LETTER

doi:10.1038/nature19806

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 Feenstra[10], Natalie R. van Zuydam[1,2], Kyle J. Gaulton[1,11], Niels Grarup[12], Jonathan P. Bradfield[13], David P. Strachan[14], Ruifang Li-Gao[15], Tarunveer S. Ahluwalia[12,16,17], Eskil Kreiner[16], Rico Rueedi[18,19], Leo-Pekka Lyytikäinen[20,21], Diana L. Cousminer[22,23,24], Ying Wu[25], Elisabeth Thiering[26,27], Carol A. Wang[6], Christian T. Have[12], Jouke-Jan Hottenga[28], Natalia Vilor-Tejedor[29,30,31], Peter K. Joshi[32], Eileen Tai Hui Boh[33], Ioanna Ntalla[34,35], Niina Pitkänen[36], Anubha Mahajan[1], Elisabeth M. van Leeuwen[8], Raimo Joro[37], Vasiliki Lagou[1,38,39], Michael Nodzenski[40], Louise A. Diver[41], Krina T. Zondervan[1,42], Mariona Bustamante[29,30,31,43], Pedro Marques-Vidal[44], Josep M. Mercader[45], Amanda J. Bennett[2], Nilufer Rahmioglu[1], Dale R. Nyholt[46], Ronald C. W. Ma[47,48,49], Claudia H. T. Tam[47], Wing Hung Tam[50], CHARGE Consortium Hematology Working Group†, Santhi K. Ganesh[51], Frank J. A. van Rooij[8], Samuel E. Jones[3], Po-Ru Loh[52,53], Katherine S. Ruth[3], Marcus A. Tuke[3], Jessica Tyrrell[3,54], Andrew R. Wood[3], Hanieh Yaghootkar[3], Denise M. Scholtens[40], Lavinia Paternoster[55,56], Inga Prokopenko[1,57], Peter Kovacs[58], Mustafa Atalay[37], Sara M. Willems[8], Kalliope Panoutsopoulou[59], Xu Wang[33], Lisbeth Carstensen[10], Frank Geller[10], Katharina E. Schraut[32], Mario Murcia[31,60], Catharina E. M. van Beijsterveldt[28], Gonneke Willemsen[28], Emil V. R. Appel[12], Cilius E. Fonvig[12,61], Caecilie Trier[12,61], Carla M. T. Tiesler[26,27], Marie Standl[26], Zoltán Kutalik[19,62], Sílvia Bonàs-Guarch[45], David M. Hougaard[63,64], Friman Sánchez[45,65], David Torrents[45,66], Johannes Waage[16], Mads V. Hollegaard[63,64]‡, Hugoline G. de Haan[15], Frits R. Rosendaal[15], Carolina Medina-Gomez[7,8,67], Susan M. Ring[55,56], Gibran Hemani[55,56], George McMahon[56], Neil R. Robertson[1,2], Christopher J. Groves[2], Claudia Langenberg[4], Jian'an Luan[4], Robert A. Scott[4], Jing Hua Zhao[4], Frank D. Mentch[13], Scott M. MacKenzie[41], Rebecca M. Reynolds[68], Early Growth Genetics (EGG) Consortium†, William L. Lowe Jr[69], Anke Tönjes[70], Michael Stumvoll[58,70], Virpi Lindi[37], Timo A. Lakka[37,71,72], Cornelia M. van Duijn[8], Wieland Kiess[73], Antje Körner[58,73], Thorkild I. A. Sørensen[55,56,74,75], Harri Niinikoski[76,77], Katja Pahkala[36,78], Olli T. Raitakari[36,79], Eleftheria Zeggini[59], George V. Dedoussis[35], Yik-Ying Teo[33,80,81], Seang-Mei Saw[33,82], Mads Melbye[10,83,84], Harry Campbell[32], James F. Wilson[32,85], Martine Vrijheid[29,30,31], Eco J. C. N. de Geus[28,86], Dorret I. Boomsma[28], Haja N. Kadarmideen[87], Jens-Christian Holm[12,61], Torben Hansen[12], Sylvain Sebert[57,88,89], Andrew T. Hattersley[3], Lawrence J. Beilin[90], John P. Newnham[6], Craig E. Pennell[6], Joachim Heinrich[26,91], Linda S. Adair[92], Judith B. Borja[93,94], Karen L. Mohlke[25], Johan G. Eriksson[95,96,97], Elisabeth Widén[22], Mika Kähönen[98,99], Jorma S. Viikari[100,101], Terho Lehtimäki[20,21], Peter Vollenweider[44], Klaus Bønnelykke[16], Hans Bisgaard[16], Dennis O. Mook-Kanamori[15,102,103], Albert Hofman[7,8], Fernando Rivadeneira[7,8,67], André G. Uitterlinden[7,8,67], Charlotta Pisinger[104], Oluf Pedersen[12], Christine Power[105], Elina Hyppönen[105,106,107], Nicholas J. Wareham[4], Hakon Hakonarson[13,23,108], Eleanor Davies[41], Brian R. Walker[68], Vincent W. V. Jaddoe[7,8,9], Marjo-Riitta Järvelin[88,89,109,110], Struan F. A. Grant[13,23,108,111], Allan A. Vaag[83,112,113], Debbie A. Lawlor[55,56], Timothy M. Frayling[3], George Davey Smith[55,56], Andrew P. Morris[1,114,115]§, Ken K. Ong[4,116]§, Janine F. Felix[7,8,9]§, Nicholas J. Timpson[55,56]§, John R. B. Perry[4]§, David M. Evans[5,55,56]§, Mark I. McCarthy[1,2,117]§ & Rachel M. Freathy[3,55]§

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

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

We identified 60 loci (of which 59 were autosomal) associated with BW at genome-wide significance ($P < 5 \times 10^{-8}$) in either the European

A list of affiliations appears in the online version of this paper.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ancestry or trans-ancestry meta-analyses (Extended Data Fig. 2a, Extended Data Table 1a and Supplementary Data; Methods). For lead single nucleotide polymorphisms (SNPs), we observed no heterogeneity in allelic effects between the three study components (Cochran's Q statistic $P > 0.00083$) (Supplementary Table 3). We found that 53 of these loci were novel in that the lead SNP mapped >2 Mb away from, and was independent ($R^2 < 0.05$ in the European (EUR) component of 1000G) of, the seven previously reported BW signals[5], all of which were confirmed in this larger analysis (Supplementary Table 4). Approximate conditional analysis in the European ancestry data indicated that three of these novel loci (near *ZBTB7B, HMGA1* and *PTCH1*) harboured multiple distinct association signals that attained genome-wide significance ($P < 5 \times 10^{-8}$) (Methods, Supplementary Table 5 and Extended Data Fig. 3).

The lead variants for most signals mapped to non-coding sequences, and at only two loci, *ADRB1* (rs7076938; $R^2 = 0.99$ with *ADRB1* G389R) and *NRIP1* (rs2229742, R448G), did the association data point to potential causal non-synonymous coding variants (Supplementary Table 6 and Methods). Lead SNPs for all but two loci (those mapping near *YKT6–GCK* and *SUZ12P1–CRLF3*) were common (minor allele frequency (MAF) $\geq$ 5%) with individually modest effects on BW ($\beta = 0.020$–0.053 standard deviations (s.d.) per allele, equivalent to 10–26 g). This was despite the much-improved coverage of low-frequency variants in this study (compared to previous HapMap 2 imputed meta-analyses, ref. 5) reflecting imputation from larger, and more complete, reference panels (Extended Data Table 1b). Indeed, all but five of the common variant association signals were tagged by variants (EUR $R^2 > 0.6$) in the HapMap 2 reference panel (Supplementary Tables 4, 5), indicating that most of the novel discoveries in the present study were driven by increased sample size[5]. Fine-mapping analysis yielded 14 regions in which fewer than ten variants contributed to the locus-specific credible sets that accounted for >99% of the posterior probability of association (Methods and Supplementary Table 7). The greatest refinement was at *YKT6–GCK*, where the credible set included only the low frequency variant rs138715366, which maps intronic to *YKT6*. These credible-set variants collectively showed enrichment for overlap with DNaseI hypersensitivity sites, particularly those generated, by ENCODE, from fetal (4.2-fold, 95% CI 1.8–10.7) and neonatal tissues (4.9-fold, 1.8–11.0) (Supplementary Fig. 1, Supplementary Table 8 and Methods).

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

Associations between fetal genotype and BW could result from indirect effects of the maternal genotype influencing BW via the intrauterine environment, given the correlation ($R \approx 0.5$) between maternal and fetal genotype. However, two lines of evidence indicated that variation in the fetal genome was the predominant driver of BW associations. First, an analysis of the global contribution of maternal versus fetal genetic variation, using a maternal genome-wide complex trait analysis (GCTA) model (ref. 6) (Methods) applied to 4,382 mother–child pairs, estimated that the child's genotype ($\sigma_C^2 = 0.24 \pm 0.11$) made a larger contribution to BW variance than either the mother's genotype ($\sigma_M^2 = 0.04 \pm 0.10$), or the covariance between the two ($\sigma_{CM} = 0.04 \pm 0.08$). Second, when we compared the point estimates of the BW-effect size dependent on maternal genotype at each of the 60 loci (as measured in up to 68,254 women[7]) with those dependent on fetal genotype (using European ancestry data from 143,677 individuals in the present study), fetal variation had a greater impact than maternal variation at 93% of the loci (55 out of 60; binomial $P = 10^{-11}$) (Supplementary Table 10, Extended Data Figs 4, 5 and Methods). The power to further disentangle maternal and fetal contributions using analyses of fetal genotype which were conditional on maternal genotype was constrained by the limited sample size available ($n = 12{,}909$ mother–child pairs) (Supplementary Table 11).

Collectively, these analyses provide evidence that the fetal genotype has a substantial impact on early growth, as measured by BW. We used these genetic associations to understand the causal relationships underlying observed associations between BW and disease, and to characterize the processes responsible.

To quantify the shared genetic contribution to BW and other health-related traits, we estimated their genetic correlations using linkage-disequilibrium score regression[8] (Methods). BW (in European ancestry samples) showed strong positive genetic correlations with anthropometric and obesity-related traits including birth length ($R_g = 0.81$, $P = 2.0 \times 10^{-44}$) and, in adults, height ($R_g = 0.41$, $P = 4.8 \times 10^{-52}$), waist circumference ($R_g = 0.18$, $P = 3.9 \times 10^{-10}$) and BMI ($R_g = 0.11$, $P = 7.3 \times 10^{-6}$). By contrast, BW showed inverse genetic correlations with indicators of adverse metabolic and cardiovascular health including coronary artery disease (CAD, $R_g = -0.30$, $P = 6.5 \times 10^{-9}$), systolic blood pressure (SBP, $R_g = -0.22$, $P = 5.5 \times 10^{-13}$) and T2D ($R_g = -0.27$, $P = 1.1 \times 10^{-6}$) (Fig. 1, Supplementary Table 12). The correlations between BW and adult cardiometabolic phenotypes are of similar magnitude, although directionally opposite, to the reported genetic correlations between adult BMI and those same cardiometabolic outcomes[8]. These findings support observational associations between a history of paternal T2D and lower BW (ref. 4), and establish more generally that the observed life-course associations between early growth and adult disease, at least in part, reflect the impact of shared genetic variants that influence both sets of phenotypes.

In an effort to estimate the extent of genetic contribution to these life-course associations, we first focused on data from the UK Biobank ($n = 57{,}715$). For many of the traits for which data were available, genetic variation contributed substantially to the life-course relationship between BW and adult phenotypes, and in some cases appeared to be the major source of covariance between the traits. For example, we estimated that 85% (95% CI = 70–99%) of the negative covariance between BW and SBP was explained by shared genetic associations captured by directly genotyped SNPs (Supplementary Table 13, Methods and Supplementary Fig. 2). For continuous cardiometabolic measures, including lipids and fasting glycaemia, for which measures are not currently available in the UK Biobank, we used data from the Northern Finland Birth Cohort ($n = 5{,}009$), and obtained similar results (Supplementary Table 13). However, these estimates were limited, not only by wide confidence intervals, but also by the assumption of a linear relationship between BW and each of the phenotypes and by the inability to explicitly model maternal genotypic effects. In other words, the inverse genetic correlations between BW and cardiometabolic traits may not exclusively reflect genetic effects mediated directly through the offspring, but also effects mediated by maternal genotype acting indirectly on the fetus via perturbation of the *in utero* environment. Nevertheless, these estimates indicate that a substantial proportion of the variance in cardiometabolic risk that correlates with BW can be attributed to the effects of common genetic variation.

To elucidate the biological pathways and processes underlying regulation of fetal growth, we first performed gene set enrichment analysis of our BW GWAS analysis using MAGENTA (Meta-Analysis Gene-set Enrichment of variaNT Associations, ref. 9) approach (Methods). Twelve pathways reached study-wide significance (false discovery rate, FDR < 0.05), including pathways involved in metabolism (insulin signalling, glycogen biosynthesis and cholesterol biosynthesis), growth (IGF signalling and growth hormone pathway) and development (chromatin remodelling) (Extended Data Table 2a). Similar pathways were detected in a complementary analysis in which we analysed empirical protein–protein interaction (PPI) data identifying

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

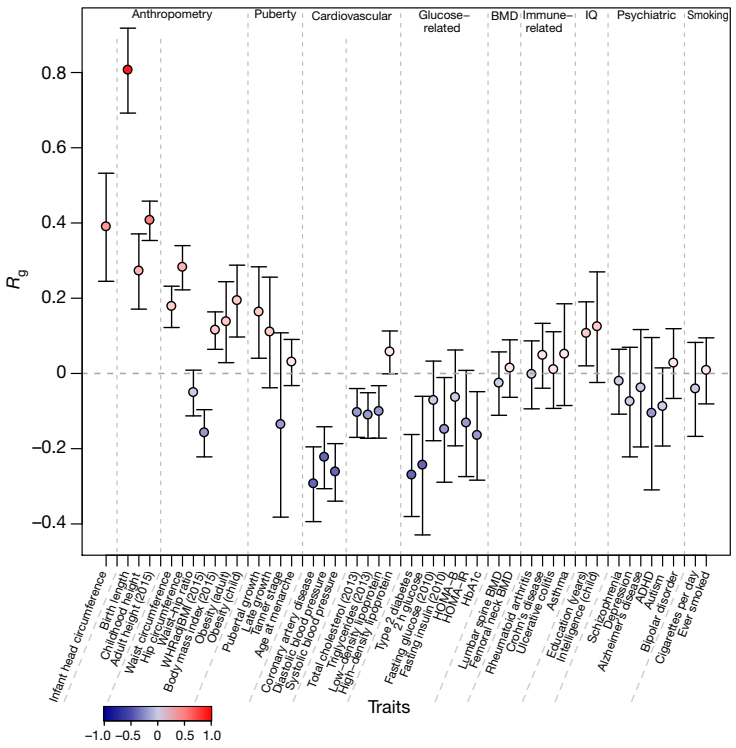


Figure 1 | Genome-wide genetic correlation between BW and a range of traits and diseases in later life. Genetic correlation (R_g) and corresponding s.e. (error bars) between BW and the traits displayed on the *x* axis were estimated using linkage-disequilibrium score regression (ref. 8). The genetic correlation estimates (R_g) are colour coded according to their intensity and direction (red for positive and blue for inverse correlation). WHRadjBMI, waist–hip ratio adjusted for body mass index; HOMA-B/IR, homeostasis model assessment of beta-cell function/insulin resistance; HbA1c, haemoglobin A1c; BMD, bone mineral density; ADHD, attention deficit hyperactivity disorder. See Supplementary Table 12 for references for each of the traits and diseases displayed.

13 PPI network modules with marked (Z score >5) enrichment for BW-association scores (Extended Data Table 2b, Extended Data Fig. 6a, b and Methods). The proteins within these modules were themselves enriched for diverse processes related to metabolism, growth and development (Extended Data Fig. 6a, b).

We also observed enrichment of BW association signals across the set of 77 imprinted genes defined by the Genotype-Tissue Expression (GTEx) project (ref. 10) ($P = 1.9 \times 10^{-4}$; Extended Data Table 2a and Supplementary Table 14). Such enrichment is consistent with the 'parental conflict' hypothesis regarding the allocation of maternal resources to the fetus[11]. Although the role of imprinted genes in fetal growth has been described in animal models and rare human disorders[12], these data provide a large-scale, systematic indication of their contribution to normal variation in BW. Of the 60 genome-wide significant loci, two (*INS–IGF2* and *RB1*) fall within (or near) imprinted regions (Extended Data Fig. 2b), with a noteworthy third signal at *DLK1* (previously fetal antigen-1; $P = 5.6 \times 10^{-8}$). Parent-of-origin specific analyses to further investigate these individual loci (comparing heterozygote versus homozygote BW variance in 57,715 unrelated individuals, and testing BW associations with paternal versus maternal alleles in 4,908 mother–child pairs; see Methods) proved, despite these sample sizes, to be underpowered (Extended Data Fig. 7 and Supplementary Tables 15, 16).

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

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

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

We focused our more detailed investigation of the mechanistic links between early growth and adult traits on two phenotypic areas: arterial blood pressure and T2D/glycaemia. Across both the overall GWAS and specifically among the 60 significant BW loci, most BW-raising alleles were associated with reduced blood pressure (Figs 1, 2); the strongest inverse associations were seen for the loci near *NT5C2, FES, NRIP1, EBF1* and *PTH1R*. However, we also observed locus-specific heterogeneity in the genetic relationships between blood pressure and BW: the SBP-raising allele at *ADRB1*[13] is associated with higher, rather than lower, BW (Extended Data Fig. 8a). When we considered the reciprocal relationship, that is, the effects on BW of blood-pressure-raising alleles at 30 reported loci for SBP[13,14], there was an excess of associations (5 out of 30 with lower BW at $P < 0.05$; binomial $P = 0.0026$; Extended Data Fig. 8a). To dissect maternal and fetal genotype effects at these loci, we tested the impact on BW of a risk score generated from the 30 SBP SNPs, restricted to the untransmitted maternal haplotype score[15] in a set of 5,201 mother–child pairs. Analysis of these loci indicated that maternal genotype effects on the intrauterine environment probably

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

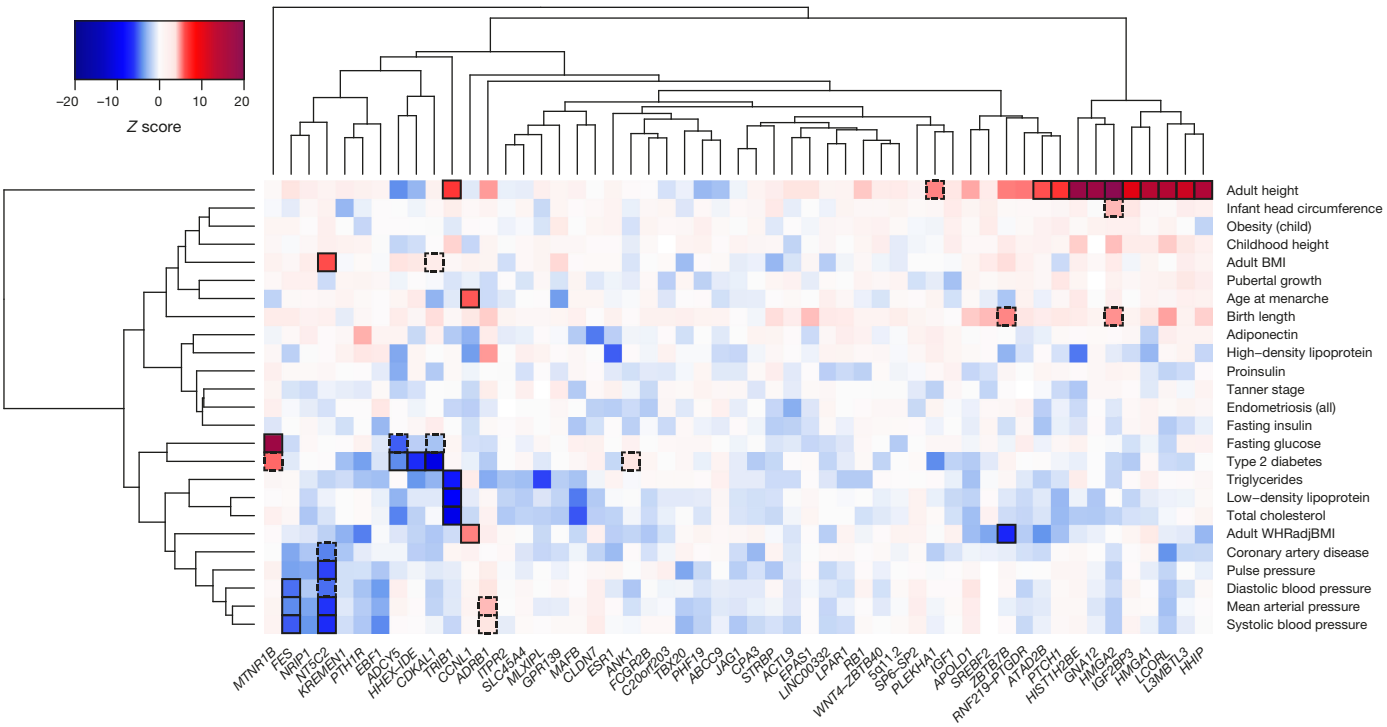


Figure 2 | Hierarchical clustering of BW loci based on similarity of overlap with adult diseases, metabolic and anthropometric traits. For the lead SNP at each BW locus (*x* axis), *Z* scores (aligned to BW-raising allele) were obtained from publicly available GWAS for various traits (*y* axis; see Supplementary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

contribute to the inverse genetic correlation between SBP and BW (Methods and Supplementary Table 18), and was consistent with the results of a wider study of >30,000 women which demonstrated associations between a maternal genetic score for SBP (conditional on fetal genotype) and lower offspring BW[16].

The blood-pressure-raising allele with the largest BW-lowering effect mapped to the *NT5C2* locus (index variant for BW, rs74233809, $R^2 = 0.98$ with index variant for blood pressure, rs11191548; ref. 14) and was also associated with lower adult BMI ($R^2 = 0.99$ with rs11191560; ref. 17). The BW-lowering allele at rs74233809 is a proxy for a recently described[18] functional variant in the nearby *CYP17A1* gene ($R^2 = 0.92$ with rs138009835). The *CYP17A1* gene encodes the cytochrome P450c17α enzyme CYP17 (ref. 19), which catalyses key steps in steroidogenesis that determine the balance between mineralocorticoid, glucocorticoid and androgen synthesis. This variant has been shown to alter transcriptional efficiency *in vitro* and is associated with increased urinary tetrahydroaldosterone excretion[18]. *CYP17A1* is expressed in fetal adrenal glands and testes from early gestation[20] as well as in the placenta[21]. These data suggest that variation in *CYP17A1* expression contributes to the observational association between low BW and adult hypertension[22].

When we analysed 45 loci associated with CAD[23], the inverse genetic correlation between CAD and BW was concentrated amongst the five CAD loci with primary blood pressure associations. This suggests that genetic determinants of blood pressure play a leading role in mediating the life-course associations between BW and CAD (Extended Data Fig. 8b, e).

Linkage-disequilibrium score regression analyses demonstrated overall inverse genetic correlation between lower BW and elevated risk of T2D (Fig. 1). However, the locus-specific heat map indicates a heterogeneous pattern across individual loci (Fig. 2). To explore this further, we tested the 84 reported T2D loci[24] for association with BW. Some T2D risk alleles (such as those at *ADCY5*, *CDKAL1* and *HHEX–IDE*) were strongly associated with lower BW, while others (including *ANK1* and *MTNR1B*) were associated with higher BW (Extended Data Fig. 8c). This was in contrast with the BW effects of 422 known height loci[25] (Extended Data Fig. 8d), which showed a strong positive correlation consistent with the overall genetic correlation between height and BW, indicating that the growth effects of many height loci start prenatally and persist into adulthood.

The contrasting associations of T2D-risk alleles with both higher and lower BW probably reflect the differential impacts, across loci, of variation in the maternal and fetal genomes. Observational data link paternal diabetes with lower offspring BW[4], indicating that the inheritance of T2D risk alleles by the fetus tends, in line with the linkage-disequilibrium score regression analysis, to reduce growth. These relationships are consistent with the precepts of the 'fetal insulin hypothesis'[26] and reflect the potential for reduced insulin secretion and/or signalling to lead to both reduced fetal growth and, many decades later, enhanced predisposition to T2D. In line with this, the inferred paternal transmitted haplotype score generated from the 84 T2D risk variants was associated with lower BW ($P = 0.045$) in 5,201 mother–child pairs (Methods and Supplementary Table 18). In contrast, maternal diabetes is observationally associated with higher offspring BW[4], reflecting the ability of maternal hyperglycaemia to stimulate fetal insulin secretion. The contribution of genotype-dependent maternal hyperglycaemia to BW is in line with the evidence, from a recent study, that maternal genotype scores for fasting glucose and T2D (conditional on fetal genotype) were causally associated with higher offspring BW[16]. It is also consistent with the observation that a subset of glucose-raising alleles is associated with higher BW[7]. For example, the T2D-risk variant at *MTNR1B* (which also has a marked effect on fasting glucose levels in non-diabetic individuals[27,28]) was amongst the subset of BW loci (5 out of 60) for which the BW effect attributable to maternal genotype exceeded that associated with the fetal genotype (maternal: $\beta = 0.048$, $P = 5.1 \times 10^{-15}$; fetal: $\beta = 0.023$, $P = 2.9 \times 10^{-8}$) (Supplementary Table 10 and Extended Data Figs 4, 5). Thus, both maternal and fetal genetic effects connect BW to later T2D risk, albeit acting in opposing

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

directions. When we categorized T2D loci using a classification of physiological functions derived from their effects on related glycaemic and anthropometric traits[27], we found that T2D-risk alleles associated with lower BW were those typically characterized by reduced insulin processing and secretion without detectable changes in fasting glucose (the 'Beta Cell' cluster in Extended Data Fig. 8f).

The *YTK6* signal at rs138715366 is notable not only because the genetic data indicate that a single low-frequency non-coding variant is driving the association signal (see above) but also because of the proximity of this signal to *GCK*. Rare coding variants in glucokinase are causal for a form of monogenic hyperglycaemia and lead to large reductions in BW when parental alleles are passed on to their offspring[29]. In addition, common non-coding variants nearby are implicated in T2D risk and fasting hyperglycaemia[28]. However, the latter variants are conditionally independent of rs138715366 (Supplementary Table 19) and show no comparable association with lower BW. Either rs138715366 acts through effector transcripts other than *GCK*, or the impact of the low-frequency SNP near *YKT6* on *GCK* expression involves tissue- and/or temporal-specific variation in regulatory impact.

In conclusion, we have identified 60 genetic loci associated with BW and used them to gain insights into the aetiology of fetal growth and into well-established, but until now poorly understood, life-course disease associations. The evidence that the relationship between early growth and later metabolic disease has an appreciable genetic component contrasts with, but is not necessarily incompatible with, the emphasis on adverse early environmental events highlighted by the fetal origins hypothesis[1]. As we have shown, these genetic effects reflect variation in both the fetal and the maternal genome: the impact of the latter on the offspring's predisposition to adult disease could be mediated, at least in part, through perturbation of the antenatal and early life environment. Future mechanistic and genetic studies should support reconciliation between these alternative, but complementary, explanations for the far-reaching life-course associations that exist between events in early life and predisposition to cardiometabolic disease several decades later.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 February; accepted 2 September 2016.
Published online 28 September 2016.

1. Barker, D. J. The developmental origins of chronic adult disease. *Acta Paediatr. Suppl.* **93,** 26–33 (2004).
2. The 1000 Genomes Project Consortium An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
3. The UK10K Project Consortium The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).
4. Tyrrell, J. S., Yaghootkar, H., Freathy, R. M., Hattersley, A. T. & Frayling, T. M. Parental diabetes and birthweight in 236,030 individuals in the UK Biobank study. *Int. J. Epidemiol.* **42,** 1714–1723 (2013).
5. Horikoshi, M. *et al.* New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat. Genet.* **45,** 76–82 (2013).
6. Eaves, L. J., Pourcain, B. S., Smith, G. D., York, T. P. & Evans, D. M. Resolving the effects of maternal and offspring genotype on dyadic outcomes in genome wide complex trait analysis ("M-GCTA"). *Behav. Genet.* **44,** 445–455 (2014).
7. Feenstra, B., *et al.* Maternal genome-wide association study identifies a fasting glucose variant associated with offspring birth weight. Preprint at: http://biorxiv.org/content/early/2015/12/11/034207 (2015).
8. Bulik-Sullivan, B. K. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
9. Segrè, A. V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** e1001058 (2010).
10. Baran, Y. *et al.* The landscape of genomic imprinting across diverse adult human tissues. *Genome Res.* **25,** 927–936 (2015).
11. Haig, D. & Westoby, M. Parent-specific gene expression and the triploid endosperm. *Am. Nat.* **134,** 147–155 (1989).
12. Peters, J. The role of genomic imprinting in biology and disease: an expanding view. *Nat. Rev. Genet.* **15,** 517–530 (2014).
13. Johnson, T. *et al.* Blood pressure loci identified with a gene-centric array. *Am. J. Hum. Genet.* **89,** 688–700 (2011).
14. International Consortium for Blood Pressure Genome-Wide Association Studies *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478,** 103–109 (2011).
15. Zhang, G. *et al.* Assessing the causal relationship of maternal height on birth size and gestational age at birth: a Mendelian randomization analysis. *PLoS Med.* **12,** e1001865 (2015).
16. Tyrrell, J. *et al.* Genetic evidence for causal relationships between maternal obesity-related traits and birth weight. *J. Am. Med. Assoc.* **315,** 1129–1140 (2016).
17. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).
18. Diver, L. A. *et al.* Common polymorphisms at the *CYP17A1* locus associate with steroid phenotype: support for blood pressure genome-wide association study signals at this locus. *Hypertension* **67,** 724–732 (2016).
19. Picado-Leonard, J. & Miller, W. L. Cloning and sequence of the human gene for P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase): similarity with the gene for P450c21. *DNA* **6,** 439–448 (1987).
20. Pezzi, V., Mathis, J. M., Rainey, W. E. & Carr, B. R. Profiling transcript levels for steroidogenic enzymes in fetal tissues. *J. Steroid Biochem. Mol. Biol.* **87,** 181–189 (2003).
21. Escobar, J. C., Patel, S. S., Beshay, V. E., Suzuki, T. & Carr, B. R. The human placenta expresses CYP17 and generates androgens *de novo*. *J. Clin. Endocrinol. Metab.* **96,** 1385–1392 (2011).
22. Reynolds, R. M. *et al.* Programming of hypertension: associations of plasma aldosterone in adult men and women with birthweight, cortisol, and blood pressure. *Hypertension* **53,** 932–936 (2009).
23. CARDIoGRAMplusC4D Consortium *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* **45,** 25–33 (2013).
24. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46,** 234–244 (2014).
25. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–1186 (2014).
26. Hattersley, A. T. & Tooke, J. E. The fetal insulin hypothesis: an alternative explanation of the association of low birth weight with diabetes and vascular disease. *Lancet* **353,** 1789–1792 (1999).
27. Dimas, A. S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63,** 2158–2171 (2014).
28. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44,** 981–990 (2012).
29. Hattersley, A. T. *et al.* Mutations in the glucokinase gene of the fetus result in reduced birth weight. *Nat. Genet.* **19,** 268–270 (1998).

Supplementary Information is available in 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: 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.La., 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.Pan., X.W., L.C., F.G., K.E.S., M.Mu., 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.Sta., Z.K., D.M.H., 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.Stu., V.Li., T.A.L., C.M.v.D., A.K., T.I.S., H.N., K.Pah., O.T.R., E.Z., G.V.D., S.-M.S., M.Me., H.C., J.F.W., M.V., J.-C.H., T.H., S.S., 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.Pi., 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.Stu., V.Li., T.A.L., C.M.v.D., W.K., A.K., T.I.S., H.N., K.Pah., O.T.R., G.V.D., Y.-Y.T., S.-M.S., M.Me., 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.Pi., O.P., C.Po., 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. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.I.M (mark.mccarthy@drl.ox.ac.uk) or R.M.F (r.freathy@ex.ac.uk).

Reviewer Information *Nature* thanks J. Whitfield and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [2]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LE, UK. [3]Institute of Biomedical and Clinical Science, University of Exeter Medical School, University of Exeter, Royal Devon and Exeter Hospital, Exeter EX2 5DW, UK. [4]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK. [5]The University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4102, Australia. [6]School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia 6009, Australia. [7]The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3015 CE, the Netherlands. [8]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3015 CE, the Netherlands. [9]Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3015 CE, the Netherlands. [10]Department of Epidemiology Research, Statens Serum Institut, Copenhagen DK-2300, Denmark. [11]Department of Pediatrics, University of California San Diego, La Jolla, San Diego, California 92093, 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 DK-2100, Denmark. [13]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [14]Population Health Research Institute, St George's University of London, Cranmer Terrace, London SW17 0RE, UK. [15]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden 2333 ZA, the Netherlands. [16]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, 2820 Gentofte, Denmark. [17]Steno Diabetes Center, Gentofte DK-2820, Denmark. [18]Department of Computational Biology, University of Lausanne, Lausanne 1011, Switzerland. [19]Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland. [20]Department of Clinical Chemistry, Fimlab Laboratories, Tampere 33520, Finland. [21]Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere 33014, Finland. [22]Institute for Molecular Medicine, Finland (FIMM), University of Helsinki, Helsinki FI-00100, Finland. [23]Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [24]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [25]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [26]Institute of Epidemiology I, Helmholtz Zentrum München- German Research Center for Environmental Health, 85764 Neuherberg, Germany. [27]Division of Metabolic and Nutritional Medicine, Dr. von Hauner Children's Hospital, University of Munich Medical Center, 80337 Munich, Germany. [28]Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Amsterdam 1081 BT, the Netherlands. [29]ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona 08003, Spain. [30]Universitat Pompeu Fabra (UPF), Barcelona 08002, Spain. [31]CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain. [32]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, UK. [33]Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore 119077, Singapore. [34]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. [35]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens 17671, Greece. [36]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku 20014, Finland. [37]Institute of Biomedicine, Physiology, University of Eastern Finland, Kuopio FI-70211, Finland. [38]KUL – University of Leuven, Department of Neurosciences, Leuven 3000, Belgium. [39]Translational Immunology Laboratory, VIB, Leuven 3000, Belgium. [40]Department of Preventive Medicine, Division of Biostatistics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. [41]Institute of Cardiovascular & Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK. [42]Endometriosis CaRe Centre, Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford OX3 9DU, UK. [43]Center for Genomic Regulation (CRG), Barcelona 08003, Spain. [44]Department of Internal Medicine, Internal Medicine, Lausanne University Hospital (CHUV), Lausanne 1011, Switzerland. [45]Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Barcelona 08034, Spain. [46]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland 4000, Australia. [47]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China. [48]Li Ka Shing Institute of Health Sciences, The Chinese University of 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, China. [51]Department of Human Genetics and Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109, USA. [52]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA. [53]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [54]European Centre for Environment and Human Health, University of Exeter, Truro TR1 3HD, UK. [55]Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol BS8 2BN, UK. [56]School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK. [57]Department of Genomics of Common Disease, School of Public Health, Imperial College London, London SW7 2AZ, UK. [58]IFB Adiposity Diseases, University of Leipzig, 04103 Leipzig, Germany. [59]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. [60]FISABIO–Universitat Jaume I–Universitat de València, Joint Research Unit of Epidemiology and Environmental Health, Valencia 46020, Spain. [61]The Children's Obesity Clinic, Department of Pediatrics, Copenhagen University Hospital Holbæk, Holbæk DK-4300, Denmark. [62]Institute of Social and Preventive Medicine, Lausanne University Hospital (CHUV), Lausanne 1010, Switzerland. [63]Danish Center for Neonatal Screening, Statens Serum Institute, Copenhagen DK-2300, Denmark. [64]Department for Congenital Disorders, Statens Serum Institute, Copenhagen DK-2300, Denmark. [65]Computer Sciences Department, Barcelona Supercomputing Center, Barcelona 08034, Spain. [66]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain. [67]Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3015 CE, the Netherlands. [68]BHF Centre for Cardiovascular Science, University of Edinburgh, Queen's Medical Research Institute, Edinburgh EH16 4TJ, UK. [69]Department of Medicine, Division of Endocrinology, Metabolism, and Molecular Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. [70]Medical Department, University of Leipzig, 04103 Leipzig, Germany. [71]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio FI-70029, Finland. [72]Kuopio Research Institute of Exercise Medicine, Kuopio FI-70100, Finland. [73]Pediatric Research Center, Department of Women´s & Child Health, University of Leipzig, 04103 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 DK-2200, Denmark. [75]Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen DK-2000, Denmark. [76]Department of Pediatrics, Turku University Hospital, Turku 20521, Finland. [77]Department of Physiology, University of Turku, Turku 20014, Finland. [78]Paavo Nurmi Centre, Sports and Exercise Medicine Unit, Department of Physical Activity and Health, Turku 20014, Finland. [79]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku 20521, Finland. [80]Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546, Singapore. [81]Life Sciences Institute, National University of Singapore, Singapore 117456, Singapore. [82]Singapore Eye Research Institute, Singapore 168751, Singapore. [83]Department of Clinical Medicine, University of Copenhagen, Copenhagen DK-2200, Denmark. [84]Department of Medicine, Stanford School of Medicine, Stanford, California 94305, USA. [85]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. [86]EMGO Institute for Health and Care Research, VU University and VU University Medical Center, Amsterdam 1081 HV, the Netherlands. [87]Department of Large Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg C DK-1870, Denmark. [88]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu 90014, Finland. [89]Biocenter Oulu, University of Oulu, Oulu 90014, Finland. [90]School of Medicine and Pharmacology, Royal Perth Hospital Unit, The University of Western Australia, Perth, Western Australia 6000, Australia. [91]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital Munich, Ludwig Maximilian University of Munich, 80336 Munich, Germany. [92]Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [93]USC-Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City 6000, Philippines. [94]Department of Nutrition and Dietetics, University of San Carlos, Cebu City 6000, Philippines. [95]National Institute for Health and Welfare, Helsinki 00271, Finland. [96]Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki 00014, Finland. [97]Folkhälsan Research Center, Helsinki 00250, Finland. [98]Department of Clinical Physiology, Tampere University Hopital, Tampere 33521, Finland. [99]Department of Clinical Physiology, University of Tampere School of Medicine, Tampere 33014, Finland. [100]Division of Medicine, Turku University Hospital, Turku 20521, Finland. [101]Department of Medicine, University of Turku, Turku 20014, Finland. [102]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden 2333 ZA, the Netherlands. [1033]Epidemiology Section, BESC Department, King Faisal Specialist Hospital and Research Centre, Riyadh 12713, Saudi Arabia. [104]Research Center for Prevention and Health Capital Region, Center for Sundhed, Rigshospitalet – Glostrup, Copenhagen University, Glostrup DK-2600, Denmark. [105]Population, Policy and Practice, UCL Institute of Child Health, University College London, London WC1N 1EH, UK. [106]Centre for Population Health Research, School of Health Sciences, and Sansom Institute, University of South Australia, Adelaide, South Australia 5001, Australia. [107]South Australian Health and Medical Research Institute, Adelaide, South Australia 5000, Australia. [108]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [109]Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London SW7 2AZ, UK. [110]Unit of Primary Care, Oulu University Hospital, Oulu 90220, Finland. [111]Division of Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [112]Department of Endocrinology, Rigshospitalet, Copenhagen DK-2100, Denmark. [113]AstraZeneca, Innovative Medicines and Early Development | Early Clinical Development, Mölndal 431 83, Sweden. [114]Department of Biostatistics, University of Liverpool, Liverpool L69 3GA, UK. [115]Estonian Genome Center, University of Tartu, Tartu 50090, Estonia. [116]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. [117]Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Oxford OX3 7LE, UK.

*These authors contributed equally to this work.

†A list of consortium members appears in the Supplementary Information.

‡Deceased.

§These authors jointly supervised this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

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

Study-level analyses. No statistical methods were used to predetermine sample size: to maximise power to detect association signals, we set out to collect the largest possible set of samples for which the combination of genome-wide genotyping data and reliable measures of BW could be made available for analysis. Within each study, BW was collected from a variety of sources, including measurements at birth by medical practitioners, obstetric records, medical registers, interviews with the mother and self-report as adults (Supplementary Table 1). BW was *Z*-score transformed separately in males and females. Individuals with extreme BW (>5 s.d. from the sex-specific study mean), monozygotic or polyzygotic siblings, or preterm births (gestational age <37 weeks, where this information was available) were excluded from downstream association analyses (Supplementary Table 1).

Within each study, stringent quality control of the GWAS genotype scaffold was carried out before imputation (Supplementary Table 2). Each scaffold was then pre-phased and imputed[30,31] up to reference panels from the 1000G project[2] or the combined 1000G and UK10K projects[3] (Supplementary Table 2). Association of BW with each variant passing established GWAS quality control filters[32] was tested in a linear regression framework, under an additive model for the allelic effect, after adjustment for study-specific covariates, including gestational age, where available (Supplementary Table 2). Where necessary, population structure was accounted for by adjustment for axes of genetic variation from principal components analysis[33] and subsequent genomic control correction[34], or inclusion of a genetic relationship matrix in a mixed model[35] (Supplementary Table 2). We calculated the genomic control inflation factor (λ) in each study to confirm that study-level population structure was accounted for before meta-analysis.

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

Genotype data from the May 2015 release were available for a subset of 152,249 participants from UK Biobank. In addition to the quality control metrics performed centrally by UK Biobank, we defined a subset of 'white European' ancestry samples using a *K*-means ($K=4$) clustering approach based on the first four genetically determined principal components. A maximum of 67,786 individuals (40,425 females and 27,361 males) with genotype and valid BW measures were available for downstream analyses. We tested for association with BW, assuming an additive allelic effect, in a linear mixed model implemented in BOLT-LMM (ref. 37) to account for cryptic population structure and relatedness. Genotyping array was included as a binary covariate in all models. Total chip heritability (that is, the variance explained by all autosomal polymorphic genotyped SNPs passing quality control) was calculated using restricted maximum likelihood (REML) implemented in BOLT-LMM (ref. 37). We additionally analysed the association between BW and directly genotyped SNPs on the X chromosome: for this analysis, we used 57,715 unrelated individuals with BW available and identified by UK Biobank as white British. We excluded SNPs with evidence of deviation from Hardy–Weinberg equilibrium ($P<1\times10^{-6}$), MAF <0.01 or overall missing rate >0.015, resulting in 19,423 SNPs for analysis in Plink v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/)[38], with the first five ancestry principal components as covariates.

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

European ancestry meta-analysis. The European ancestry meta-analysis consisted of two components: (i) 75,891 individuals from 30 GWAS from Europe, USA and Australia; and (ii) 67,786 individuals of white European origin from the UK Biobank. In the first component, we combined sex-specific BW association summary statistics across studies in a fixed-effects meta-analysis, implemented in GWAMA (ref. 39) and applied a second round of genomic control[34] ($\lambda_{GC}=1.001$). Subsequently, we combined association summary statistics from this component with the UK Biobank in a European ancestry fixed-effects meta-analysis, implemented in GWAMA (ref. 39). Variants failing GWAS quality control filters in the UK Biobank, reported in less than 50% of the total sample size in the first component, or with MAF <0.1%, were excluded from the European ancestry meta-analysis. We aggregated X-chromosome association summary statistics from the UK Biobank (19,423 SNPs) with corresponding statistics from the European GWAS component using fixed effects *P*-value-based meta-analysis in METAL (ref. 40) (max $n=99{,}152$).

We were concerned that self-reported BW as adults in the UK Biobank would not be comparable with that obtained from more stringent collection methods used in other European ancestry GWAS. In addition, the UK Biobank lacked information on gestational age for adjustment, which could have an impact on strength of association compared with the results obtained from other European ancestry GWAS. However, we observed no evidence of heterogeneity in BW allelic effects at lead SNPs between the two components of European ancestry meta-analysis, using Cochran's *Q* statistic[41] implemented in GWAMA (ref. 39) after Bonferroni correction ($P>0.00083$) (Supplementary Table 3). We tested for heterogeneity in allelic effects between studies within the European component using Cochran's *Q*. At loci demonstrating evidence of heterogeneity, we confirmed that association signals were not driven by outlying studies by visual inspection of forest plots. We performed sensitivity analyses to assess the impact of covariate adjustment (gestational age and population structure) on heterogeneity.

We were also concerned that overlap of individuals (duplicated or related) between the two components of the European ancestry meta-analysis might lead to false positive association signals. We performed bivariate linkage-disequilibrium score regression[8] using the two components of the European ancestry meta-analysis and observed a genetic covariance intercept of 0.0156 ± 0.0058 (s.e.), indicating a maximum of 1,119 duplicate individuals. Univariate linkage-disequilibrium score regression[8] of the European ancestry meta-analysis estimated the intercept as 1.0426, which may indicate population structure or relatedness that was not adequately accounted for in the analysis. To assess the impact of this inflation on the European ancestry meta-analysis, we expanded the standard errors of BW allelic effect size estimates and re-calculated association *P* values. On the basis of this adjusted analysis, only the lead SNP at *MTNR1B* dropped below genome-wide significance (rs10830963, $P=5.5\times10^{-8}$).

Trans-ancestry meta-analysis. The trans-ancestry meta-analysis combined the two European ancestry components with an additional 10,104 individuals from six GWAS from diverse ancestry groups: African American, Chinese, Filipino, Surinamese, Turkish and Moroccan. Within each GWAS, we first combined sex-specific BW association summary statistics in a fixed-effects meta-analysis, implemented in GWAMA (ref. 39) and applied a second round of genomic control[34]. Subsequently, we combined association summary statistics from the six non-European GWAS and the two European ancestry components in a trans-ancestry fixed-effects meta-analysis, implemented in GWAMA (ref. 39). Variants failing GWAS quality control filters in the UK Biobank, reported in less than 50% of the total sample size in the first component, or with MAF <0.1%, were excluded from the trans-ancestry meta-analysis. We tested for heterogeneity in allelic effects between ancestries using Cochran's *Q* (ref. 41).

Approximate conditional analysis. We searched for multiple distinct BW association signals in each of the established and novel loci, defined as 1 Mb up- and down-stream of the lead SNP from the trans-ancestry meta-analysis, through approximate conditional analysis. We applied GCTA (ref. 42) to identify 'index SNPs' for distinct association signals attaining genome-wide significance ($P<5\times10^{-8}$) in the European ancestry meta-analysis using a reference sample of 5,000 individuals of white British origin, randomly selected from the UK Biobank, to approximate patterns of linkage disequilibrium between variants in these regions. Note that we performed approximate conditioning on the basis of only the European ancestry meta-analysis because GCTA cannot accommodate linkage-disequilibrium variation between diverse populations.

Prioritizing candidate genes in each BW locus. We combined a number of approaches to prioritize the most likely candidate gene(s) in each BW locus. Expression quantitative trait loci (eQTLs) were obtained from the Genotype Tissue Expression (GTEx) Project[43], the GEUVADIS project[44] and eleven other studies[45–55] using HaploReg v4 (ref. 56). We interrogated coding variants for each BW lead SNP and its proxies (EUR $R^2>0.8$) using Ensembl[57] and HaploReg. Their likely functional consequences were predicted by SIFT (ref. 58) and PolyPhen2 (ref. 59). Biological candidacy was assessed by presence in significantly enriched gene set pathways from MAGENTA analyses (see below for details). We extracted all genes within 300 kb of all lead BW SNPs and searched for connectivity between

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

any genes using STRING (ref. 60). If two or more genes between two separate BW loci were connected, they were given an increased prior for both being plausible candidates. We also applied protein–protein interaction (PPI) analysis (see below for details) to all genes within 300 kb of each lead BW SNPs and ranked the genes based on the score for connectivity with the surrounding genes.

Evaluation of imputation quality of the low-frequency variant at the *YKT6–GCK* locus. At the *YKT6–GCK* locus, the lead SNP (rs138715366) was found at a low frequency in European ancestry populations (MAF = 0.92%) and was even rarer in other ancestry groups (MAF = 0.23% in African Americans, otherwise monomorphic) and was not present in the HapMap reference panel[61]. To assess the accuracy of imputation for this low-frequency variant, we genotyped rs138715366 in the Northern Finland Birth Cohort (NFBC) 1966 (Supplementary Table 1). Of the 5,009 samples in the study, 4,704 were successfully imputed and genotyped (or sequenced) for rs138715366. The overall concordance rate between imputed and directly assayed genotypes was 99.8% and for directly assayed heterozygote calls was 75.0%.

Fine-mapping analyses. We investigated linkage-disequilibrium differences between populations contributing to the trans-ancestry meta-analysis and to take advantage of the improved coverage of common and low-frequency variation offered by 1000G or 1000G and UK10K combined imputation to localize variants driving each distinct association signal achieving locus-wide significance. For each distinct signal, we used MANTRA (ref. 62) to construct 99% credible sets of variants[63] that together account for 99% of the posterior probability of driving the association. MANTRA incorporates a prior model of relatedness between studies, based on mean pair-wise allele frequency differences across loci, to account for heterogeneity in allelic effects (Supplementary Table 3). MANTRA has been demonstrated, by simulation, to improve localization of causal variants compared with either a fixed- or random-effects trans-ancestry meta-analysis[62,64].

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

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

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

where the summation is over all variants mapping within the (conditional) meta-analysis across the locus. In this expression, Λ_j is the Bayes' factor in favour of association from the MANTRA analysis. A 99% credible set[63] was then constructed by: (i) ranking all variants according to their Bayes' factor, Λ_j; and (ii) including ranked variants until their cumulative posterior probability exceeds 0.99.

Genomic annotation. We used genomic annotations of DNaseI hypersensitive sites (DHS) from the ENCODE (ref. 65 project and protein coding genes from GENCODE (ref. 66). We filtered cell types that are cancer cell lines (karyotype 'cancer' from https://genome.ucsc.edu/ENCODE/cellTypes.html), and merged data from multiple samples from the same cell type. This resulted in 128 DHS cell-type annotations, as well as 4 gene-based annotations (coding exon, 5′UTR, 3′UTR and 1 kb upstream of the transcription start site (TSS)). First, we tested for the effect of each cell type DHS and gene annotation individually using the Bayes' factors for all variants in the 62 credible sets using fgwas (ref. 67). Second, we categorized the annotations into 'genic', 'fetal DHS', 'embryonic DHS', 'stem cell DHS', 'neonatal DHS' and 'adult DHS' based on the description fields from ENCODE, and tested for the effect of each category individually as described above using fgwas. Third, we then tested the effect of each category by including all categories in a joint model using fgwas. For each of the three analyses, we obtained the estimated effects and 95% confidence intervals (CI) for each annotation, and considered an annotation enriched if the 95% CI did not overlap zero.

Estimation of genetic variance explained. The 'variance explained' statistic was calculated using the REML method implemented in GCTA (ref. 68). We considered the variance explained by two sets of SNPs: (i) lead SNPs of all 62 distinct association signals at the 59 established and novel autosomal BW loci identified in the European-specific or trans-ancestry meta-analyses; (ii) lead SNPs of 55 distinct association signals at the 52 novel autosomal BW loci (Extended Data Table 1a and Supplementary Table 7). The 'variance explained' was calculated in samples of European ancestry in the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study[69] (independent of the meta-analysis) and two studies that were part of the European ancestry meta-analysis: NFBC1966 and Generation R (Supplementary Table 1). In each study, the genetic relationship matrix was estimated for each set of SNPs and was tested individually against BW (males and females combined) with study specific covariates. These analyses provided an estimate and s.e. for the variance explained by each of the given sets of SNPs.

Examining the relative effects on BW of maternal and fetal genotype at the 60 identified loci. We performed four sets of analyses. First, we used GWAS data from 4,382 mother–child pairs in the Avon Longitudinal Study of Parents and Children (ALSPAC) study to fit a 'maternal-GCTA model'[6] to estimate the extent to which the maternal genome might influence offspring BW independent of the fetal genome. The maternal-GCTA model uses genome-wide genetic similarity between mothers and offspring to partition the phenotypic variance in BW into components due to the maternal genotype, the child's genotype, the covariance between the two and environmental sources of variation.

Second, we compared associations with BW of the fetal versus maternal genotype at each of the 60 BW loci. The maternal allelic effect on offspring BW was obtained from a maternal GWAS meta-analysis of 68,254 European mothers from the EGG Consortium ($n = 19{,}626$)[7] and the UK Biobank ($n = 48{,}628$). In the UK Biobank, mothers were asked to report the BW of their first child. Women of European ancestry with genotype data available in the May 2015 data release were included, and those with reported BW equivalent to <2.5 kg or >4.5 kg were excluded. No information on gestational age or gender of child was available. BW of first child was associated with maternal factors such as smoking status, BMI and height in the expected directions. Of the 68,254 women included in the maternal GWAS, 13% were mothers of individuals included in the current fetal European ancestry GWAS, and a further ~45% were themselves (with their own BW) included in the fetal GWAS.

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

Fourth, we used the method of Zhang *et al.*[15] to test associations between BW and the maternal untransmitted, maternal transmitted and inferred paternal transmitted haplotype score of 422 height SNPs[25], 30 SBP SNPs[13,14] and 84 T2D SNPs[24] in 5,201 mother–child pairs from the ALSPAC study.

Linkage-disequilibrium score regression. The use of linkage-disequilibrium score regression to estimate the genetic correlation between two traits/diseases has been described in detail elsewhere[70]. Briefly, the linkage-disequilibrium score is a measure of how much genetic variation each variant tags; if a variant has a high linkage-disequilibrium score then it is in high linkage disequilibrium with many nearby polymorphisms. Variants with high linkage-disequilibrium scores are more likely to contain more true signals and hence provide more chance of overlap with genuine signals between GWAS. The linkage-disequilibrium score regression method uses summary statistics from the GWAS meta-analysis of BW and the other traits of interest, calculates the cross-product of test statistics at each SNP, and then regresses the cross-product on the linkage-disequilibrium score. Bulik-Sullivan *et al.*[70] show that the slope of the regression is a function of the genetic covariance between traits:

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

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

Summary statistics from the GWAS meta-analysis for traits and diseases of interest were downloaded from the relevant consortium website. The summary statistics files were reformatted for linkage-disequilibrium score regression analysis using the munge_sumstats.py python script provided on the developer's website (https://github.com/bulik/ldsc). For each trait, we filtered the summary statistics to the subset of HapMap 3 SNPs[71], as advised by the developers, to ensure that no bias was introduced due to poor imputation quality. Summary statistics from the European-specific BW meta-analysis were used because of the variable linkage-disequilibrium structure between ancestry groups. Where the sample size for each SNP was included in the results file this was flagged using N-col; if no sample size was available then the maximum sample size reported in the reference for the GWAS meta-analysis was used. SNPs were excluded for the

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

following reasons: MAF < 0.01; ambiguous strand; duplicate rsID; non-autosomal SNPs; reported sample size less than 60% of the total available. Once all files were reformatted, we used the ldsc.py python script, also on the developers' website, to calculate the genetic correlation between BW and each of the traits and diseases. The European linkage-disequilibrium score files calculated from the 1000G reference panel and provided by the developers were used for the analysis. Where multiple GWAS meta-analyses had been conducted on the same phenotype (that is, over a period of years), the genetic correlation with BW was estimated using each set of summary statistics and presented in Supplementary Table 12. The phenotypes with multiple GWAS included height, BMI, waist–hip ratio (adjusted for BMI), total cholesterol, triglycerides, high density lipoprotein (HDL) and low density lipoprotein (LDL). The estimate of the genetic correlation between the multiple GWAS meta-analyses on the same phenotype were comparable and the later GWAS had a smaller standard error due to the increased sample size, so only the genetic correlation between BW and the most recent meta-analyses were presented in Fig. 2.

In the published GWAS for blood pressure[14] the phenotype was adjusted for BMI. Caution is needed when interpreting the genetic correlation between BW and BMI-adjusted SBP owing to the potential for collider bias[72]. Since BMI is associated with both blood pressure and BW, it is possible that the use of a blood pressure genetic score adjusted for BMI might bias the genetic correlation estimate towards a more negative value. To verify that the inverse genetic correlation with BW ($r_g = -0.26$, s.e. = 0.05, $P = 6.5 \times 10^{-9}$) was not due to collider bias caused by the BMI adjustment of the phenotype, we obtained an alternative estimate using UK Biobank GWAS data for SBP that was unadjusted for BMI and obtained a similar result ($R_g = -0.22$, s.e. = 0.03, $P = 5.5 \times 10^{-13}$). The SBP phenotype in the UK Biobank was prepared as follows. Two blood pressure readings were taken at assessment, approximately 5 min apart. We included all individuals with an automated blood pressure reading (taken using an automated Omron blood pressure monitor). Two valid measurements were available for most participants (averaged to create a blood pressure variable, or alternatively a single reading was used if only one was available). Individuals were excluded if the two readings differed by more than 4.56 s.d. Blood pressure measurements more than 4.56 s.d. away from the mean were excluded. We accounted for blood pressure medication use by adding 15 mm Hg to the SBP measure. Blood pressure was adjusted for age, sex and centre location and then inverse rank normalized. We performed the GWAS on 127,698 individuals of British descent using BOLT-LMM (ref. 37), with genotyping array as covariate.

Estimating the proportion of the BW-adult traits covariance attributable to genotyped SNPs. We estimated the phenotypic, genetic and residual correlations as well as the genetic and residual covariance between BW and several quantitative traits and/or disease outcomes in the UK Biobank using directly genotyped SNPs and the REML method implemented in BOLT-LMM (ref. 37). The traits examined included T2D, SBP, diastolic blood pressure, CAD, height, BMI, weight, waist–hip ratio, hip circumference, waist circumference, obesity, overweight, age at menarche, asthma, and smoking. Where phenotypes were not available (for example, serum blood measures are not currently available in the UK Biobank), we obtained estimates using the NFBC1966 study (for correlations/covariance between BW and triglycerides, total cholesterol, HDL, LDL, fasting glucose and fasting insulin). In the UK Biobank analysis, we used 57,715 unrelated individuals with BW available and identified by the UK Biobank as white British. SNPs with evidence of deviation from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$), MAF < 0.05 or overall missing rate >0.015 were excluded, resulting in 328,928 SNPs for analysis. We included the first five ancestry principal components as covariates. In the NFBC1966 analysis, 5,009 individuals with BW were enrolled. Genotyped SNPs that passed quality control (Supplementary Table 2) were included, resulting in 324,895 SNPs for analysis. The first three ancestry principal components and sex were included as covariates.

Gene set enrichment analysis. Meta-analysis gene-set enrichment of variant associations (MAGENTA) was used to explore pathway-based associations using summary statistics from the trans-ancestry meta-analysis. MAGENTA implements a gene set enrichment analysis (GSEA) based approach, as previously described[9]. Briefly, each gene in the genome was mapped to a single index SNP with the lowest *P* value within a 110 kb upstream and 40 kb downstream window. This *P* value, representing a gene score, was then corrected for confounding factors such as gene size, SNP density and linkage-disequilibrium-related properties in a regression model. Genes within the HLA-region were excluded from analysis due to difficulties in accounting for gene density and linkage-disequilibrium patterns. Each mapped gene in the genome was then ranked by its adjusted gene score. At a given significance threshold (95th and 75th percentiles of all gene scores), the observed number of gene scores in a given pathway, with a ranked score above the specified threshold percentile, was calculated. This observed statistic was then compared to 1,000,000 randomly permuted pathways of identical size. This generates an empirical GSEA *P* value for each pathway. Significance was attained when an individual pathway reached a FDR < 0.05 in either analysis. In total, 3,216 pre-defined biological pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with BW. The MAGENTA software was also used for enrichment testing of custom gene sets.

Protein–protein interaction network analyses. We used the integrative protein-interaction-network-based pathway analysis (iPINBPA) method[73]. Briefly, we generated gene-wise *P* values from the trans-ancestry meta-analysis using VEGAS2 (ref. 74), which mapped the SNPs to genes and accounted for possible confounders, such as linkage-disequilibrium between markers. The empirical gene-wise *P* values were calculated using simulations from the multivariate normal distribution. Those that were nominally significant ($P \leq 0.01$) were selected as 'seed genes', and were collated within a high confidence version of inweb3 (ref. 75) to weight the nodes in the network following a guilt-by-association approach. In a second step, a network score was defined by the combination of the *Z* scores derived from the gene-wise *P* values with node weights using the Liptak-Stouffer method[76]. A heuristic algorithm was then applied to extensively search for modules enriched in genes with low *P* values. The modules were further normalized using a null distribution of 10,000 random networks. Only those modules with *Z* score > 5 were selected. Finally, the union of all modules constructed a BW-overall PPI network. Both the proteins on the individual modules and on the overall BW-PPI were interrogated for enrichment in Gene Ontology terms (biological processes) using a hypergeometric test. Terms were considered as significant when the adjusted *P* value, following the Benjamini–Hochberg procedure, was below 0.05.

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

Parent-of-origin specific associations. We first searched for evidence of parent-of-origin effects in the UK Biobank samples by comparing variance between heterozygotes and homozygotes using Quicktest (ref. 77). In this analysis, we used only unrelated individuals identified genetically as of white British origin ($n = 57{,}715$). Principal components were generated using these individuals and the first five were used to adjust for population structure as covariates in the analysis, in addition to a binary indicator for genotyping array.

We also examined 4,908 mother–child pairs in ALSPAC and determined the parental origin of the alleles where possible[78]. Briefly, the method used mother–child pairs to determine the parent of origin of each allele. For example, if the mother/child genotypes were AA/Aa, the child's maternal/paternal allele combination was A/a. For the situation where both mother and child were heterozygous, the child's maternal/paternal alleles could not be directly specified. However, the parental origin of the alleles could be determined by phasing the genotype data and comparing maternal and child haplotypes. We then tested these alleles for association with BW adjusting for sex and gestational age.

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

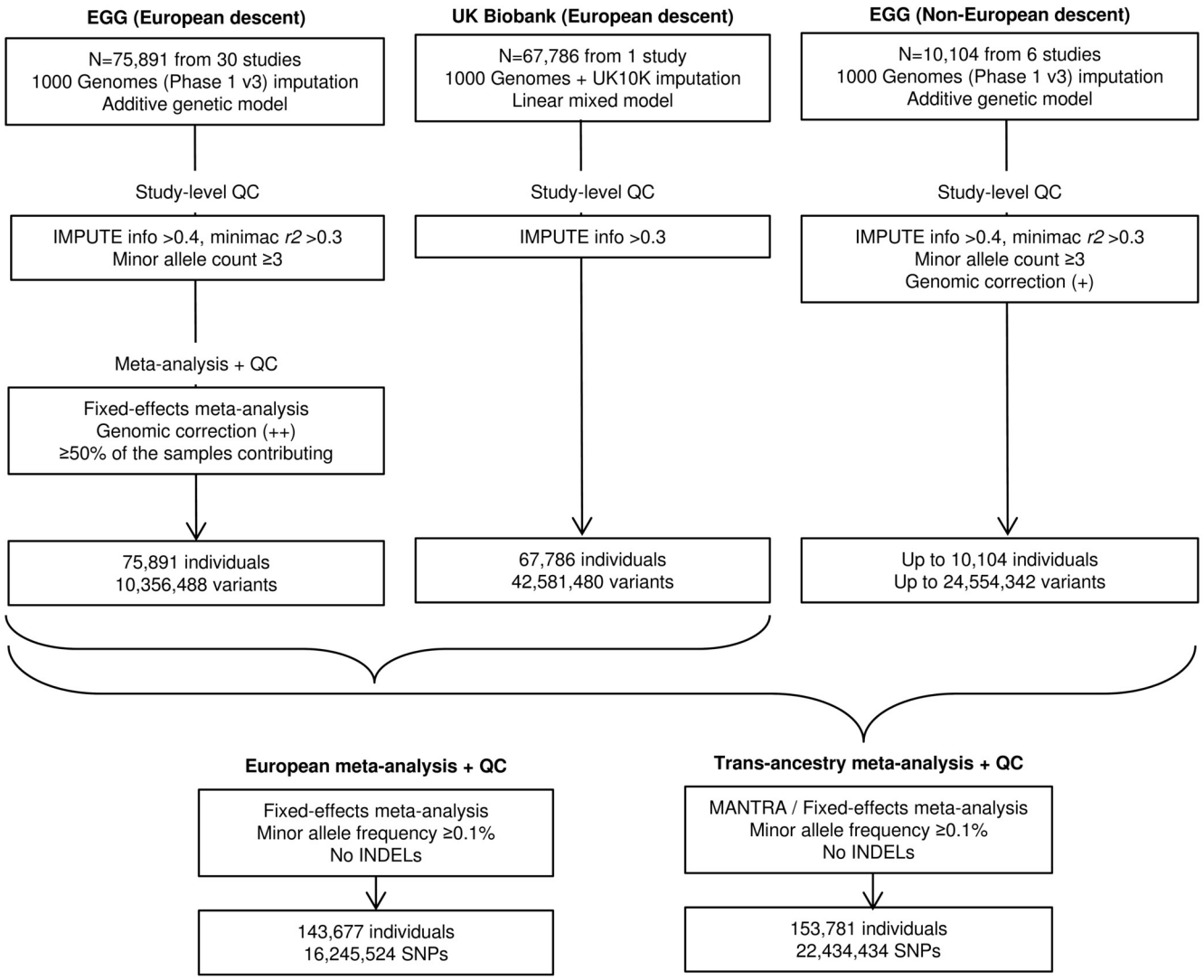
Hierarchical clustering of BW loci. To explore the different patterns of association between BW and other anthropometric/metabolic/endocrine traits and diseases, we performed hierarchical clustering analysis. The lead SNP (or proxy, EUR $R^2 > 0.6$) at the 60 BW loci was queried in publicly available GWAS meta-analysis datasets or in GWAS results obtained through collaboration[79]. Results were available for 53 of those loci and the extracted *Z* score (allelic effect/s.e., Supplementary Table 17) was aligned to the BW-raising allele. We performed two dimensional clustering by trait and by locus. We computed the Euclidean distance amongst *Z* scores of the extracted traits and loci and performed complete hierarchical clustering implemented in the pvclust package (http://www.sigmath.es.osaka-u.ac.jp/shimo-lab/prog/pvclust/) in R v3.2.0 (http://www.R-project.org/). Clustering uncertainty was measured by multiscale bootstrap resampling estimated from 1,000 replicates. We used $\alpha = 0.05$ to define distinct clusters and, based on the bootstrap analysis, calculated the Calinski index to identify the number of well-supported clusters (cascadeKM function, vegan package, http://CRAN.R-project.org/package=vegan). Clustering was visualized by constructing dendrograms and a heat map.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Separately from the hierarchical clustering analysis, we queried the lead SNP at *EPAS1* in a GWAS of haematological traits[80] because variation at that locus has previously been implicated in BW and adaptation to hypoxia at high altitudes in Tibetans[81,82] (Supplementary Table 17).

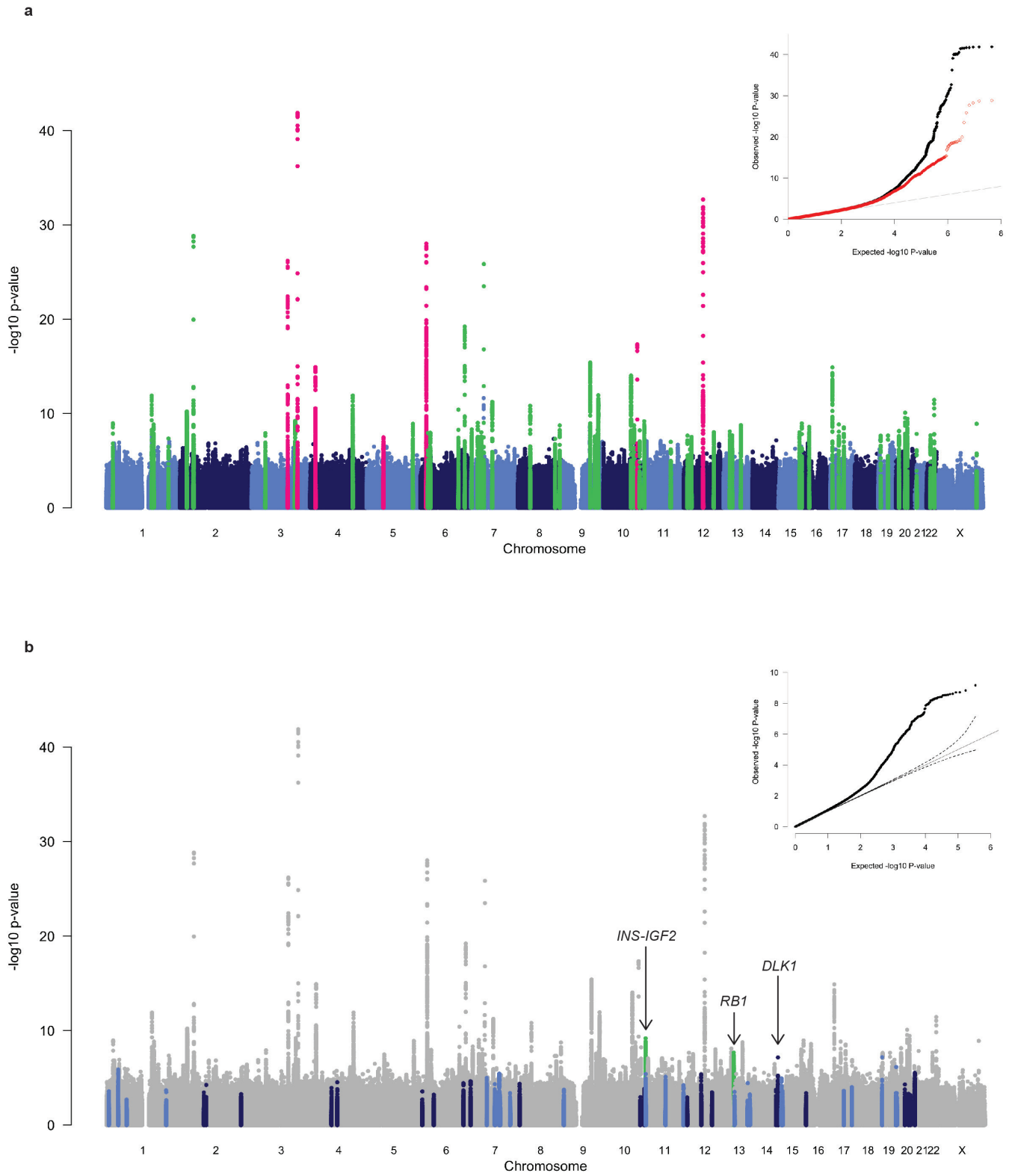
30. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11,** 499–511 (2010).
31. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44,** 955–959 (2012).
32. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9,** 1192–1212 (2014).
33. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
34. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
35. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42,** 348–354 (2010).
36. Allen, N. E., Sudlow, C., Peakman, T. & Collins, R. UK Biobank data: come and get it. *Sci. Transl. Med.* **6,** 224ed4 (2014).
37. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).
38. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
39. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11,** 288 (2010).
40. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
41. Ioannidis, J. P., Patsopoulos, N. A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One* **2,** e841 (2007).
42. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369–375, S1–S3 (2012).
43. GTEx Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).
44. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501,** 506–511 (2013).
45. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464,** 773–777 (2010).
46. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6,** e107 (2008).
47. Gibbs, J. R. *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6,** e1000952 (2010).
48. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39,** 1217–1224 (2007).
49. Li, Q. *et al.* Expression QTL-based analyses reveal candidate causal genes and loci across five tumor types. *Hum. Mol. Genet.* **23,** 5294–5302 (2014).
50. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45,** 1238–1243 (2013).
51. Zou, F. *et al.* Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.* **8,** e1002707 (2012).
52. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8,** e1003029 (2012).
53. Koopmann, T. T. *et al.* Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS One* **9,** e97380 (2014).
54. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343,** 1246949 (2014).
55. Grundberg, E. *et al.* Global analysis of the impact of environmental perturbation on cis-regulation of gene expression. *PLoS Genet.* **7,** e1001279 (2011).
56. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40,** D930–D934 (2012).
57. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42,** D749–D755 (2014).
58. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–1081 (2009).
59. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).
60. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–D452 (2015).
61. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52–58 (2010).
62. Morris, A. P. Transethnic meta-analysis of genome-wide association studies. *Genet. Epidemiol.* **35,** 809–822 (2011).
63. The Wellcome Trust Case Control Consortium. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44,** 1294–1301 (2012).
64. Wang, X. *et al.* Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **22,** 2303–2311 (2013).
65. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
66. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22,** 1760–1774 (2012).
67. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94,** 559–573 (2014).
68. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42,** 565–569 (2010).
69. Urbanek, M. *et al.* The chromosome 3q25 genomic region is associated with measures of adiposity in newborns in a multi-ethnic genome-wide association study. *Hum. Mol. Genet.* **22,** 3583–3596 (2013).
70. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–1241 (2015).
71. The International HapMap Consortium *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).
72. Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96,** 329–339 (2015).
73. Wang, L., Mousavi, P. & Baranzini, S. E. iPINBPA: an integrative network-based functional module discovery tool for genome-wide association studies. *Pac. Symp. Biocomput.* 255–266 (2015).
74. Mishra, A. & Macgregor, S. VEGAS2: software for more flexible gene-based testing. *Twin Res. Hum. Genet.* **18,** 86–91 (2015).
75. Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25,** 309–316 (2007).
76. Whitlock, M. C. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.* **18,** 1368–1373 (2005).
77. Hoggart, C. J. *et al.* Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index. *PLoS Genet.* **10,** e1004508 (2014).
78. Wang, S., Yu, Z., Miller, R. L., Tang, D. & Perera, F. P. Methods for detecting interactions between imprinted genes and environmental exposures using birth cohort designs with mother-offspring pairs. *Hum. Hered.* **71,** 196–208 (2011).
79. Painter, J. N. *et al.* Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat. Genet.* **43,** 51–54 (2011).
80. Ganesh, S. K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* **41,** 1191–1198 (2009).
81. Xu, X. H. *et al.* Two functional loci in the promoter of EPAS1 gene involved in high-altitude adaptation of Tibetans. *Sci. Rep.* **4,** 7465 (2014).
82. Huerta-Sánchez, E. *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512,** 194–197 (2014).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



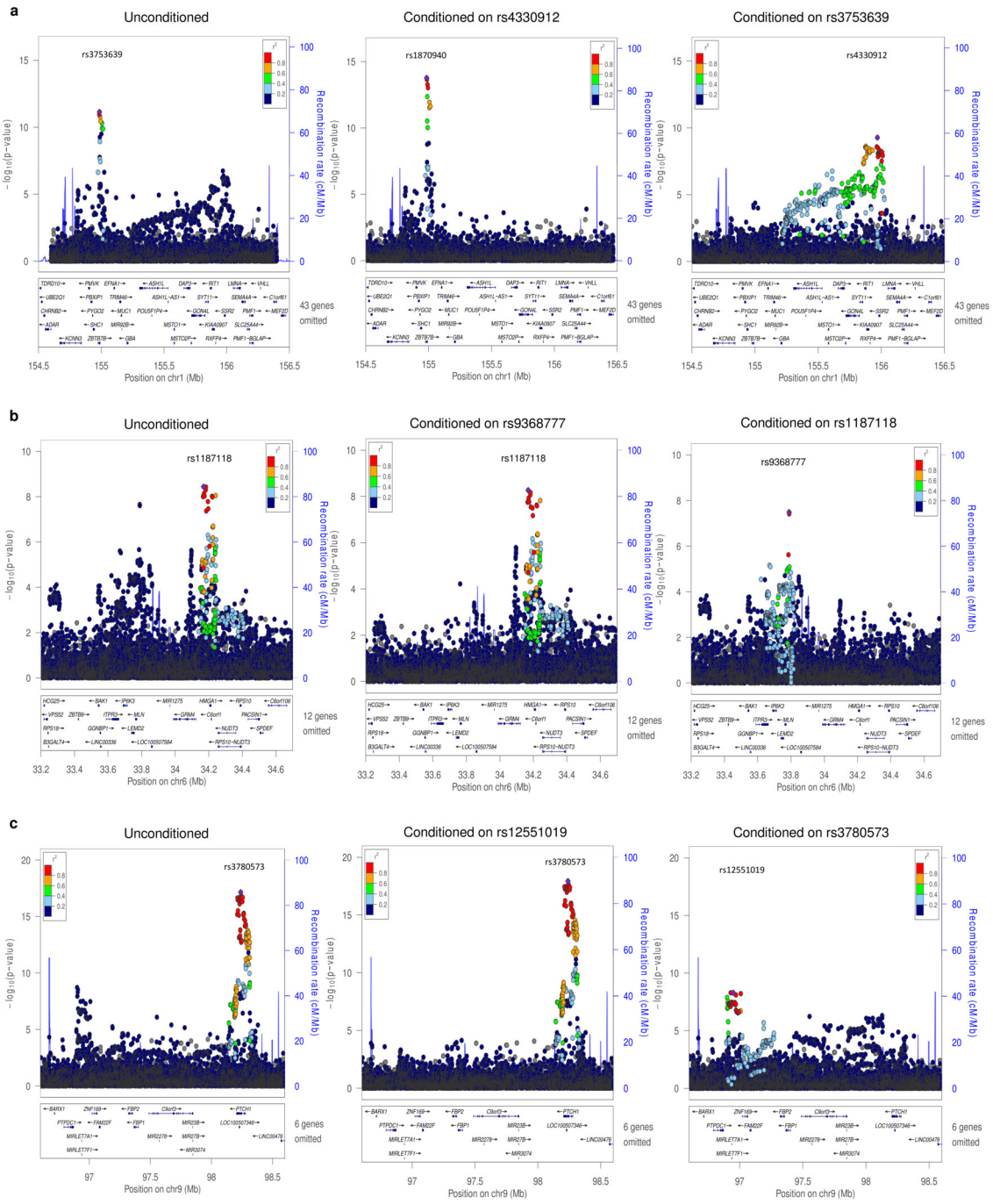
Extended Data Figure 1 | Flow chart of the study design.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



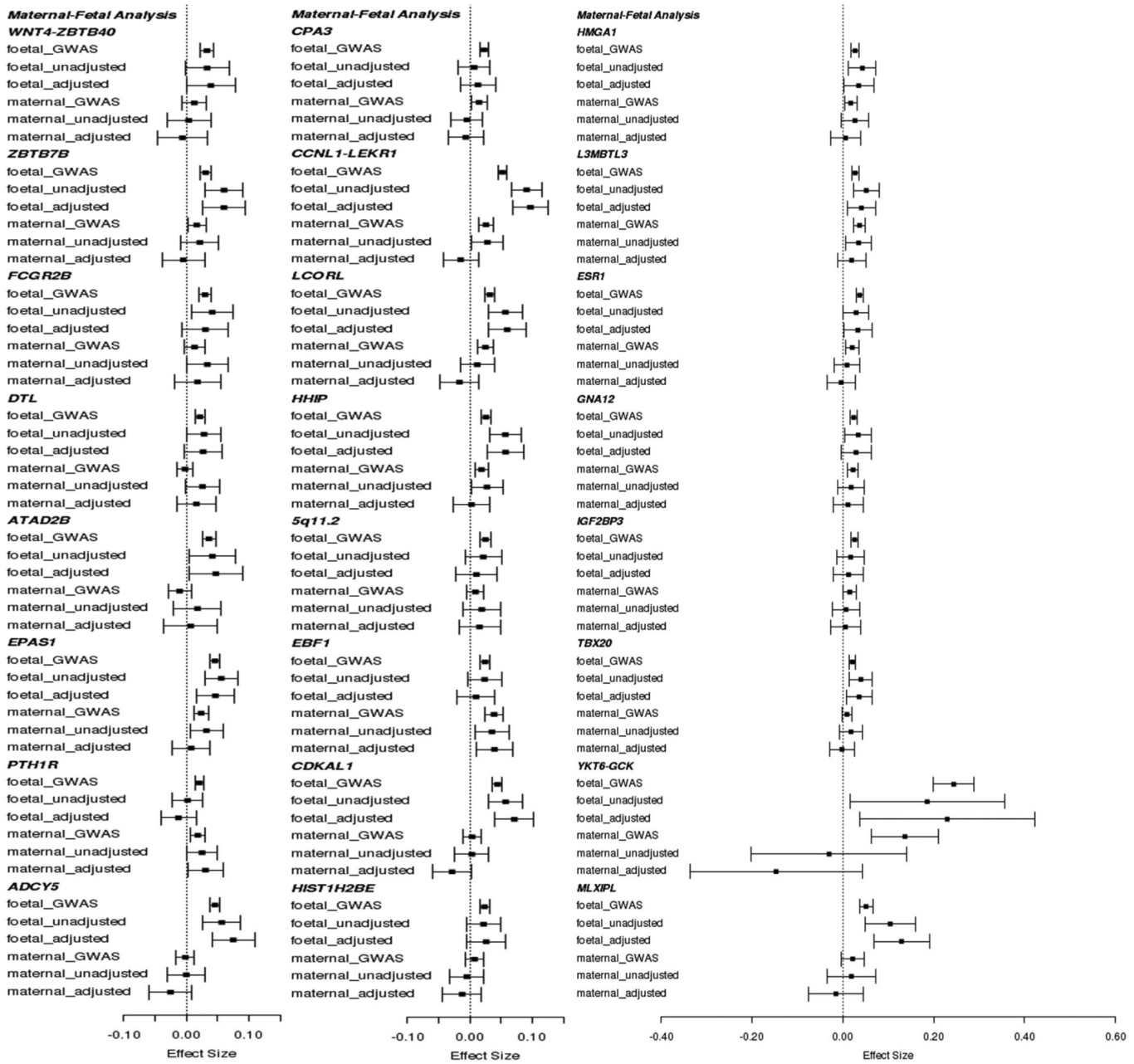
Extended Data Figure 2 | Manhattan and quantile–quantile (QQ) plots of the trans-ancestry meta-analysis for BW. a, Manhattan (main panel) and QQ (top right) plots of genome-wide association results for BW from trans-ancestry meta-analysis of up to 153,781 individuals. The association *P* value (on $-\log_{10}$ scale) for each of up to 22,434,434 SNPs (*y* axis) was 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 signals[5]. **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.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



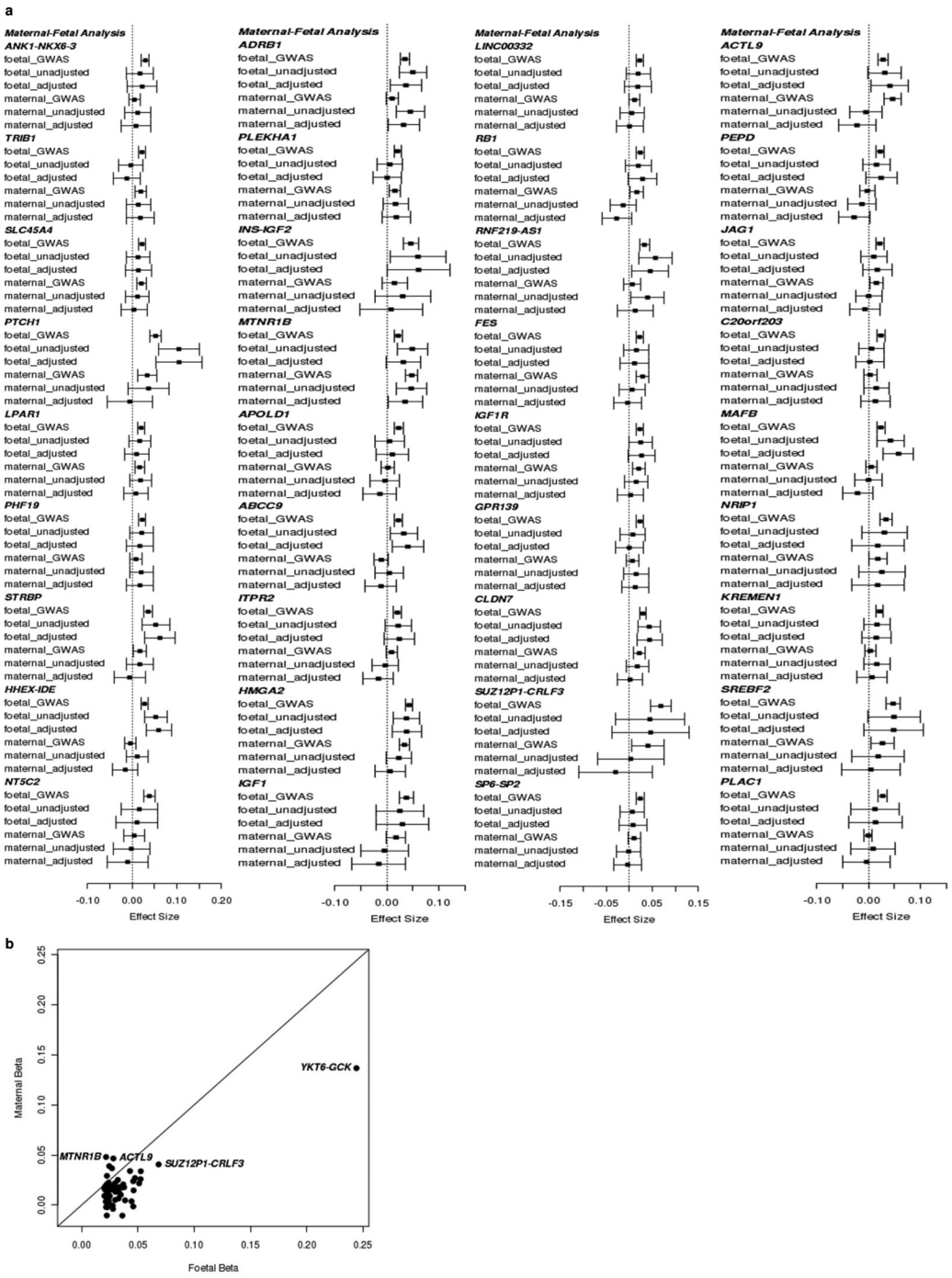
Extended Data Figure 3 | Regional plots for multiple distinct signals at three BW loci. Regional plots for each locus, *ZBTB7B* (**a**), *HMGA1* (**b**) and *PTCH1* (**c**), are displayed from: the unconditional European-specific meta-analysis of up to 143,677 individuals (left); the approximate conditional 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 were 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) were plotted to reflect the local linkage-disequilibrium structure around the index SNPs and their correlated proxies. SNPs were coloured in reference to linkage-disequilibrium 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.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



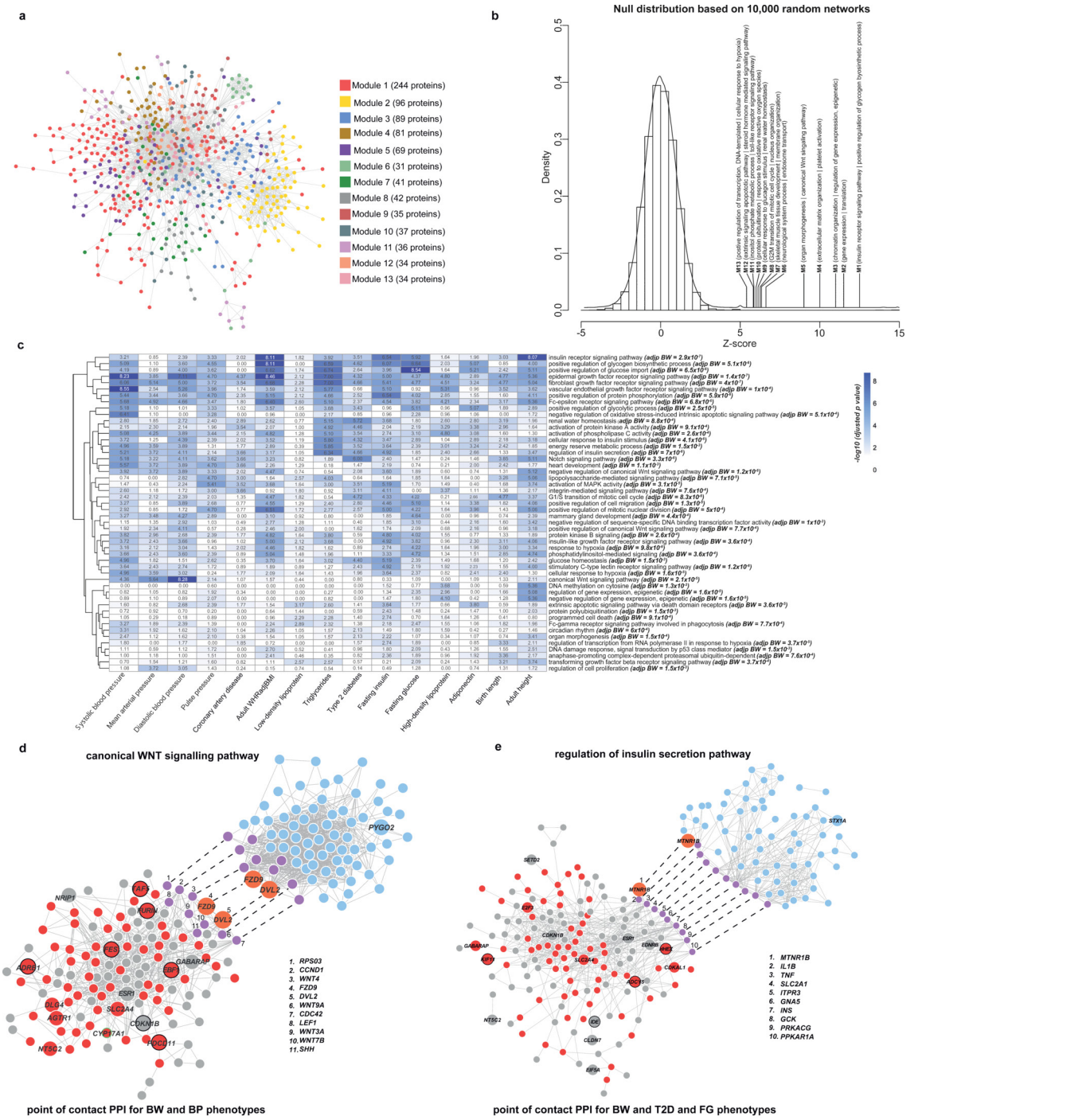
Extended Data Figure 4 | Comparison of fetal effect sizes and maternal effect sizes at 60 known and novel birth weight loci, for the first 24 loci. The remaining loci are shown in Extended Data Fig. 5a. For each BW locus, the following six effect sizes (with 95% CI) are shown, all aligned to the same BW-raising allele: fetal_GWAS, fetal allelic effect on BW (from European ancestry meta-analysis of up to $n = 143,677$ individuals); fetal_unadjusted, fetal allelic effect on BW (unconditioned in $n = 12,909$ mother–child pairs); fetal_adjusted, fetal 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 fetal genotype, $n = 12,909$). The 60 BW loci were ordered by chromosome and position (Supplementary Tables 10, 11). These plots illustrate that, in large GWAS of BW, fetal effect size estimates are larger than those of maternal at 55 out of 60 identified loci (binomial $P = 1 \times 10^{-11}$), suggesting that most of the associations are driven by the fetal genotype. In conditional analyses that modelled the effects of both maternal and fetal genotypes ($n = 12,909$ mother–child pairs), confidence intervals around the estimates were wide, precluding inference about the likely contribution of maternal versus fetal genotype at individual loci.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



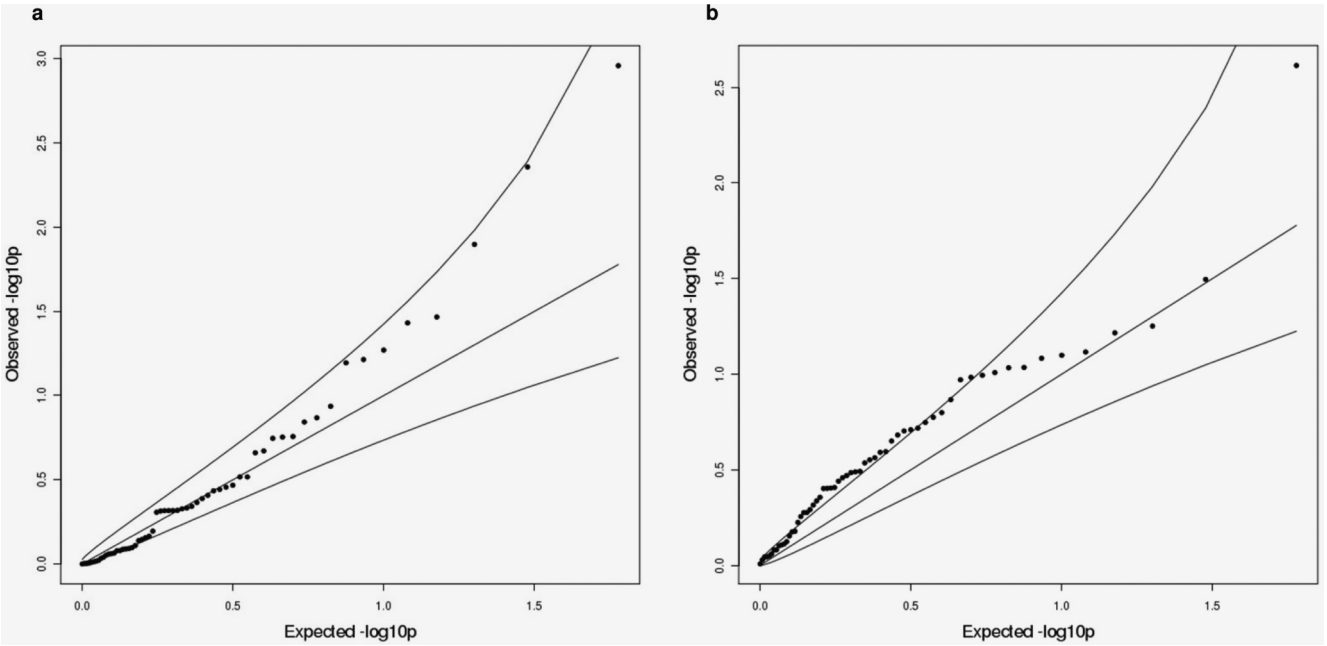
Extended Data Figure 5 | Comparison of fetal effect sizes and maternal effect sizes at 60 known and novel birth weight loci, for the remaining 36 loci. a, Continued from Extended Data Fig. 4. **b**, The scatter plot illustrates the difference between the fetal (*x* axis) and maternal (*y* axis) effect sizes in the overall maternal versus fetal GWAS results.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



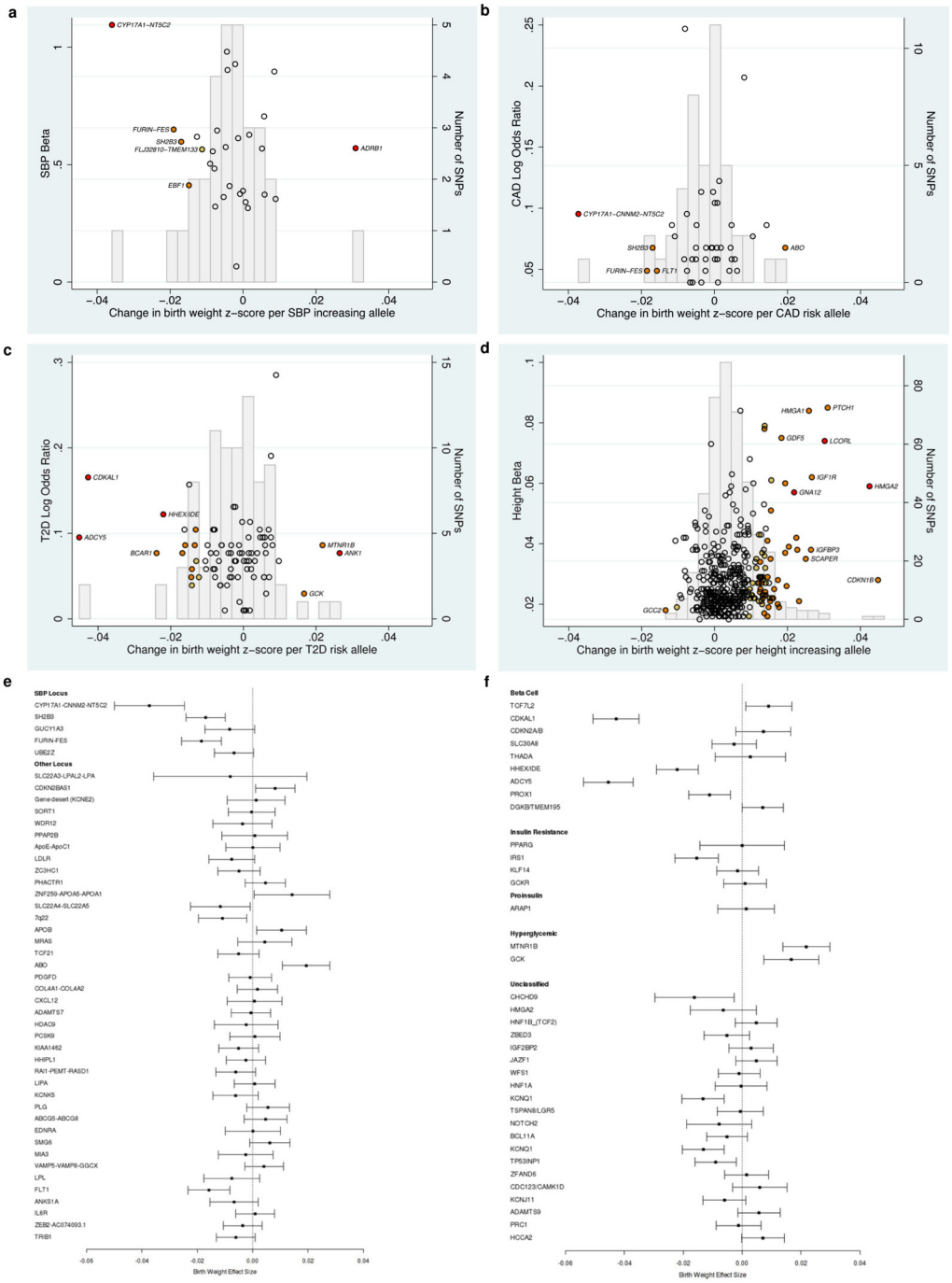
Extended Data Figure 6 | Protein–protein Interaction (PPI) Network analysis. a, The largest global component of BW PPI network containing 13 modules is shown. **b**, The histogram shows the null distribution of *Z* scores of BW PPI networks based on 10,000 random networks, and where the *Z* scores for the 13 BW modules (M1–13) lie. For each module, the two most significant GO terms are shown. **c**, A heat map is shown, which takes the top 50 biological processes over-represented in the global BW PPI network (listed at the right of the plot), and displays the 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 T2D/fasting glucose (FG). Red nodes indicate those 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 300 kb from the lead SNP), and have a black border if they, amongst all BW loci, have a stronger (top 5) association with at least one of the pairing adult traits.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Quantile–Quantile (QQ) plots of variance comparison between heterozygotes and homozygotes analysis in 57,715 UK Biobank samples and parent-of-origin specific analysis in 4,908 ALSPAC mother–child pairs at 59 autosomal BW loci plus *DLK1*. **a**, QQ plot from the Quicktest analysis (ref. 77) 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 versus 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.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 8 | Summary of previously reported loci for SBP, CAD, T2D and adult height and their effect on birth weight. a–d, Effect sizes (left *y* axis) of previously reported 30 SBP loci[13,14], 45 CAD loci[23], 84 T2D loci[24] and 422 adult height loci[25] were plotted against effects on BW (*x* axis). Effect sizes were aligned to the adult trait (or risk)-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 were 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 blood pressure association. **f**, Effect sizes (with 95% CI) on BW of 32 known T2D loci were plotted, subdivided by previously reported categories derived from detailed adult physiological data[27]. Heterogeneity in BW effect sizes between five T2D loci groups with different mechanistic categories was substantial (Cochran's *Q* statistic $P_{het} = 1.2 \times 10^{-9}$). In pairwise comparisons, the 'beta cell' group of variants differed from the other four groups: fasting hyperglycaemia ($P_{het} = 3 \times 10^{-11}$), insulin resistance ($P_{het} = 0.002$), proinsulin ($P_{het} = 0.78$) and unclassified ($P_{het} = 0.02$) groups. All of the BW effect sizes plotted in the forest plots were aligned to the trait (or risk)-raising allele.

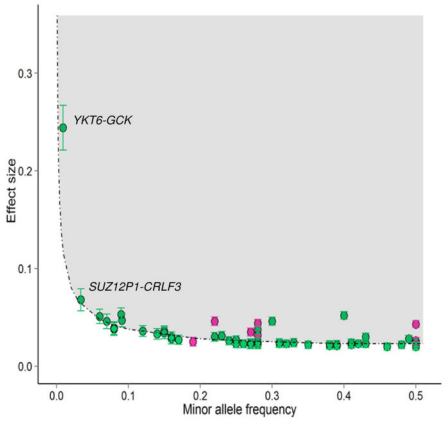
© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Sixty loci associated with BW ($P < 5 \times 10^{-8}$) in European ancestry meta-analysis of up to 143,677 individuals and/or trans-ancestry meta-analysis of up to 153,781 individuals

a

Locus	Lead SNP	Chr.	Position (bp, b37)	Alleles Effect/Other	EAF	European ancestry β (SE)	European ancestry *P*-value	Trans-ancestry β (SE)	Trans-ancestry *P*-value
Previously reported loci									
CCNL1-LEKR1	rs13322435	3	156,795,468	A/G	0.59	0.053 (0.004)	3.7×10^{-41}	0.052 (0.004)	1.3×10^{-42}
HMGA2	rs1351394	12	66,351,826	T/C	0.48	0.044 (0.004)	1.9×10^{-32}	0.043 (0.004)	2.0×10^{-33}
CDKAL1	rs35261542	6	20,675,792	C/A	0.73	0.044 (0.004)	4.4×10^{-27}	0.044 (0.004)	9.7×10^{-29}
ADCY5	rs11719201	3	123,068,744	T/C	0.23	0.046 (0.004)	2.4×10^{-26}	0.046 (0.004)	6.4×10^{-27}
ADRB1	rs7076938	10	115,789,375	T/C	0.73	0.036 (0.004)	4.7×10^{-18}	0.035 (0.004)	4.7×10^{-18}
LCORL	rs925098	4	17,919,811	G/A	0.28	0.034 (0.004)	5.4×10^{-16}	0.032 (0.004)	1.3×10^{-15}
5q11.2	rs854037	5	57,091,783	A/G	0.80	0.027 (0.005)	2.2×10^{-8}	0.025 (0.005)	3.5×10^{-8}
Novel loci									
EPAS1	rs1374204	2	46,484,205	T/C	0.70	0.047 (0.004)	6.2×10^{-29}	0.046 (0.004)	1.5×10^{-29}
YKT6-GCK	rs138715366	7	44,246,271	C/T	0.99	0.241 (0.023)	7.2×10^{-26}	0.244 (0.023)	1.4×10^{-26}
ESR1	rs1101081	6	152,032,917	C/T	0.73	0.038 (0.004)	1.6×10^{-19}	0.037 (0.004)	6.1×10^{-20}
PTCH1	rs28510415	9	98,245,026	G/A	0.09	0.056 (0.007)	1.5×10^{-17}	0.053 (0.006)	4.0×10^{-18}
CLDN7	rs113086489	17	7,171,356	T/C	0.55	0.031 (0.004)	9.1×10^{-16}	0.030 (0.004)	1.3×10^{-15}
HHEX-IDE	rs61862780	10	94,468,643	T/C	0.52	0.028 (0.004)	3.0×10^{-14}	0.028 (0.004)	9.5×10^{-15}
STRBP	rs700059	9	125,824,055	G/A	0.16	0.033 (0.005)	4.7×10^{-10}	0.036 (0.005)	1.2×10^{-12}
HHIP	rs6537307	4	145,601,863	G/A	0.48	0.025 (0.004)	9.5×10^{-12}	0.026 (0.004)	1.3×10^{-12}
ZBTB7B	rs3753639	1	154,986,091	C/T	0.23	0.031 (0.004)	7.3×10^{-12}	0.031 (0.004)	1.3×10^{-12}
SREBF2	rs62240962	22	42,259,524	C/T	0.92	0.047 (0.007)	9.7×10^{-12}	0.047 (0.007)	3.7×10^{-12}
MLXIPL	rs62466330	7	73,056,805	C/T	0.07	0.049 (0.008)	1.2×10^{-10}	0.051 (0.007)	5.9×10^{-12}
ANK1-NKX6-3	rs13266210	8	41,533,514	A/G	0.79	0.031 (0.005)	1.3×10^{-11}	0.030 (0.004)	1.6×10^{-11}
L3MBTL3	rs1415701	6	130,345,835	G/A	0.73	0.025 (0.004)	2.6×10^{-9}	0.027 (0.004)	4.0×10^{-11}
ATAD2B	rs7575873	2	23,962,647	A/G	0.88	0.038 (0.006)	1.3×10^{-11}	0.036 (0.006)	6.2×10^{-11}
C20orf203	rs28530618	20	31,275,581	A/G	0.50	0.026 (0.004)	7.7×10^{-12}	0.024 (0.004)	8.4×10^{-11}
MAFB	rs6016377	20	39,172,728	T/C	0.45	0.024 (0.004)	9.5×10^{-10}	0.024 (0.004)	3.7×10^{-10}
CPA3	rs10935733	3	148,622,968	T/C	0.42	0.022 (0.004)	9.2×10^{-9}	0.023 (0.004)	6.2×10^{-10}
INS-IGF2	rs72851023	11	2,130,620	T/C	0.07	0.048 (0.008)	2.9×10^{-10}	0.046 (0.007)	6.8×10^{-10}
IGF2BP3	rs11765649	7	23,479,013	T/C	0.76	0.027 (0.004)	5.8×10^{-10}	0.026 (0.004)	1.0×10^{-9}
WNT4-ZBTB40	rs2473248	1	22,536,643	C/T	0.87	0.033 (0.006)	1.1×10^{-8}	0.033 (0.005)	1.1×10^{-9}
IGF1R	rs7402982	15	99,193,269	A/G	0.42	0.023 (0.004)	2.3×10^{-9}	0.023 (0.004)	1.1×10^{-9}
PLAC1	rs11096402	X	133,827,868	G/A	0.25	0.028 (0.005)	1.3×10^{-9}	N/A	N/A
EBF1	rs7729301	5	157,886,953	A/G	0.72	0.024 (0.004)	1.6×10^{-8}	0.025 (0.004)	1.3×10^{-9}
SUZ12P1-CRLF3	rs144843919	17	29,037,339	G/A	0.96	0.066 (0.012)	1.4×10^{-8}	0.068 (0.011)	1.5×10^{-9}
FCGR2B	rs72480273	1	161,644,871	C/A	0.17	0.031 (0.005)	8.0×10^{-10}	0.030 (0.005)	1.5×10^{-9}
RNF219-AS1	rs1819436	13	78,580,283	C/T	0.87	0.033 (0.006)	6.3×10^{-8}	0.033 (0.005)	1.8×10^{-9}
NT5C2	rs74233809	10	104,913,940	C/T	0.08	0.037 (0.007)	5.2×10^{-8}	0.039 (0.006)	1.8×10^{-9}
SLC45A4	rs12543725	8	142,247,979	G/A	0.60	0.023 (0.004)	1.2×10^{-8}	0.022 (0.004)	1.9×10^{-9}
GPR139	rs1011939	16	19,992,996	G/A	0.31	0.022 (0.004)	1.3×10^{-7}	0.024 (0.004)	2.7×10^{-9}
SP6-SP2	rs12942207	17	45,968,294	C/T	0.30	0.022 (0.004)	5.1×10^{-8}	0.024 (0.004)	3.0×10^{-9}
GNA12	rs798489	7	2,801,803	C/T	0.74	0.023 (0.004)	2.0×10^{-8}	0.024 (0.004)	5.0×10^{-9}
PHF19	rs7847628	9	123,631,225	G/A	0.67	0.023 (0.004)	1.0×10^{-8}	0.023 (0.004)	5.4×10^{-9}
PLEKHA1	rs2421016	10	124,167,512	T/C	0.48	0.021 (0.004)	1.8×10^{-8}	0.021 (0.004)	6.1×10^{-9}
JAG1	rs6040076	20	10,658,882	C/G	0.51	0.023 (0.004)	2.0×10^{-8}	0.022 (0.004)	7.2×10^{-9}
LINC00332	rs2324499	13	40,662,001	G/C	0.67	0.022 (0.004)	7.3×10^{-8}	0.023 (0.004)	8.3×10^{-9}
IGF1	rs7964361	12	102,994,878	A/G	0.08	0.039 (0.007)	4.7×10^{-8}	0.038 (0.007)	9.7×10^{-9}
FES	rs12906125	15	91,427,612	G/A	0.69	0.023 (0.004)	1.7×10^{-8}	0.023 (0.004)	1.0×10^{-8}
TBX20	rs6959887	7	35,295,365	A/G	0.61	0.023 (0.004)	1.5×10^{-8}	0.021 (0.004)	1.0×10^{-8}
HMGA1	rs7742369	6	34,165,721	G/A	0.19	0.028 (0.005)	1.0×10^{-8}	0.027 (0.005)	1.1×10^{-8}
HIST1H2BE	rs9379832	6	26,186,200	A/G	0.71	0.023 (0.004)	6.6×10^{-8}	0.024 (0.004)	1.2×10^{-8}
PTH1R	rs2242116	3	46,941,116	A/G	0.39	0.022 (0.004)	1.4×10^{-8}	0.021 (0.004)	1.2×10^{-8}
NRIP1	rs2229742	21	16,339,172	G/C	0.87	0.036 (0.006)	2.2×10^{-9}	0.034 (0.006)	1.5×10^{-8}
RB1	rs2854355	13	48,882,363	G/A	0.26	0.023 (0.004)	9.8×10^{-8}	0.024 (0.004)	2.2×10^{-8}
KREMEN1	rs134594	22	29,468,456	C/T	0.35	0.023 (0.004)	1.0×10^{-8}	0.022 (0.004)	2.2×10^{-8}
APOLD1	rs11055034	12	12,890,626	C/A	0.73	0.022 (0.004)	1.8×10^{-7}	0.023 (0.004)	2.3×10^{-8}
PEPD	rs10402712	19	33,926,013	A/G	0.27	0.022 (0.004)	4.4×10^{-7}	0.023 (0.004)	2.3×10^{-8}
ACTL9	rs61154119	19	8,787,750	T/G	0.84	0.028 (0.005)	1.1×10^{-7}	0.028 (0.005)	2.3×10^{-8}
LPAR1	rs2150052	9	113,945,067	T/A	0.50	0.021 (0.004)	2.2×10^{-8}	0.020 (0.004)	2.8×10^{-8}
ITPR2	rs12823128	12	26,872,730	T/C	0.56	0.021 (0.004)	1.9×10^{-8}	0.020 (0.004)	3.2×10^{-8}
DTL	rs61830764	1	212,289,976	A/G	0.36	0.022 (0.004)	5.6×10^{-8}	0.022 (0.004)	4.5×10^{-8}
TRIB1	rs6989280	8	126,508,746	G/A	0.70	0.022 (0.004)	2.2×10^{-7}	0.022 (0.004)	5.0×10^{-8}
MTNR1B	rs10830963	11	92,708,710	G/C	0.27	0.023 (0.004)	2.9×10^{-8}	0.022 (0.004)	1.0×10^{-7}
ABCC9	rs139975827	12	22,068,161	G/A	0.63	0.025 (0.004)	1.1×10^{-8}	0.022 (0.004)	1.0×10^{-7}

b



a, Effects (β values) were aligned to the BW-raising allele. Effect allele frequency (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; b37, build 37; EAF, effect allele frequency; SE, standard error. **b**, The effect of the lead SNP (absolute value of β, *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.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Gene set enrichment analysis and protein–protein interaction (PPI) analysis

a. Gene set enrichment analysis

Database	Gene set	Number of genes (mapped to MAGENTA)	95th percentile enrichment cutoff			75th percentile enrichment cutoff		
			P	FDR	Expected (observed) number of genes	*P*	FDR	Expected (observed) number of genes
GOTERM	Positive regulation of glycogen biosynthetic process	10 (10)	5.6×10^{-5}	0.005	1 (5)	3.6×10^{-3}	0.18	3 (7)
GOTERM	Insulin-like growth factor receptor binding	13 (13)	2.4×10^{-5}	0.006	1 (6)	0.02	0.35	3 (7)
GOTERM	Positive regulation of glucose import	22 (22)	1.0×10^{-4}	0.019	1 (7)	0.02	0.36	6 (10)
GOTERM	Insulin receptor signalling pathway	35 (34)	2.8×10^{-5}	0.022	2 (9)	4.3×10^{-3}	0.27	9 (16)
GOTERM	Chromatin remodelling complex	11 (9)	9.0×10^{-4}	0.036	0 (4)	0.16	0.55	2 (4)
KEGG	Glycosphyngolipid biosynthesis globo-series	14 (13)	2.6×10^{-3}	0.037	1 (4)	0.21	0.48	3 (5)
KEGG	Melanoma	71 (67)	1.6×10^{-3}	0.037	3 (10)	0.05	0.35	17 (23)
KEGG	Terpenoid backbone biosynthesis	15 (15)	5.9×10^{-3}	0.039	1 (1)	0.15	0.44	4 (6)
KEGG	Type 2 Diabetes Mellitus	47 (45)	2.2×10^{-3}	0.040	2 (8)	0.14	0.46	11 (15)
Panther	Cholesterol biosynthesis	11 (11)	1.8×10^{-3}	0.040	1 (4)	0.29	0.64	3 (4)
BIOCARTA	Growth hormone pathway	28 (27)	3.0×10^{-4}	0.044	1 (7)	0.11	0.25	7 (10)
KEGG	Oocyte meiosis	114 (108)	1.0×10^{-3}	0.048	5 (14)	0.07	0.45	27 (34)
Custom gene set of imprinted genes								
GTEX	Imprinted genes (All)	77 (72)	1.9×10^{-4}	-	4 (12)	0.11	-	18 (23)
GTEX	Imprinted genes (Primary)	38 (35)	6.9×10^{-3}	-	2 (6)	0.14	-	9 (12)
GTEX	Imprinted genes (Primary + Suggestive)	55 (50)	0.010	-	3 (7)	0.25	-	13 (15)

b. Protein-protein interaction analysis

Database	Pathway	Number of genes (overlapped with PPI network)	*Z* score	*P*	adjusted *P*[a]
GOTERM	Epidermal growth factor receptor signalling pathway	198 (31)	7.97	3.3×10^{-10}	1.4×10^{-7}
GOTERM	Insulin receptor signalling pathway	151 (26)	7.90	1.1×10^{-9}	2.9×10^{-7}
GOTERM	Stimulatory C-type lectin receptor signalling pathway	121 (22)	7.59	7.5×10^{-9}	1.2×10^{-6}
GOTERM	Negative regulation of canonical Wnt signalling pathway	152 (25)	7.46	6.2×10^{-9}	1.2×10^{-6}
GOTERM	Notch signalling pathway	129 (22)	7.21	2.6×10^{-8}	3.3×10^{-6}
GOTERM	Cellular response to insulin stimulus	71 (16)	7.62	3.7×10^{-8}	4.1×10^{-6}
GOTERM	Positive regulation of glycogen biosynthetic process	15 (8)	9.39	5.3×10^{-8}	5.1×10^{-6}
GOTERM	Positive regulation of protein phosphorylation	114 (20)	7.03	6.8×10^{-8}	5.9×10^{-6}
GOTERM	Positive regulation of glucose import	27 (10)	8.42	8.3×10^{-8}	6.5×10^{-6}
GOTERM	Fc-epsilon receptor signalling pathway	186 (26)	6.58	9.6×10^{-8}	6.8×10^{-6}

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 shown for custom sets of imprinted genes: Primary, genes identified as highly likely to be imprinted in the GTEx database (tested $n = 38$); Primary + suggestive, genes identified as highly likely and suggestively imprinted in GTEx ($n = 55$); All, the above plus genes selected from the literature where imprinting status is consistent in GTEx ($n = 77$). **b**, The results of a complementary analysis of empirical PPI data, displaying the top 10 most significant pathways enriched for BW-association scores.

[a]*P* value is adjusted for multiple correction using the Benjamini–Hochberg method.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.