a form of
 245 monogenic hyperglycaemia and lead to large reductions in BW when parental alleles are passed to
 246 their offspring²⁹. In addition, common non-coding variants nearby are implicated in T2D-risk and
 247 fasting hyperglycaemia²⁸. However, the latter variants are conditionally independent of rs138715366
 248 (**Supplementary Table 19**) and show no comparable association with lower BW. Either rs138715366
 249 acts through effector transcripts other than *GCK*, or the impact of the low-frequency SNP near *YTK6*
 250 on *GCK* expression involves tissue- and/or temporal-specific variation in regulatory impact.

251

252 In conclusion, we have identified 60 genetic loci associated with BW and used these to gain insights
 253 into the aetiology of foetal growth and into well-established, but until now poorly understood,
 254 lifecourse disease associations. The evidence that the relationship between early growth and later

255 metabolic disease has an appreciable genetic component contrasts with, but is not necessarily
256 incompatible with, the emphasis on adverse early environmental events highlighted by the Foetal
257 Origins Hypothesis¹. As we have shown, these genetic effects reflect variation in both the foetal and
258 the maternal genome: the impact of the latter on the offspring's predisposition to adult disease
259 could be mediated, at least in part, through perturbation of the antenatal and early life
260 environment. Future mechanistic and genetic studies should support reconciliation between these
261 alternative, but complementary, explanations for the far-reaching lifecourse associations that exist
262 between events in early life and predisposition to cardiometabolic disease several decades later.
263
264
265

266 **REFERENCES**

- 267
- 268 1. Barker DJ. The developmental origins of chronic adult disease. *Acta Paediatr Suppl* **93**, 26-33
- 269 (2004).
- 270 2. The 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, *et al.* An
- 271 integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
- 272 3. The UK10K Project Consortium, Walter K, Min JL, Huang J, Crooks L, *et al.* The UK10K project
- 273 identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).
- 274 4. Tyrrell JS, Yaghootkar H, Freathy RM, Hattersley AT, Frayling TM. Parental diabetes and
- 275 birthweight in 236 030 individuals in the UK biobank study. *Int J Epidemiol* **42**, 1714-1723
- 276 (2013).
- 277 5. Horikoshi M, Yaghootkar H, Mook-Kanamori DO, Sovio U, Taal HR, *et al.* New loci associated
- 278 with birth weight identify genetic links between intrauterine growth and adult height and
- 279 metabolism. *Nat Genet* **45**, 76-82 (2013).
- 280 6. Eaves LJ, Pourcain BS, Smith GD, York TP, Evans DM. Resolving the effects of maternal and
- 281 offspring genotype on dyadic outcomes in genome wide complex trait analysis ("M-GCTA").
- 282 *Behav Genet* **44**, 445-455 (2014).
- 283 7. Feenstra B, Beaumont RN, Cavadino A, Tyrrell J, McMahon G, *et al.* Maternal genome-wide
- 284 association study identifies a fasting glucose variant associated with offspring birth weight. (in
- 285 preparation) doi: <http://dx.doi.org/10.1101/034207>
- 286 8. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, *et al.* LD Score regression distinguishes
- 287 confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295
- 288 (2015).
- 289 9. Segrè AV, DIAGRAM Consortium, MAGIC investigators, Groop L, Mootha VK, *et al.* Common
- 290 inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes
- 291 or related glycemc traits. *PLoS Genet* **6**, e1001058 (2010).
- 292 10. Baran Y, Subramaniam M, Biton A, Tukiainen T, Tsang EK, *et al.* The landscape of genomic
- 293 imprinting across diverse adult human tissues. *Genome Res* **25**, 927-936 (2015).
- 294 11. Haig D and Westoby M. Parent-Specific Gene Expression and the Triploid Endosperm. *The*
- 295 *American Naturalist* **134**, 147-155 (1989).
- 296 12. Peters J. The role of genomic imprinting in biology and disease: an expanding view. *Nat Rev*
- 297 *Genet* **15**, 517-530 (2014).
- 298 13. Johnson T, Gaunt TR, Newhouse SJ, Padmanabhan S, Tomaszewski M, *et al.* Blood pressure loci
- 299 identified with a gene-centric array. *Am J Hum Genet* **89**, 688-700 (2011).
- 300 14. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB,
- 301 Munroe PB, Rice KM, Bochud M, *et al.* Genetic variants in novel pathways influence blood
- 302 pressure and cardiovascular disease risk. *Nature* **478**, 103-109 (2011).
- 303 15. Zhang G, Bacelis J, Lengyel C, Teramo K, Hallman M, *et al.* Assessing the Causal Relationship of
- 304 Maternal Height on Birth Size and Gestational Age at Birth: A Mendelian Randomization
- 305 Analysis. *PLoS Med* **12**, e1001865 (2015).
- 306 16. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, *et al.* Genetic evidence for causal
- 307 relationships between maternal obesity-related traits and birth weight. *JAMA* **315**, 1129-1140
- 308 (2016).
- 309 17. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, *et al.* Genetic studies of body mass index yield
- 310 new insights for obesity biology. *Nature* **518**, 197-206 (2015).
- 311 18. Diver LA, MacKenzie SM, Fraser R, McManus F, Freel EM, *et al.* Common Polymorphisms at the
- 312 *CYP17A1* locus associate with steroid phenotype: support for blood pressure GWAS signals at
- 313 this locus. *Hypertension* **67**, 724-732 (2016).
- 314 19. Picado-Leonard J, Miller WL. Cloning and sequence of the human gene for P450c17 (steroid 17
- 315 alpha-hydroxylase/17,20 lyase): similarity with the gene for P450c21. *DNA* **6**, 439-448 (1987).

- 316 20. Pezzi V, Mathis JM, Rainey WE, Carr BR. Profiling transcript levels for steroidogenic enzymes in
317 fetal tissues. *J Steroid Biochem Mol Biol* **87**, 181-189 (2003).
- 318 21. Escobar JC, Patel SS, Beshay VE, Suzuki T, Carr BR. The human placenta expresses CYP17 and
319 generates androgens *de novo*. *J Clin Endocrinol Metab* **96**, 1385-1392 (2011).
- 320 22. Reynolds RM, Walker BR, Phillips DI, Dennison EM, Fraser R, *et al*. Programming of
321 hypertension: associations of plasma aldosterone in adult men and women with birth weight,
322 cortisol, and blood pressure. *Hypertension* **53**, 932-936 (2009).
- 323 23. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, *et al*. Large-
324 scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25-
325 33 (2013).
- 326 24. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic
327 Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes
328 (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes
329 Genetic Exploration by Nex-generation sequencing in muilti-Ethnic Samples (T2D-GENES)
330 Consortium, Mahajan A, *et al*. Genome-wide trans-ancestry meta-analysis provides insight into
331 the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-244 (2014).
- 332 25. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, *et al*. Defining the role of common variation in
333 the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-1186
334 (2014).
- 335 26. Hattersley AT, Tooke JE. The fetal insulin hypothesis: an alternative explanation of the
336 association of low birthweight with diabetes and vascular disease. *Lancet* **353**, 1789-1792
337 (1999).
- 338 27. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, *et al*. Impact of type 2 diabetes susceptibility
339 variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158-
340 2171 (2014).
- 341 28. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, *et al*. Large-scale association analysis
342 provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat*
343 *Genet* **44**, 981-990 (2012).
- 344 29. Hattersley AT, Beards F, Ballantyne E, Appleton M, Harvey R, *et al*. Mutations in the glucokinase
345 gene of the fetus result in reduced birth weight. *Nat Genet* **19**, 268-270 (1998).
- 346
- 347

348 **FIGURE LEGENDS**

349

350 **Figure 1 | Genome-wide genetic correlation between birth weight and a range of traits and**
351 **diseases in later life.** Genetic correlation (r_g) and corresponding standard error between BW and the
352 traits displayed on the x axis are estimated using LD Score regression⁸. The genetic correlation
353 estimates (r_g) are colour coded according to their intensity and direction (red for positive and blue
354 for inverse correlation). WHRadjBMI=waist-hip ratio adjusted for body mass index, HOMA-
355 B/IR=homeostatic model assessment of beta-cell function/insulin resistance, HbA1c=Hemoglobin
356 A1c, BMD=bone mineral density, ADHD=attention deficit hyperactivity disorder. See Supplementary
357 Table 12 for references for each of the traits and diseases displayed.

358

359 **Figure 2 | Hierarchical clustering of birth weight loci based on similarity of overlap with adult**
360 **diseases, metabolic and anthropometric traits.** For the lead SNP at each BW locus (x-axis), z-scores
361 (aligned to BW-raising allele) were obtained from publicly available GWAS for various traits (y-axis;
362 see Supplementary Table 17). A positive z-score (red) indicates a positive association between the
363 BW-raising allele and the outcome trait, while a negative z-score (blue) indicates an inverse
364 association. BW loci and traits are clustered according to the Euclidean distance amongst z-scores
365 (see Methods). Squares are outlined with a solid black line if the BW locus is significantly ($P < 5 \times 10^{-8}$)
366 associated with the trait in publicly available GWAS, or with a dashed line if reported significant
367 elsewhere. WHRadjBMI=waist-hip ratio adjusted for body mass index.

368

369 **METHODS**

370

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

375

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

383 Within each study, stringent quality control of the GWAS genotype scaffold was undertaken,
 384 prior to imputation (**Supplementary Table 2**). Each scaffold was then pre-phased and imputed^{30,31}
 385 up to reference panels from the 1000G² or 1000G and UK10K Project³ (**Supplementary Table 2**).
 386 Association of BW with each variant passing established GWAS quality control filters³² was tested in
 387 a linear regression framework, under an additive model for the allelic effect, after adjustment for
 388 study-specific covariates, including gestational age, where available (**Supplementary Table 2**).
 389 Where necessary, population structure was accounted for by adjustment for axes of genetic
 390 variation from principal components analysis³³ and subsequent genomic control correction³⁴, or
 391 inclusion of a genetic relationship matrix in a mixed model³⁵ (**Supplementary Table 2**). We calculated
 392 the genomic control inflation factor (λ) in each study to confirm that study-level population structure
 393 was accounted for prior to meta-analysis.

394

395 **Preparation, quality control and genetic analysis in UK Biobank samples.** UK Biobank phenotype
 396 data were available for 502,655 participants³⁶. All participants in the UK Biobank were asked to recall
 397 their BW, of which 279,971 did so at either the baseline or follow-up assessment visit. Of these,
 398 7,686 participants reported being part of multiple births and were excluded from downstream
 399 analyses. Ancestry checks, based on self-reported ancestry, resulted in the exclusion of 8,998
 400 additional participants reported not to be white European. Of those individuals reporting BW at
 401 baseline and follow-up assessments, 393 were excluded because the two reported values differed by
 402 more than 0.5 kg. For those reporting different values (≤ 0.5 kg) between baseline and follow-up, we
 403 took the baseline measure forward for downstream analyses. We then excluded 36,716 individuals
 404 reporting values <2.5 kg or >4.5 kg as implausible for live term births before 1970. In total 226,178
 405 participants had data relating to BW that matched these inclusion criteria.

406 Genotype data from the May 2015 release were available for a subset of 152,249
 407 participants from UK Biobank. In addition to the quality control metrics performed centrally by UK
 408 Biobank, we defined a subset of “white European” ancestry samples using a K-means (K=4)
 409 clustering approach based on the first four genetically determined principal components. A
 410 maximum of 67,786 individuals (40,425 females and 27,361 males) with genotype and valid BW
 411 measures were available for downstream analyses. We tested for association with BW, assuming an
 412 additive allelic effect, in a linear mixed model implemented in BOLT-LMM³⁷ to account for cryptic
 413 population structure and relatedness. Genotyping array was included as a binary covariate in all
 414 models. Total chip heritability (i.e. the variance explained by all autosomal polymorphic genotyped
 415 SNPs passing quality control) was calculated using Restricted Maximum Likelihood (REML)
 416 implemented in BOLT-LMM³⁷. We additionally analysed the association between BW and directly
 417 genotyped SNPs on the X chromosome: for this analysis, we used 57,715 unrelated individuals with
 418 BW available and identified by UK Biobank as white British. We excluded SNPs with evidence of
 419 deviation from Hardy-Weinberg Equilibrium ($P < 1 \times 10^{-6}$), MAF < 0.01 or overall missing rate > 0.015,

420 resulting in 19,423 SNPs for analysis in Plink v1.07 (<http://pngu.mgh.harvard.edu/purcell/plink/>)³⁸,
421 with the first five ancestry principal components as covariates.

422 In both the full UK Biobank sample and our refined sample, we observed that BW was
423 associated with sex, year of birth and maternal smoking ($P < 0.0015$, all in the expected directions),
424 confirming more comprehensive previous validation of self-reported BW⁴. We additionally verified
425 that BW associations with lead SNPs at seven established loci⁵ based on self-report in UK Biobank
426 were consistent with those previously published.

427

428 **European ancestry meta-analysis.** The European ancestry meta-analysis consisted of two
429 components: (i) 75,891 individuals from 30 GWAS from Europe, USA and Australia; and (ii) 67,786
430 individuals of white European origin from UK Biobank. In the first component, we combined sex-
431 specific BW association summary statistics across studies in a fixed-effects meta-analysis,
432 implemented in GWAMA³⁹ and applied a second round of genomic control³⁴ ($\lambda_{GC} = 1.001$).
433 Subsequently, we combined association summary statistics from this component with UK Biobank in
434 a European ancestry fixed-effects meta-analysis, implemented in GWAMA³⁹. Variants failing GWAS
435 quality control filters in UK Biobank, reported in less than 50% of the total sample size in the first
436 component, or with $MAF < 0.1\%$, were excluded from the European ancestry meta-analysis. We
437 aggregated X-Chromosome association summary statistics from UK Biobank (19,423 SNPs) with
438 corresponding statistics from the European GWAS component using fixed effects P -value based
439 meta-analysis in METAL⁴⁰ (max $N = 99,152$).

440 We were concerned that self-reported BW as adults in UK Biobank would not be comparable
441 with that obtained from more stringent collection methods used in other European ancestry GWAS.
442 In addition, UK Biobank lacked information on gestational age for adjustment, which could have an
443 impact on difference in strength of association compared to the results obtained from other
444 European ancestry GWAS. However, we observed no evidence of heterogeneity in BW allelic effects
445 at lead SNPs between the two components of European ancestry meta-analysis, using Cochran's Q
446 statistic⁴¹, implemented in GWAMA³⁹, after Bonferroni correction ($P > 0.00083$) (**Supplementary**
447 **Table 3**). We tested for heterogeneity in allelic effects between studies within the European
448 component using Cochran's Q . At loci demonstrating evidence of heterogeneity, we confirmed that
449 association signals were not being driven by outlying studies by visual inspection of forest plots. We
450 performed sensitivity analyses to assess the impact of covariate adjustment (gestational age and
451 population structure) on heterogeneity.

452 We were also concerned that overlap of individuals (duplicated or related) between the two
453 components of the European ancestry meta-analysis might lead to false positive association signals.
454 We performed bivariate LD Score regression⁸ using the two components of the European ancestry
455 meta-analysis and observed a genetic covariance intercept of 0.0156 (SE 0.0058), indicating a
456 maximum of 1,119 duplicate individuals. Univariate LD Score regression⁸ of the European ancestry
457 meta-analysis estimated the intercept as 1.0426, which may indicate population structure or
458 relatedness that is not adequately accounted for in the analysis. To assess the impact of this inflation
459 on the European ancestry meta-analysis, we expanded the standard errors of BW allelic effect size
460 estimates and re-calculated association P -values. On the basis of this adjusted analysis, the lead SNP
461 only at *MTNR1B* dropped below genome-wide significance ($rs10830963$, $P = 5.5 \times 10^{-8}$).

462

463 **Trans-ancestry meta-analysis.** The trans-ancestry meta-analysis combined the two European
464 ancestry components with an additional 10,104 individuals from six GWAS from diverse ancestry
465 groups: African American, Chinese, Filipino, Surinamese, Turkish and Moroccan. Within each GWAS,
466 we first combined sex-specific BW association summary statistics in a fixed-effects meta-analysis,
467 implemented in GWAMA³⁹ and applied a second round of genomic control³⁴. Subsequently, we
468 combined association summary statistics from the six non-European GWAS and the two European
469 ancestry components in a trans-ancestry fixed-effects meta-analysis, implemented in GWAMA³⁹.
470 Variants failing GWAS quality control filters in UK Biobank, reported in less than 50% of the total

471 sample size in the first component, or with $MAF < 0.1\%$, were excluded from the trans-ancestry meta-
 472 analysis. We tested for heterogeneity in allelic effects between ancestries using Cochran's Q^{41} .

473

474 **Approximate conditional analysis.** We searched for multiple distinct BW association signals in each
 475 of the established and novel loci, defined as 1Mb up- and down-stream of the lead SNP from the
 476 trans-ancestry meta-analysis, through approximate conditional analysis. We applied GCTA⁴² to
 477 identify "index SNPs" for distinct association signals attaining genome-wide significance ($P < 5 \times 10^{-8}$) in
 478 the European ancestry meta-analysis using a reference sample of 5,000 individuals of white British
 479 origin, randomly selected from UK Biobank, to approximate patterns of linkage disequilibrium (LD)
 480 between variants in these regions. Note that we performed approximate conditioning on the basis of
 481 only the European ancestry meta-analysis because GCTA cannot accommodate LD variation between
 482 diverse populations.

483

484 **Prioritising candidate genes in each BW locus.** We combined a number of approaches to prioritise
 485 the most likely candidate gene(s) in each BW locus. Expression quantitative trait loci (eQTLs) were
 486 obtained from the Genotype Tissue Expression (GTEx) Project⁴³, the GEUVADIS Project⁴⁴ and eleven
 487 other studies⁴⁵⁻⁵⁵ using HaploReg v4⁵⁶. We interrogated coding variants for each BW lead SNP and its
 488 proxies (EUR $r^2 > 0.8$) using Ensembl⁵⁷ and HaploReg. Their likely functional consequences were
 489 predicted by SIFT⁵⁸ and PolyPhen2⁵⁹. Biological candidacy was assessed by presence in significantly
 490 enriched gene set pathways from MAGENTA analyses (see below for details). We extracted all genes
 491 within 300 kb of all lead BW SNPs and searched for connectivity between any genes using STRING⁶⁰.
 492 If two or more genes between two separate BW loci were connected, they were given an increased
 493 prior for both being plausible candidates. We also applied protein-protein interaction (PPI) analysis
 494 (see below for details) to all genes within 300 kb of each lead BW SNPs and ranked the genes based
 495 on the score for connectivity with the surrounding genes.

496

497 **Evaluation of imputation quality of low-frequency variant at the *YKT6-GCK* locus.** At the *YKT6-GCK*
 498 locus, the lead SNP (rs138715366) is of low-frequency in European ancestry populations
 499 ($MAF = 0.92\%$) and even rarer in other ancestry groups ($MAF = 0.23\%$ in African Americans, otherwise
 500 monomorphic) and is not present in the HapMap reference panel⁶¹. To assess the accuracy of
 501 imputation for this low-frequent variant, we genotyped rs138715366 in the Northern Finland Birth
 502 Cohort (NFBC) 1966 (**Supplementary Table 1**). Of the 5,009 samples in the study, 4,704 were
 503 successfully imputed and genotyped (or sequenced) for rs138715366. The overall concordance rate
 504 between imputed and directly assayed genotypes was 99.8% and for directly assayed heterozygote
 505 calls was 75.0%.

506

507 **Fine-mapping analyses.** We sought to leverage LD differences between populations contributing to
 508 the trans-ancestry meta-analysis and to take advantage of the improved coverage of common and
 509 low-frequency variation offered by 1000G or 1000G and UK10K combined imputation to localise
 510 variants driving each distinct association signal achieving locus-wide significance. For each distinct
 511 signal, we used MANTRA⁶² to construct 99% credible sets of variants⁶³ that together account for 99%
 512 of the posterior probability of driving the association. MANTRA incorporates a prior model of
 513 relatedness between studies, based on mean pair-wise allele frequency differences across loci, to
 514 account for heterogeneity in allelic effects (**Supplementary Table 3**). MANTRA has been
 515 demonstrated, by simulation, to improve localisation of causal variants compared with either a
 516 fixed- or random-effects trans-ancestry meta-analysis^{62,64}.

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

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

523

$$524 \pi_{Cj} = \frac{\Lambda_j}{\sum_k \Lambda_k},$$

525

526 where the summation is over all variants mapping within the (conditional) meta-analysis across the
 527 locus. In this expression, Λ_j is the Bayes' factor (BF) in favour of association from the MANTRA
 528 analysis. A 99% credible set⁶³ was then constructed by: (i) ranking all variants according to their BF,
 529 Λ_j ; and (ii) including ranked variants until their cumulative posterior probability exceeds 0.99.

530

531 **Genomic annotation.** We used genomic annotations of DNaseI hypersensitive sites (DHS) from the
 532 ENCODE⁶⁵ project and protein coding genes from GENCODE⁶⁶. We filtered cell types that are cancer
 533 cell lines (karyotype 'cancer' from <https://genome.ucsc.edu/ENCODE/cellTypes.html>), and merged
 534 data from multiple samples from the same cell type. This resulted in 128 DHS cell-type annotations,
 535 as well as 4 gene-based annotations (coding exon, 5UTR, 3UTR and 1kb upstream of TSS). First, we
 536 tested for the effect of each cell type DHS and gene annotation individually using the Bayes' factors
 537 for all variants in the 62 credible sets using fgwas⁶⁷. Second, we categorised the annotations into
 538 'genic', 'foetal DHS', 'embryonic DHS', 'stem cell DHS', 'neonatal DHS' and 'adult DHS' based on the
 539 description fields from ENCODE, and tested for the effect of each category individually as described
 540 above using fgwas. Third, we then tested the effect of each category by including all categories in a
 541 joint model using fgwas. For each of the three analyses, we obtained the estimated effects and 95%
 542 confidence intervals (CI) for each annotation, and considered an annotation enriched if the 95% CI
 543 did not overlap zero.

544

545 **Estimation of genetic variance explained.** Variance explained was calculated using the REML
 546 method implemented in GCTA⁶⁸. We considered the variance explained by two sets of SNPs: (i) lead
 547 SNPs of all 62 distinct association signals at the 59 established and novel autosomal BW loci
 548 identified in the European-specific or trans-ancestry meta-analyses; (ii) lead SNPs of 55 distinct
 549 association signals at the 52 novel autosomal BW loci (**Extended Data Table 1a** and **Supplementary**
 550 **Table 7**). Variance explained was calculated in samples of European ancestry in the Hyperglycemia
 551 and Adverse Pregnancy Outcome (HAPO) study⁶⁹ (independent of the meta-analysis) and two studies
 552 that were part of the European ancestry meta-analysis: NFBC1966 and Generation R
 553 (**Supplementary Table 1**). In each study, the genetic relationship matrix was estimated for each set
 554 of SNPs and was tested individually against BW (males and females combined) with study specific
 555 covariates. These analyses provided an estimate and standard error for the variance explained by
 556 each of the given sets of SNPs.

557

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

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

566 Second, we compared associations with BW of the foetal versus maternal genotype at each
 567 of the 60 BW loci. The maternal allelic effect on offspring BW was obtained from a maternal GWAS
 568 meta-analysis of 68,254 European mothers from the EGG Consortium ($n=19,626$)⁷ and the UK
 569 Biobank ($n=48,628$). In the UK Biobank, mothers were asked to report the BW of their first child.
 570 Women of European ancestry with genotype data available in the May 2015 data release were
 571 included, and those with reported BW equivalent to <2.5 kg or >4.5 kg were excluded. No

572 information on gestational age or gender of child was available. BW of first child was associated with
 573 maternal factors such as smoking status, BMI and height in the expected directions. Of the 68,254
 574 women included in the maternal GWAS, 13% were mothers of individuals included in the current
 575 foetal European ancestry GWAS, and a further approximately 45% were themselves (with their own
 576 BW) included in the foetal GWAS.

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

582 Fourth, we used the method of Zhang et al¹⁵ to test associations between BW and the
 583 maternal untransmitted, maternal transmitted and inferred paternal transmitted haplotype score of
 584 422 height SNPs²⁵, 30 SBP SNPs^{13,14} and 84 T2D SNPs²⁴ in 5,201 mother-child pairs from the ALSPAC
 585 study.

586
 587 **LD Score Regression.** The use of LD Score regression to estimate the genetic correlation between
 588 two traits/diseases has been described in detail elsewhere⁷⁰. Briefly, “LD Score” is a measure of how
 589 much genetic variation each variant tags; if a variant has a high LD Score then it is in high LD with
 590 many nearby polymorphisms. Variants with high LD Scores are more likely to contain more true
 591 signals and hence provide more chance of overlap with genuine signals between GWAS. The LD
 592 score regression method uses summary statistics from the GWAS meta-analysis of BW and the other
 593 traits of interest, calculates the cross-product of test statistics at each SNP, and then regresses the
 594 cross-product on the LD Score. Bulik-Sullivan et al⁷⁰ show that the slope of the regression is a
 595 function of the genetic covariance between traits:

596

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

597 where N_i is the sample size for study i , ρ_g is the genetic covariance, M is the number of SNPs in the
 598 reference panel with MAF between 5% and 50%, l_j is the LD score for SNP j , N_s quantifies the number
 599 of individuals that overlap both studies, and ρ is the phenotypic correlation amongst the N_s
 600 overlapping samples. Thus, if there is sample overlap (or cryptic relatedness between samples), it
 601 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

602 estimates of the genetic covariance will not be biased by sample overlap. Likewise, population
 603 stratification will affect the intercept but will have minimal impact on the slope (i.e. intuitively since
 604 population stratification does not correlate with linkage disequilibrium between nearby markers).

605 Summary statistics from the GWAS meta-analysis for traits and diseases of interest were
 606 downloaded from the relevant consortium website. The summary statistics files were reformatted
 607 for LD Score regression analysis using the `munge_sumstats.py` python script provided on the
 608 developer’s website (<https://github.com/bulik/ldsc>). For each trait, we filtered the summary
 609 statistics to the subset of HapMap 3 SNPs⁷¹, as advised by the developers, to ensure that no bias was
 610 introduced due to poor imputation quality. Summary statistics from the European-specific BW meta-
 611 analysis were used because of the variable LD structure between ancestry groups. Where the sample
 612 size for each SNP was included in the results file this was flagged using `--N-col`; if no sample size was
 613 available then the maximum sample size reported in the reference for the GWAS meta-analysis was
 614 used. SNPs were excluded for the following reasons: MAF<0.01; ambiguous strand; duplicate rsID;
 615 non-autosomal SNPs; reported sample size less than 60% of the total available. Once all files were
 616 reformatted, we used the `ldsc.py` python script, also on the developers’ website, to calculate the
 617 genetic correlation between BW and each of the traits and diseases. The European LD Score files

618 that were calculated from the 1000G reference panel and provided by the developers were used for
 619 the analysis. Where multiple GWAS meta-analyses had been conducted on the same phenotype (i.e.
 620 over a period of years), the genetic correlation with BW was estimated using each set of summary
 621 statistics and presented in **Supplementary Table 12**. The phenotypes with multiple GWAS included
 622 height, BMI, waist-hip ratio (adjusted for BMI), total cholesterol, triglycerides, high density
 623 lipoprotein (HDL) and low density lipoprotein (LDL). The estimate of the genetic correlation between
 624 the multiple GWAS meta-analyses on the same phenotype were comparable and the later GWAS
 625 had a smaller standard error due to the increased sample size, so only the genetic correlation
 626 between BW and the most recent meta-analyses were presented in **Fig. 2**.

627 In the published GWAS for BP¹⁴ the phenotype was adjusted for BMI. Caution is needed
 628 when interpreting the genetic correlation between BW and BMI-adjusted SBP due to the potential
 629 for collider bias⁷². Since BMI is associated with both BP and BW, it is possible that the use of a BP
 630 genetic score adjusted for BMI might bias the genetic correlation estimate towards a more negative
 631 value. To verify that the inverse genetic correlation with BW ($r_g = -0.26$, $SE = 0.05$, $P = 6.5 \times 10^{-9}$) was not
 632 due to collider bias caused by the BMI adjustment of the phenotype, we obtained an alternative
 633 estimate using UK Biobank GWAS data for SBP that was unadjusted for BMI and obtained a similar
 634 result ($r_g = -0.22$, $SE = 0.03$, $P = 5.5 \times 10^{-13}$). The SBP phenotype in UK Biobank was prepared as follows.
 635 Two BP readings were taken at assessment, approximately 5 minutes apart. We included all
 636 individuals with an automated BP reading (taken using an automated Omron BP monitor). Two valid
 637 measurements were available for most participants (averaged to create a BP variable, or
 638 alternatively a single reading was used if only one was available). Individuals were excluded if the
 639 two readings differed by more than 4.56 SD. BP measurements more than 4.56 SD away from the
 640 mean were excluded. We accounted for BP medication use by adding 15 mmHg to the SBP measure.
 641 BP was adjusted for age, sex and centre location and then inverse rank normalised. We performed
 642 the GWAS on 127,698 individuals of British descent using BOLT-LMM³⁷, with genotyping array as
 643 covariate.

644
 645 **Estimating the proportion of the BW-adult traits covariance attributable to genotyped SNPs.** We
 646 estimated the phenotypic, genetic and residual correlations as well as the genetic and residual
 647 covariance between BW and several quantitative traits/disease outcomes in UK Biobank using
 648 directly genotyped SNPs and the REML method implemented in BOLT-LMM³⁷. The traits examined
 649 included T2D, SBP, diastolic BP, CAD, height, BMI, weight, waist-hip ratio, hip circumference, waist
 650 circumference, obesity, overweight, age at menarche, asthma, and smoking. Where phenotypes
 651 were not available (e.g. serum blood measures are not currently available in UK Biobank), we
 652 obtained estimates using the NFBC1966 study (for correlations/covariance between BW and
 653 triglycerides, total cholesterol, HDL, LDL, fasting glucose and fasting insulin). In the UK Biobank
 654 analysis, we used 57,715 unrelated individuals with BW available and identified by UK Biobank as
 655 white British. SNPs with evidence of deviation from Hardy-Weinberg Equilibrium ($P < 1 \times 10^{-6}$),
 656 $MAF < 0.05$ or overall missing rate > 0.015 were excluded, resulting in 328,928 SNPs for analysis. We
 657 included the first five ancestry principal components as covariates. In the NFBC1966 analysis, 5,009
 658 individuals with BW were enrolled. Genotyped SNPs that passed quality control (**Supplementary**
 659 **Table 2**) were included, resulting in 324,895 SNPs for analysis. The first three ancestry principal
 660 components and sex were included as covariates.

661
 662 **Gene set enrichment analysis.** Meta-Analysis Gene-set Enrichment of variaNT Associations
 663 (MAGENTA) was used to explore pathway-based associations using summary statistics from the
 664 trans-ancestry meta-analysis. MAGENTA implements a gene set enrichment analysis (GSEA) based
 665 approach, as previously described⁹. Briefly, each gene in the genome is mapped to a single index SNP
 666 with the lowest P -value within a 110 kb upstream and 40 kb downstream window. This P -value,
 667 representing a gene score, is then corrected for confounding factors such as gene size, SNP density
 668 and LD-related properties in a regression model. Genes within the HLA-region were excluded from

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

679

680 **Protein-Protein interaction network analyses.** We used the integrative Protein-Interaction-
 681 Network-Based Pathway Analysis (iPINBPA) method⁷³. Briefly, we generated gene-wise *P*-values
 682 from the trans-ancestry meta-analysis using VEGAS2⁷⁴, which map the SNPs to genes and account for
 683 possible cofounders, such as LD between markers. The empirical gene-wise *P*-values are calculated
 684 using simulations from the multivariate normal distribution. Those that were nominally significant
 685 ($P \leq 0.01$) were selected as “seed genes”, and were collated within high confidence version of
 686 inweb3⁷⁵, to weight the nodes in the network following a guilt-by-association approach. In a second
 687 step, a network score was defined by the combination of the z-scores derived from the gene-wise *P*-
 688 values with node weights using the Liptak-Stouffer method⁷⁶. A heuristic algorithm was then applied
 689 to extensively search for modules enriched in genes with low *P*-values. The modules were further
 690 normalised using a null distribution of 10,000 random networks. Only those modules with z-score >5
 691 were selected. Finally, the union of all modules constructed a BW-overall PPI network. Both the
 692 proteins on the individual modules and on the overall BW-PPI were interrogated for enrichment in
 693 Gene Ontology Terms (Biological Processes) using a Hypergeometric test. Terms were considered as
 694 significant when adjusted *P*-value, following Benjamini-Hochberg procedure, was below 0.05.

695

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

702

703 **Parent-of-origin specific associations.** We first searched for evidence of parent of origin effects in
 704 the UK Biobank samples by comparing variance between heterozygotes and homozygotes using
 705 Quicktest⁷⁷. In this analysis, we used only unrelated individuals identified genetically as of white
 706 British origin (n=57,715). Principal components were generated using these individuals and the first
 707 five were used to adjust for population structure as covariates in the analysis, in addition to a binary
 708 indicator for genotyping array.

709 We also examined 4,908 mother-child pairs in ALSPAC and determined the parental origin of
 710 the alleles where possible⁷⁸. Briefly, the method uses mother-child pairs to determine the parent of
 711 origin of each allele. For example, if the mother/child genotypes are AA/Aa, the child’s
 712 maternal/paternal allele combination is A/a. For the situation where both mother and child are
 713 heterozygous, the child’s maternal/paternal alleles cannot be directly specified. However, the
 714 parental origin of the alleles can be determined by phasing the genotype data and comparing
 715 maternal and child haplotypes. We then tested these alleles for association with BW adjusting for
 716 sex and gestational age.

717

718 Statistical power in these currently available sample sizes is insufficient to rule out
 719 widespread parent-of-origin effects across the regions tested. Using the mean beta (0.034 SD) and
 MAF (0.28) of the identified loci, we estimate that we would need at least 200,000 unrelated

720 individuals or 70,000 mother-child pairs for 80% power to detect parent-of-origin effects at
721 $P < 0.00085$.

722

723 **Hierarchical clustering of BW loci.** To explore the different patterns of association between BW and
724 other anthropometric/metabolic/endocrine traits and diseases, we performed hierarchical clustering
725 analysis. The lead SNP (or proxy, $EUR r^2 > 0.6$) at the 60 BW loci was queried in publicly available
726 GWAS meta-analysis datasets or in GWAS result obtained through collaboration⁷⁹. Results were
727 available for 53 of those loci and the extracted z-score (allelic effect/SE, **Supplementary Table 17**)
728 was aligned to the BW-raising allele. We performed two dimensional clustering by trait and by locus.
729 We computed the Euclidean distance amongst z-scores of the extracted traits/loci and performed
730 complete hierarchical clustering implemented in the pvclust package ([http://www.sigmath.es.osaka-](http://www.sigmath.es.osaka-u.ac.jp/shimo-lab/prog/pvclust/)
731 [u.ac.jp/shimo-lab/prog/pvclust/](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
732 measured by multiscale bootstrap resampling estimated from 1,000 replicates. We used $\alpha = 0.05$ to
733 define distinct clusters and, based on the bootstrap analysis, calculated the Calinski index to identify
734 the number of well-supported clusters (cascadeKM function, Vegan package, [http://CRAN.R-](http://CRAN.R-project.org/package=vegan)
735 [project.org/package=vegan](http://CRAN.R-project.org/package=vegan)). Clustering was visualised by constructing dendrograms and a heatmap.

736 Separately from the hierarchical clustering analysis, we queried the lead SNP at *EPAS1* in a
737 GWAS of haematological traits⁸⁰ because variation at that locus has previously been implicated in
738 BW and adaptation to hypoxia at high altitudes in Tibetans^{81,82} (**Supplementary Table 17**).

739

740

741 **ADDITIONAL REFERENCES FOR METHODS**

742

- 743 30. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*
744 **11**, 499-511 (2010).
- 745 31. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype
746 imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**, 955-959
747 (2012).
- 748 32. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, *et al.* Quality control and conduct
749 of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192-1212 (2014).
- 750 33. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, *et al.* Principal components
751 analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909
752 (2006).
- 753 34. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
- 754 35. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, *et al.* Variance component model to account
755 for sample structure in genome-wide association studies. *Nat Genet* **42**, 348-354 (2010).
- 756 36. Allen NE, Sudlow C, Peakman T, Collins R. UK Biobank. UK biobank data: come and get it. *Sci*
757 *Transl Med* **6**, 224ed4 (2014).
- 758 37. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsón BJ, Finucane HK, *et al.* Efficient Bayesian
759 mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290
760 (2015).
- 761 38. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR *et al.* PLINK: a toolset for whole-
762 genome association and population-based linkage analysis. *Am J Hum Genet* **81**, 559-575
763 (2007).
- 764 39. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC*
765 *Bioinformatics* **11**, 288 (2010).
- 766 40. Willer CJ, Li Y, and Abecasis GR. METAL: fast and efficient meta-analysis of genomewide
767 association scans. *Bioinformatics* **26**, 2190-2191 (2010).
- 768 41. Ioannidis J, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide
769 association investigations. *PLoS One* **2**, e0000841 (2007).

- 770 42. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits
771 (GIANT) Consortium, *et al.* Conditional and joint multiple-SNP analysis of GWAS summary
772 statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375 (2012).
- 773 43. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
774 multitissue gene. *Science* **348**, 648-660 (2015).
- 775 44. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, *et al.* Transcriptome and
776 genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-511 (2013).
- 777 45. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, *et al.* Transcriptome
778 genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777
779 (2010).
- 780 46. Schadt EE, Molony C, Chudin E, Hao K, Yang X, *et al.* Mapping the genetic architecture of gene
781 expression in human liver. *PLoS Biol* **6**, e107 (2008).
- 782 47. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, *et al.* Abundant quantitative
783 trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* **6**,
784 e1000952 (2010).
- 785 48. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, *et al.* Population genomics of human gene
786 expression. *Nat Genet* **9**, 1217-1224 (2007).
- 787 49. Li Q, Stram A, Chen C, Kar S, Gayther S, *et al.* Expression QTL-based analyses reveal candidate
788 causal genes and loci across five tumor types. *Hum Mol Genet* **23**, 5294-5302 (2014).
- 789 50. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, *et al.* Systematic identification of
790 trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).
- 791 51. Zou F, Chai HS, Younkin CS, Allen M, Crook J, *et al.* Brain expression genome-wide association
792 study (eGWAS) identifies human disease-associated variants. *PLoS Genet* **8**, e1002707 (2012).
- 793 52. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, *et al.* Lung eQTLs to help reveal the molecular
794 underpinnings of asthma. *PLoS Genet* **8**, e1003029 (2012).
- 795 53. Koopmann TT, Adriaens ME, Moerland PD, Marsman RF, Westerveld ML, *et al.* Genome-wide
796 identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS One* **9**, e97380
797 (2014).
- 798 54. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, *et al.* Innate immune activity
799 conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**,
800 1246949 (2014).
- 801 55. Grundberg E, Adoue V, Kwan T, Ge B, Duan QL, *et al.* Global analysis of the impact of
802 environmental perturbation on cis-regulation of gene expression. *PLoS Genet* **7**, e1001279
803 (2011).
- 804 56. Ward LD and Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and
805 regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**,
806 D930-934 (2012).
- 807 57. Flicek P, Amode MR, Barrell D, Beal K, Billis K, *et al.* Ensembl 2014. *Nucleic Acids Res* **42**, D749-
808 755 (2014).
- 809 58. Kumar P, Henikoff S, Ng P. Predicting the effects of coding non-synonymous variants on protein
810 function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081 (2009).
- 811 59. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, *et al.* A method and server for
812 predicting damaging missense mutations. *Nat Methods* **7**, 248-249 (2010).
- 813 60. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, *et al.* STRING v10: protein-protein
814 interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-452 (2015).
- 815 61. The International HapMap 3 Consortium. Integrating common and rare genetic variation in
816 diverse human populations. *Nature* **467**, 52-58 (2010).
- 817 62. Morris AP. Transethnic meta-analysis of genome-wide association studies. *Genet Epidemiol* **35**,
818 809-822 (2011).

- 819 63. Wellcome Trust Case Control Consortium, Maller JB, McVean G, Byrnes J, Vukcevic D, *et al.*
820 Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* **44**,
821 1294-1301 (2012).
- 822 64. Wang X, Chua HX, Chen P, Ong RT, Sim X, *et al.* Comparing methods for performing trans-ethnic
823 meta-analysis of genome-wide association studies. *Hum Mol Genet* **22**, 2303-2311 (2013).
- 824 65. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human
825 genome. *Nature* **489**, 57-74 (2012).
- 826 66. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, *et al.* GENCODE: the reference
827 human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774 (2012).
- 828 67. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18
829 human traits. *Am J Hum Genet* **94**, 559–573 (2014).
- 830 68. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, *et al.* Common SNPs explain a large
831 proportion of the heritability for human height. *Nat Genet* **42**, 565-569 (2010).
- 832 69. Urbanek M, Hayes MG, Armstrong LL, Morrison J, Lowe LP *et al.* The chromosome 3q25
833 genomic region is associated with measures of adiposity in newborns in a multi-ethnic genome-
834 wide association study. *Hum Mol Genet* **22**, 3583-3596 (2013).
- 835 70. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, *et al.* An atlas of genetic correlations
836 across human diseases and traits. *Nat Genet* **47**, 1236-1241 (2015).
- 837 71. The International HapMap Consortium. A second generation haplotype map of over 3.1 million
838 SNPs. *Nature* **449**, 851-861 (2007).
- 839 72. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can
840 bias effect estimates in genome-wide association studies. *Am J Hum Genet* **96**, 329-339 (2015).
- 841 73. Wang L, Mousavi P, Baranzini SE. iPINBPA: an integrative network-based functional module
842 discovery tool for genome-wide association studies. *Pac Symp Biocomput* 255-266 (2015).
- 843 74. Mishra A and Macgregor S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res*
844 *Hum Genet* **18**, 86-91 (2015).
- 845 75. Lage K, Karlberg EO, Størling ZM, Olason PI, Pedersen AG, *et al.* A human phenome-interactome
846 network of protein complexes implicated in genetic disorders. *Nat Biotechnol* **25**, 309–316
847 (2007).
- 848 76. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior
849 to Fisher's approach. *J Evol Biol* **18**, 1368-1373 (2005).
- 850 77. Hoggart, CJ, Venturini G, Mangino M, Gomez F, Ascari G, *et al.* Novel Approach Identifies SNPs
851 in SLC2A10 and KCNK9 with Evidence for Parent-of-Origin Effect on Body Mass Index. *PLoS*
852 *Genet* **10**, 1–12 (2014).
- 853 78. Wang S, Yu Z, Miller RL, Tang D & Perera FP. Methods for detecting interactions between
854 imprinted genes and environmental exposures using birth cohort designs with mother-offspring
855 pairs. *Hum Hered* **71**, 196–208 (2011).
- 856 79. Painter JN, Anderson CA, Nyholt DR, Macgregor S, Lin J, *et al.* Genome-wide association study
857 identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet* **43**, 51-54 (2011).
- 858 80. Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, *et al.* Multiple loci influence erythrocyte
859 phenotypes in the CHARGE Consortium. *Nat Genet* **41**, 1191-1198 (2009).
- 860 81. Xu XH, Huang XW, Qun L, Li YN, Wang Y, *et al.* Two functional loci in the promoter of EPAS1
861 gene involved in high-altitude adaptation of Tibetans. *Sci Rep* **4**, 7465 (2014).
- 862 82. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, *et al.* Altitude adaptation in Tibetans caused
863 by introgression of Denisovan-like DNA. *Nature* **512**, 194-197 (2014).
- 864
- 865
- 866

ENDNOTES

Supplementary Information is linked to the online version of the paper.

Acknowledgements Full acknowledgements and supporting grant details 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., 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

Summary statistics from the meta-analyses are available at <http://egg-consortium.org/>. Reprints and permissions information is available at www.nature.com/reprints. One of the authors discloses competing financial interests: Krina Zondervan has a scientific collaboration with Bayer HealthCare Ltd. and Population Diagnostics Inc. Correspondence and requests for materials should be addressed to mark.mccarthy@drl.ox.ac.uk and r.freathy@ex.ac.uk.

EXTENDED DATA LEGENDS

Extended Data Figure 1 | Flow chart of the study design.

Extended Data Figure 2 | 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 BW from trans-ancestry meta-analysis of up to 153,781 individuals. The association P -value (on $-\log_{10}$ scale) for each of up to 22,434,434 SNPs (y axis) is plotted against the genomic position (NCBI Build 37; x axis). Association signals that reached genome-wide significance ($P < 5 \times 10^{-8}$) 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 signals⁵. **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 ($P < 5 \times 10^{-8}$) 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, depending on chromosome number (odd or even). 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.

Extended Data Figure 3 | Regional plots for multiple distinct signals at three birth weight loci, *ZBTB7B* (a), *HMGAI* (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 Figure 4 | Comparison of foetal effect sizes and maternal effect sizes at 60 known and novel birth weight loci (continues to Extended Data Figure 5). For each BW locus, the following six effect sizes (with 95% CI) are shown, all aligned to the same BW-raising allele: **foetal_GWAS** = foetal allelic effect on BW (from European ancestry meta-analysis of up to $n=143,677$ individuals); **foetal_unadjusted** = foetal allelic effect on BW (unconditioned in $n=12,909$ mother-child pairs); **foetal_adjusted** = foetal effect (conditioned on maternal genotype, $n=12,909$); **maternal_GWAS** = maternal allelic effect on offspring BW (from meta-analysis of up to $n=68,254$ European mothers)⁷; **maternal_unadjusted** = maternal allelic effect on offspring BW (unconditioned, $n=12,909$); **maternal_adjusted** = maternal effect (conditioned on foetal genotype, $n=12,909$). The 60 BW loci are ordered by chromosome and position (Supplementary Tables 10, 11). These plots illustrate that in large GWAS of BW, foetal effect size estimates are larger than those of maternal at 55/60 identified loci (binomial $P=1 \times 10^{-11}$), suggesting that most of the associations are 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.

Extended Data Figure 5 | Comparison of foetal effect sizes and maternal effect sizes at 60 known and novel birth weight loci. **a**, Continued from Extended Data Figure 4. **b**, 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 Figure 6 | Protein-Protein Interaction (PPI) Network analysis. **a**, Illustrates the largest global component of 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 BW modules (M1-13) lie. For each module, the two most significant GO terms are depicted. **c**, Illustrates a heatmap which takes the top 50 biological processes over-represented in the global BW PPI network (listed at the right of the plot), and displays extent of enrichment for the various 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 BW 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; orange nodes are genes in BW loci that overlap with both the pathway and PoC. Large nodes correspond to genes in BW loci (within 300kb from the lead SNP), 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 Figure 7 | 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*. **a**, QQ plot from the Quicktest⁷⁷ 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 BW and maternally transmitted vs. paternally transmitted alleles in 4,908 mother-child pairs from the ALSPAC study (Methods, Supplementary Tables 15, 16). In both panels, the black dots represent lead SNPs at 59 identified autosomal BW loci and a further sub-genome-wide significant signal for BW near *DLK1* (rs6575803; $P=5.6 \times 10^{-8}$). The grey lines represent expected P values and their 95% confidence intervals under the null distribution for the 60 SNPs. Both results show trends in favour of imprinting effects at BW loci: however, despite the large sample size, these analyses were underpowered (see Methods) and much larger sample sizes are required for definitive analysis.

Extended Data Figure 8 | 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^{13,14}, 45 CAD loci²³, 84 T2D loci²⁴ and 422 adult height loci²⁵ are plotted against effects on BW (x axis). Effect sizes are aligned to the adult trait-raising allele. The colour of each dot indicates BW association P value: red, $P < 5 \times 10^{-8}$; orange, $5 \times 10^{-8} \leq P < 0.001$; yellow, $0.001 \leq P < 0.01$; white, $P \geq 0.01$. The superimposed grey frequency histogram shows the number of SNPs (right y axis) in each category of BW 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. CAD loci with a larger effect on BW concentrated amongst loci with primary BP association. **f**, Effect sizes (with 95% CI) on BW of 32 known T2D loci are plotted, subdivided by previously reported categories derived from detailed adult physiological data²⁷. Heterogeneity in BW effect sizes between five T2D loci groups with different mechanistic categories was substantial ($P_{\text{het}}=1.2 \times 10^{-9}$). In pairwise comparisons, the “beta cell” group of variants differed from the other four groups: fasting hyperglycaemia ($P_{\text{het}}=3 \times 10^{-11}$),

insulin resistance ($P_{\text{het}}=0.002$), proinsulin ($P_{\text{het}}=0.78$) and unclassified ($P_{\text{het}}=0.02$) groups. All of the BW effect sizes plotted in the forest plots are aligned to the trait (or risk)-raising allele.

Extended Data Table 1 | Sixty loci associated with birth weight ($P<5\times 10^{-8}$) in European ancestry meta-analysis of up to 143,677 individuals and/or trans-ancestry meta-analysis of up to 153,781 individuals. a, Effects (beta values) are aligned to the BW-raising allele. EAF was obtained from the trans-ancestry meta-analysis, except for *PLAC1*, for which the EAF was obtained from the European ancestry meta-analysis due to lack of X chromosome data from the non-European studies. Chr., chromosome; bp, base pair; EAF, effect allele frequency; SE, standard error. **b,** The effect of the lead SNP (absolute value of beta, y axis) is given as a function of minor allele frequency (x axis) for 60 known (pink) and novel (green) BW loci from the trans-ancestry meta-analysis. 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.

Extended Data Table 2 | 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. **a,** The top results (FDR<0.05 at the 95th percentile enrichment threshold) from a total of 3,216 biological pathways tested for enrichment of multiple modest associations with BW. Additionally, results are presented for custom sets of imprinted genes. **b,** The results of a complementary analysis of empirical PPI data, displaying the top 10 most significant pathways enriched for BW-association scores.