GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI


Published in:
Science Advances

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2019 the authors. This is an open access article published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

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

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI

Alexassender Couto Alves1,2*, N. Maneka G. De Silva1+, Ville Karhunen1, Ulla Sovio3,4, Shikta Das1,5, H. Rob Taal6,7, Nicole M. Warrington8,9, Alexandra M. Lewin1,10, Marika Kaakinen11,12,13, Diana L. Cousminer14,15,16, Elisabeth Thiering17,18, Nicholas J. Timpson19,20, Tom A. Bond1, Estelle Lowry21, Christopher D. Brown22, Xavier Estivill23,24,25,26,27, Virpi Lindi28, Jonathan P. Bradfield29, Frank Geller30,31, Doug Speed30,31, Lachlan J. M. Coin1,32, Marie Loh1,33, Sheila J. Barton34,35, Lawrence J. Bellin36, Hans Bisgaard37, Klaus Bønnelykke37, Rohia Alili38, Ida J. Hatoum39,38,39,40, Katharina Schramm41,42, Rufus Cartwright43, Marie-Aline Charles44, Vincenzo Salerno1, Karine Clément38,44, Annique J. A. Claringbould45, BIOS Consortium, Cornelia M. van Duijn46, Elena Moltchanova47, Johan G. Eriksson48,49,50, Cathy Elks51, Bjarke Feenstra29, Claudia Flexeder17, Stephen Franks45, Timothy M. Frayling52, Rachel M. Freathy52, Paul Elliott53,54, Elisabeth Widén16, Hakon Hakonarson14,18,35,56, Andrew T. Hattersley52, Alina Rodriguez1,57, Marco Banterle10, Joachim Heinrich17, Barbara Heude44, John W. Holloway58, Albert Hofman6,66, Elina Hyppönen59,60,61, Hazel Insik34,62, Lee M. Kaplan39,40, Asa K. Hedman53,64, Esa Läära65, Holger Prokisch41,42, Harald Grallert66,67, Timo A. Lakka15,68,69, Debbie A. Lawlor19,20, Mads Melbye29,70,71, Tanveerun S. Ahluwalia37, Marcella Marinelli25,26,72, Iona Y. Millwood73,74, Lyle J. Palmer75, Craig E. Pennell8, John R. Perry51, Susan M. Ring19,20,76, Markku J. Savolainen77, Fernando Rivadeneira46,78, Marie Stanisl9, Jordi Sunyer24,25,26,72, Carla M. T. Tiesler17,18, Andre G. Uitterlinden36,78, William Schierding79, Justin M. O’Sullivan79,80, Inga Prokopenko11,13,63,81, Karl-Heinz Herzig82,83,84,85, George Davey Smith19,20, Paul O’Reilly1,86, Janine F. Felix67,46, Jessica L. Buxton87, Alexandra I. F. Blakemore88,89, Ken K. Ong41, Vincent W. V. Jaddoe6,46, Struan F. A. Grant14,28,55,56†, Sylvain Sebert1,21,82†, Mark I. McCarthy63,81,90†‡, Marjo-Riitta Järvelin1,2,12,21,82,84,88,91†‡, Early Growth Genetics (EGG) Consortium

Early childhood growth patterns are associated with adult health, yet the genetic factors and the developmental stages involved are not fully understood. Here, we combine genome-wide association studies with modeling of longitudinal growth traits to study the genetics of infant and child growth, followed by functional, pathway, genetic correlation, risk score, and colocalization analyses to determine how developmental timings, molecular pathways, and genetic determinants of these traits overlap with those of adult health. We found a robust overlap between the genetics of child and adult body mass index (BMI), with variants associated with adult BMI acting as early as 4 to 6 years old. However, we demonstrated a completely distinct genetic makeup for peak BMI during infancy and childhood, followed by functional, pathway, and environmental factors driving child growth may have a lasting influence on maintaining health. (4). Within this framework, identification of the genetic determinants of the critical periods in child development is important for understanding the mechanisms underpinning adult health and preventing disease development.

To date, we have gained considerable insights into the shared genetic makeup of childhood and adult BMI (5, 6). These previous studies were designed to identify genetic variants associated with BMI and obesity acting through the ages of 2 to 18 years. However, BMI does not remain constant, or follow a linear pattern throughout life, particularly not from birth until the age of adiposity rebound (AR) (7, 8). On the contrary, the BMI trajectory in healthy individuals (fig. S1) encompasses three periods characterized by (i) a rapid increase in BMI up to the age of 9 months [adiposity peak (AP)], (ii) a rapid decline in BMI up to the age of 5 to 6 years [adiposity rebound timepoint (AR)], followed by (iii) a steady increase until early adulthood, when BMI growth rate decelerates. We have yet to determine whether changes in timing, velocity, or amplitude of this trajectory, during infancy and childhood, are influenced by specific genetic factors, acting at different developmental stages. The identification of genetic determinants of early growth traits is a fundamental step toward understanding the etiology of obesity and could be important in informing future strategies to prevent and treat it.

The present study set out to model sex-specific individual postnatal growth velocity and BMI curves in children using high-density longitudinal data collected from primary health care or clinical research.
visits. We first conducted a genome-wide association study (GWAS) on six harmonized early growth traits: peak height velocity (PHV), peak weight velocity (PWV), age at AR (Age-AR), and BMI at AR (BMI-AR). We then analyzed the GWAS summary statistics for these six early growth traits to gain insights into the genes and molecular pathways involved and to assess the overlap between the genetic etiology of early growth traits and adult phenotypes. In particular, we tracked the changes in the genetic determinants of BMI occurring throughout infancy, later childhood, and adulthood.

## Results

We conducted two-stage meta-analyses of GWASs on six early growth traits: PHV (in centimeters per month), PWV (in kilograms per month), Age-AR (in years), BMI-AP (in kilograms per square meter), Age-AR (in years), and BMI-AR (in kilograms per square meter). Figure S2 summarizes the study design, while participant characteristics, genotyping arrays, imputation and quality control for the discovery, and follow-up studies are presented in tables S1 and S2 and fig. S3. In the discovery stage (stage 1), we meta-analyzed GWAS from four...
population-based studies comprising between 6051 and 7215 term-born children of European ancestry that had both genetic and early growth trait data (stage 1; see Methods, table S1, and fig. S4). From the stage 1 inverse variance meta-analyses, we selected a total of eight loci with either $P < 1 \times 10^{-7}$ or $P < 1 \times 10^{-5}$ in/near genes known to be associated with obesity and metabolic traits in published GWAS or candidate gene studies. Table S3 shows selection criteria, false discovery rate (FDR), and bias-reduced effect size estimates for the selected single-nucleotide polymorphisms (SNPs). In stage 2 meta-analysis, we followed up these results in up to 16,550 term-born children from up to 11 additional studies (stage 2; see Methods and table S2). In the combined stage 1 + 2 meta-analysis of the discovery and follow-up studies (including up to 22,769 children), we identified a common variant in each of the four independent loci, associated at $P < 5 \times 10^{-8}$ with one or more of the early growth traits (Table 1, Fig. 1, and fig. S5).

**AR SNPs associate with adult BMI**

Three of the four SNPs were associated with Age-AR and BMI-AR. These three variants were previously associated ($P < 5 \times 10^{-8}$) with adult BMI and adult weight in the literature (stage 4) and in the UK Biobank PheWAS (phenome-wide association study) (9) (table S5), as well as with several adiposity-related phenotypes in PhenoScanner (10) (see Methods). The lead SNPs at each of these three loci were the following: rs1421085 at the locus harboring FTO (encoding a 2-oxoglutarate–dependent demethylase) and rs2817419 at the locus harboring TFAP2B (encoding transcription factor AP-2(B)) associated with Age-AR, and rs10938397 near GNPDA2 (encoding adiposity regulating glucosamine-6-phosphate deaminase) locus associated with BMI-AR (Table 1 and fig. S5). Each lead SNP (rs1421085, rs2817419, and rs10938397) associated with Age-AR and BMI-AR explains approximately 0.2% of variance in the relevant early growth trait (see Methods).

**A new variant in LEPR/LEPROT associated with BMI-AP**

The BMI-AP–associated SNP rs9436303 (Fig. 1 and Table 1) at the locus harboring LEPR/LEPROT (encoding the leptin receptor and the leptin receptor overlapping transcript) is novel. This novel variant is robustly associated with BMI-AP after applying a conservative bias-reducing correction for winner’s curse and a multiple testing correction for six phenotypes ($\alpha' = 10^{-8}$; see Methods and table S3). The risk allele (G) of this variant increases both BMI-AP and adult plasma soluble leptin receptor levels ($P = 1.19 \times 10^{-9}$) (table S4) (11). The LEPR/LEPROT locus is in a chromosomal region, 1p31.3, that harbors another independent signal [rs11208659: minor allele frequency (MAF) = 0.06; distance = 82.6 kilo–base pairs; $R^2 = 0.01$] associated with early-onset obesity (12), but our SNP rs9436303 is associated with BMI-AP independently of this variant [linkage disequilibrium (LD) $R^2 = 0.01$ and see conditional analysis in table S6]. There was some effect heterogeneity between studies for this variant (fig. S6, A and D), but excluding the two studies with inflated estimates eliminated heterogeneity ($I^2 = 0$) in the stage 1 + 2 meta-analysis (fig. S6, A and D).

### Table 1. Summary statistics of the eight independent SNPs associated with PWV in infancy, BMI-AP in infancy, Age-AR, and BMI-AR in discovery (stage 1) and follow-up (stage 2) and in combined meta-analyses.

<table>
<thead>
<tr>
<th>Index SNP</th>
<th>Chromosome position*</th>
<th>In/near gene</th>
<th>Effect allele/other allele</th>
<th>Effect allele frequency</th>
<th>Effect size (SE)</th>
<th>Stage 1 (n = 7,215)</th>
<th>Stage 2 (n = 16,550)</th>
<th>Combined (n = 22,769)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWV (kg/month)$^\dagger$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2860323</td>
<td>chr2:614210</td>
<td>TMEM18</td>
<td>G/A</td>
<td>0.12</td>
<td>0.09 (0.02)</td>
<td>5.9 x 10^{-8}</td>
<td>0.02 (0.02)</td>
<td>4.7 x 10^{-7}</td>
</tr>
<tr>
<td>BMI-AP (kg/m2)$^\dagger$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9436303</td>
<td>chr1:65430991</td>
<td>LEPR/LEPROT</td>
<td>G/A</td>
<td>0.22</td>
<td>0.13 (0.02)</td>
<td>4.7 x 10^{-8}</td>
<td>0.05 (0.01)</td>
<td>6.7 x 10^{-7}</td>
</tr>
<tr>
<td>rs10515235</td>
<td>chr5:96323352</td>
<td>PCSK1</td>
<td>A/G</td>
<td>0.21</td>
<td>0.09 (0.02)</td>
<td>9.7 x 10^{-7}</td>
<td>0.03 (0.01)</td>
<td>1.5 x 10^{-6}</td>
</tr>
<tr>
<td>Age-AR (years)$^\dagger$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1421085</td>
<td>chr16:53767042</td>
<td>FTO</td>
<td>C/T</td>
<td>0.25</td>
<td>-0.10 (0.02)</td>
<td>6.1 x 10^{-8}</td>
<td>-0.13 (0.01)</td>
<td>7.1 x 10^{-7}</td>
</tr>
<tr>
<td>rs2956578</td>
<td>chr5:36497552</td>
<td>Intergenic region$^\ddagger$</td>
<td>G/A</td>
<td>0.31</td>
<td>0.11 (0.02)</td>
<td>6.7 x 10^{-8}</td>
<td>0.00 (0.01)</td>
<td>8.3 x 10^{-7}</td>
</tr>
<tr>
<td>rs2817419</td>
<td>chr6:50845193</td>
<td>TFAP2B</td>
<td>A/G</td>
<td>0.76</td>
<td>-0.10 (0.02)</td>
<td>2.9 x 10^{-6}</td>
<td>-0.07 (0.01)</td>
<td>1.8 x 10^{-6}</td>
</tr>
<tr>
<td>BMI-AP (kg/m2)$^\dagger$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10938397</td>
<td>chr4:45180510</td>
<td>GNPDA2</td>
<td>G/A</td>
<td>0.35</td>
<td>0.09 (0.02)</td>
<td>5.4 x 10^{-6}</td>
<td>0.05 (0.01)</td>
<td>3.1 x 10^{-4}</td>
</tr>
<tr>
<td>rs2055816</td>
<td>chr11:85406487</td>
<td>DLG2</td>
<td>C/T</td>
<td>0.25</td>
<td>-0.13 (0.02)</td>
<td>1.4 x 10^{-7}</td>
<td>-0.03 (0.02)</td>
<td>1.8 x 10^{-7}</td>
</tr>
</tbody>
</table>

*SNP positions are according to dbSNP build 147. $^\dagger$The effect size is the change in SDs per effect allele from linear regression, adjusted for child’s sex and principal components (PCs) assuming an additive genetic model. BMI-AP was additionally adjusted for gestational age (GA). PWV, BMI-AP, and BMI-AR were log-transformed because of skewness in their distribution. Original phenotype measurement units are denoted in parentheses. None of the loci for PHV passed the selection criteria for stage 2 follow-up. $^\ddagger$Intergenic region between RANBP3L and SLC1A3.
Fig. 1. Regional association and forest plot of the novel genome-wide significant locus associated with BMI-AP. Purple diamond indicates the most significantly associated SNP in stage 1 meta-analysis, and circles represent the other SNPs in the region, with coloring from blue to red corresponding to \( r^2 \) values from 0 to 1 with the index SNP. The SNP position refers to the National Center for Biotechnology Information (NCBI) build 36. Estimated recombination rates are from HapMap build 36. Forest plots from the meta-analysis for each of the identified loci are plotted abreact. Effect size (95% confidence interval (CI)) in each individual study, discovery, follow-up, and combined meta-analysis stages is presented from fixed-effects models (heterogeneity of the SNP rs9436303 in LEPR/LEPROT; see fig. S6). At this locus, there was heterogeneity between the studies in discovery \( (I^2 = 72.1\%, P = 0.01) \) and combined stage \( (I^2 = 59.3\%, P = 0.002) \) fixed-effects meta-analyses that was mainly due to LISA-D, EDEN, and the larger well-defined NFBC1966 study (fig. S6, A and D). Removing the studies that showed inflated results from meta-analyses did not change the point estimates (fig. S6, C, F, and G). Both fixed- and random-effects models gave very similar results (fig. S6, B and E).

C and F) without a substantial impact on effect sizes or significance levels. This SNP explains 0.3% of variance in BMI-AP (see Methods).

The SNP rs9436303 overlaps a regulatory region in a LEPR intron and is downstream from a processed transcript of LEPROT gene (table S7). LEPROT and LEPR overlap and share the same promoter but encode distinct transcripts with specific biological functions (13). The known biological function and molecular mechanism of the proteins encoded by the nearest genes in the four loci discovered are given in table S8. However, as with most GWAS-identified loci, the expression of these genes may not necessarily be influenced by the underlying causal variant(s) tagged by the GWAS SNP, so we sought further evidence that the BMI-AP–associated variants influence expression in the following section.

**Cis colocalization of GWAS and expression quantitative trait locus signals**

To identify GWAS and expression quantitative trait loci (eQTLs) signals that share the same causal variants, we performed Bayesian colocalization analyses (14) using our stage 1 GWAS meta-analysis summary statistics and eQTL data from 44 postmortem tissues generated by the Genotype-Tissue Expression (GTEx) consortium (see Methods) (15). The lead GWAS variants with high (>95%) posterior probability (PP) of colocalization were followed-up in five separate studies (see Methods) using cis-eQTL data from five ex vivo tissues and combined with genomic annotation data (tables S9 and S10). In these analyses, we found high PPs of colocalization with local causal variants (>95%) driving the expression of LEPR and LEPROT (Table 2 and fig. S7). The colocalization results for each gene are markedly tissue specific (Fig. 2 and fig. S8). In ex vivo samples, the LEPR/LEPROT variant was in high LD with the top eQTLs of LEPR and LEPROT genes in omental fat, subcutaneous fat, and whole blood (table S9). Direct lookup of LEPR/LEPROT variant in eQTL data indicated that the G allele of this variant that raised BMI-AP in our GWAS up-regulated the NM017526 transcript of LEPROT and down-regulated the AK023598 transcript from the same gene in adult tissues (table S10). This observation was consistent across two different eQTL studies and four tissues, suggesting the involvement of alternative splicing of a cassette exon. The LEPR/LEPROT variant overlapped DNA binding motifs of transcription factors and regulatory regions, as well as enhancer and promoter histone marks in multiple tissues (fig. S9). In Avon Longitudinal Study of Parents and Children (ALSPAC), the same LEPR/LEPROT variant was associated with higher DNA methylation levels of a LEPR intron measured in blood samples taken from mother and offspring. In particular, associations were found during mother’s pregnancy and in offspring’s adolescence, but not at offspring’s birth, at childhood, or in mother’s middle age (table S11) (16). This observation might be consistent with the regulation of a constitutively expressed transcript, which is also supported by evidence that lower LEPR intron DNA methylation levels were associated with higher serum leptin concentrations (17). Together, these results suggest that shared causal variants in these loci regulate BMI trajectory at AP, orchestrate changes in gene expression in different tissues, and modulate methylation of the nearest genes during mother’s pregnancy and at specific developmental stages of the offspring.

**Genetic determinants of adult BMI overlap with those determining AR but not AP**

In our study, Age-AR and BMI-AR have moderate to very strong genetic correlations with adult BMI and other adult adiposity-related
phenotypes, but BMI-AP does not (see Methods, Fig. 3, and table S12). Age-AR and BMI-AR had genetic correlations with multiple (more than four) adult complex phenotypes, including adult waist circumference ($r_g = -0.62$; BMI-AR $r_g = 0.48$) and adult body fat percentage (Age-AR $r_g = -0.49$; BMI-AR $r_g = 0.44$). Adult BMI and adult obesity had strong genetic correlations with BMI-AR ($r_g = 0.64$ and $r_g = 0.66$) and Age-AR ($r_g = -0.72$ and $r_g = -0.75$) but weak correlation with BMI-AP ($r_g = 0.29$ and $r_g = 0.33$). The traits with genetic and phenotypic correlations that were directionally consistent (note S1) are reported in table S13. Genetic correlations of Age-AP with other traits could not be quantified because of low mean $\chi^2$ of the GWAS summary statistics. In summary, genetic correlation analyses suggest that the genetic factors influencing adult BMI, body fat percentage, waist circumference, and obesity are also associated with BMI-AR and Age-AR, but their overlap with BMI-AP is either absent or weak.

**Genetic risk score for adult BMI is associated with Age-AR and BMI-AR but not with Age-AP and BMI-AP**

To gain further insight into the observed genetic correlations with adult BMI and to understand the developmental timing of the adult BMI-associated variants, we constructed an adult BMI genetic risk score (GRS) based on the 97 adult BMI SNPs identified by the Genetic Investigation of Anthropometric Traits (GIANT) consortium (18) (Fig. 4 and table S14) and applied it to the six early growth traits (see Methods). The adult BMI variants and the GRS were consistently and robustly associated with Age-AR ($h^2_{grs} = 0.035$, $P = 2.6 \times 10^{-48}$) and BMI-AR ($h^2_{grs} = 0.030$, $P = 1.7 \times 10^{-41}$) but not with other early

---

### Table 2. GWAS loci colocalized with eQTL in postmortem tissues from the GTEx data.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Nearest gene</th>
<th>Trait</th>
<th>GWAS SNP</th>
<th>GWAS SNP $P$ value</th>
<th>Tissue</th>
<th>eQTL SNP</th>
<th>eQTL $P$ value</th>
<th>eQTL gene</th>
<th>Top eQTL SNP ($R^2$)</th>
<th>Colocalization PP (%)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LEPR/LEPROT</td>
<td>BMI-AP</td>
<td>rs9436303</td>
<td>$8.3 \times 10^{-9}$</td>
<td>Thyroid</td>
<td>rs9436301</td>
<td>$7.9 \times 10^{-7}$</td>
<td>LEPROT</td>
<td>rs9436745 (0.78)</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Esophagus muscularis</td>
<td>rs1887285</td>
<td>$1.6 \times 10^{-6}$</td>
<td>LEPROT</td>
<td>rs9436745 (0.78)</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cell EBV-transformed lymphocytes</td>
<td>rs1887285</td>
<td>$1.2 \times 10^{-7}$</td>
<td>LEPROT</td>
<td>rs77948204 (0.22)</td>
<td>96</td>
</tr>
<tr>
<td>6</td>
<td>TFAP2B</td>
<td>Age-AR</td>
<td>rs2817419</td>
<td>$4.4 \times 10^{-11}$</td>
<td>Testis</td>
<td>rs2635727</td>
<td>$2.9 \times 10^{-7}$</td>
<td>TFAP2B</td>
<td>rs2635727 (0.91)</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sun-exposed skin lower leg</td>
<td>rs2635727</td>
<td>$4.2 \times 10^{-6}$</td>
<td>TFAP2B</td>
<td>rs2635727 (0.91)</td>
<td>98</td>
</tr>
</tbody>
</table>

*R^2* values between GWAS SNP and GTEx top eQTL SNP for each gene (eGene) are shown for reference. Only results with a ** posterior probability (PP) of a shared causal variant of >95% are reported.

---

**Fig. 2. Tissue-specific posterior probabilities (PPs) of colocalization for LEPR and LEPROT.** PP of eQTL and GWAS SNP sharing a causal variant regulating the gene expression levels of (A) LEPR and (B) LEPROT. Colocalization reported for GTEx eQTLs data in 34 tissues that express at least one of the genes. Bar plot color-coded according to the $-\log_{10} P$ value eQTL direct lookup in the corresponding GTEx tissue of the GWAS SNP. LEPR and LEPROT eQTLs colocalized with BMI-AP variant rs9436303.
growth traits (Fig. 4 and table S15). In the remaining four early growth traits, the GRS explained a negligible proportion of variance ($h^2_{grs} < 0.001$), and the adult BMI variants had inconsistent genetic effects (fig. S10 and table S15). In particular, the adult BMI variant effects on BMI-AP and PWV were highly heterogeneous ($P_{het} < 2 \times 10^{-4}$), with evidence of horizontal pleiotropy (MR-PRESSO; $P < 2 \times 10^{-4}$). This suggests that, in contrast with their effects on Age-AR and BMI-AR, the top loci associated with adult BMI do not have robust associations with the remaining four early growth traits. Thus, the underlying genetic determinants of adult BMI might differ from those influencing BMI-AP. Together, these data indicate that many GWAS variants associated with adult BMI have effects that begin in later childhood (4 to 6 years), as early as the Age-AR but not as early as AP (around 9 months).

Gene set analyses suggest little overlap between pathways and networks controlling AP and AR

To combine information on the effects of common variants in biological pathways and networks underlying early growth, we applied a gene set enrichment analysis [Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA)] (19) to the discovery stage GWAS results (see Methods). We identified enrichment of gene sets (tables S16 and S17) but did not find evidence for overlap of enriched pathways and networks among early growth traits. Age-AR–associated regions are involved in the insulin-like growth factor 1 (IGF-1) signaling pathway (FDR < 0.05). The IGF-1 signaling pathway has a well-established role both in growth and in the regulation of energy metabolism through the activation of phosphatidylinositol 3-kinase (PI3K)/AKT pathway via either the insulin or the IGF-1 receptors (20).

SNP heritability of Age-AR and BMI-AR is larger than BMI-AP

We estimated the chip SNP heritability (the proportion of variance explained by common SNPs) for the six early growth traits using LD score regression (LDSC) (see Methods). The heritability estimates for BMI-AR ($h^2_{snp} = 0.38$), Age-AR ($h^2_{snp} = 0.36$), PWV ($h^2_{snp} = 0.32$), and BMI-AP ($h^2_{snp} = 0.29$) were statistically significant ($P < 0.05$; Table 3). LDSC and SumHer (21) SNP heritability estimates (table S18) ranked these phenotypic heritabilities in a similar manner. The BMI-AP and BMI-AR estimates compared well with LDSC estimates for adult BMI ($h^2_{snp} = 0.27$) in a much larger sample of the UK Biobank ($N = 152,736$). Twin and family study heritability estimates for BMI-AP ($h^2 = 0.75$ to 0.78) (22, 23) and BMI-AR ($h^2 = 0.4$ to 0.6) (24, 25) were higher than the SNP heritability estimated here.
and later measures of BMI overlaps or differs, and contributes to the developmental timings at which the genetic makeup of early growth traits, determined by modeling growth trajectories using high-density variants at four independent loci associated with three early growth patterns so far. In the present study, we identified four well-established growth and BMI trajectories, and the ratios of the SNP heritability obtained from LDSC and offspring GWAS SNPs with P < 0.05 are plotted with a solid circle and labeled with the nearest gene name. The scatter plots of the other early growth phenotypes are given in fig. S10.

However, the ratio of the SNP heritability obtained from LDSC and the total heritability obtained from family and twin studies suggests that a considerable (39 to 95%; see Methods) proportion of BMI heritability can be attributed to common variants. As the LDSC heritability estimates of BMI-AP, BMI-AR, and adult BMI are comparable, the differences in the genetic etiology observed in our study cannot be trivially attributed to large disparities in the variance explained by genetic factors. Hence, together, these data suggest that distinct, heritable developmental processes control the BMI trajectory at AP and AR.

DISCUSSION
There are few reports of studies investigating the genetic bases of these well-established growth and BMI trajectories (26, 27), and to our knowledge, our study is the largest genome-wide meta-analyses of early growth traits so far. In the present study, we identified four variants at four independent loci associated with three early growth trajectories, determined by modeling growth trajectories using high-density longitudinal data for height and weight. Our study provides insights into the developmental timings at which the genetic makeup of early and later measures of BMI overlaps or differs, and contributes to understanding the mechanisms and molecular pathways of early growth patterns.

The three common variants at FTO, TFAP2B, and GNPD1A2, associated with timing of adiposity rebound and/or BMI-AR, are robustly associated with adult BMI and other adiposity traits. In contrast, the newly discovered variant at the LEPR/LEP1R locus associated with BMI-AP did not associate with other growth traits reported here, or in previous studies on childhood/adult BMI and obesity. This may indicate that genetic variants involved in adult BMI only start influencing BMI after AP and are robustly associated with child BMI by the time of AR. This is further corroborated by two additional lines of evidence provided by our study: (i) We observed strong genetic correlations of adult BMI, body fat percentage, and waist circumference with Age-AR and BMI-AR but not with Age-AP and BMI-AP, and (ii) the GRS constructed using adult BMI variants was robustly associated with Age-AR and BMI-AR but not with Age-AP and BMI-AP.

The difference in the genetic determinants of BMI-AP and BMI-AR and onward may be attributed to three factors: (i) BMI explains a relatively small propotion of body fat percentage (R^2 < 0.3) in infancy (0 months < age ≤ 7 months) (28) but increasingly larger proportions (0.36 < R^2 ≤ 0.8) in childhood (2 years < age < 18 years) (29, 30) and adulthood (R^2 ≈ 0.8; age, >18 years) (31); (ii) the genes involved in the regulation of BMI during infancy seem to differ from those acting in later childhood onward, which suggests distinct biological processes acting throughout these developmental stages; and (iii) sustained changes in the infant environment after weaning and onward may progressively unmask the effects of adult BMI variants. Consistent with this view, there is some evidence that infants’ and children’s environment modifies the effect of genetic factors. In particular, having been breastfed modifies the strength of association of the FTO variant with BMI (32) and with BMI growth trajectories (27). On the other hand, the adult risk alleles of the FTO and MC4R variants are not associated with increased infant BMI (26), but FTO’s strength of association with BMI progressively increases in later childhood (4 to 11 years) (24). Likewise, BMI heritability increases throughout childhood up to young adulthood (4 to 19 years) (22, 24, 25), as offspring BMI starts resembling adult BMI as an anthropometric marker of adiposity, and as the shared environment between adults and offspring progressively increases. Consistently, BMI heritability increased between AP and AR, and a considerable proportion of heritability was explained by common variants in our study. The increase in BMI heritability with age might be explained by correlations between genotype and environment. Small genetic differences are magnified as children progressively select, modify, and create environments correlated with their genetic propensities, which, in turn, unmask the effects of other genetic variants in a feedforward loop. These processes gradually may increase the phenotype variance explained by genetic factors and thereby increase BMI heritability. All in all, our study supports the accrual of shared genetic determinants between later childhood and adult BMI (5, 6), but not with infant BMI.

In our study, the IGF-1 pathway that links diet with growth was enriched for variants associated with Age-AR, but not Age-AP, in the MAGENTA analysis. Higher IGF-1 levels, via genetic and/or nutritional factors, can reduce growth hormone (GH) levels via a negative feedback (33). Subsequent lower circulating levels of GH can suppress lipolysis and contribute to fat accumulation (34), changing BMI trajectories and Age-AR, and, thereby, increasing risk of obesity and metabolic disorders. The regulation of the GH/IGF-1 pathway is further corroborated by the strong genetic correlations of BMI-AP, BMI-AR, and adult BMI with Age-AP and BMI-AR, and (ii) the GRS constructed using adult BMI variants is robustly associated with Age-AR and BMI-AR but not with Age-AP and BMI-AR.
axis is modulated by leptin and adiponectin levels, two hormones regulated by LEPR/LEPROT and TFAP2B genes, respectively (35). The variant at LEPR/LEPROT colocalized with causal variants regulating the expression of LEPR and LEPROT in different tissues. LEPROT and the LEPR genes share the same promoter but encode distinct transcripts (13). LEPROT is cotranscribed with the LEPR, and both are expressed in multiple tissues with different functionalities. LEPR is widely distributed in peripheral tissues, shows signaling capability, and is thought to transport leptin across the blood-brain barrier (25). Some LEPR isoforms may function in leptin clearance or buffering (soluble LEPR). In our eQTL data, the G allele that raises BMI-AP up-regulates the NM017526 transcript of LEPROT but down-regulates AK023598 transcript from the same gene in adult tissues. This observation was consistent across the different eQTL studies and tissues, suggesting that this variant may regulate the alternative splicing of a cassette exon in adult blood and subcutaneous and omental adipose tissue. In addition, the LEPR/LEPROT variant was associated with methylation levels in the LEPR intron during mother’s pregnancy and at specific developmental stages of the offspring. Together, this functional analysis suggests that distinct molecular mechanisms in different tissues are involved in the expression regulation of these genes at different developmental stages.

LEPROT and the LEPR downstream mechanisms involved on the regulation of BMI are likely to be developmental stage dependent. In humans, loss-of-function mutations in the LEPR markedly increase weight of infants after birth that persists through adulthood (25). Some LEPR isoforms may function in leptin clearance or buffering (soluble LEPR). In our eQTL data, the G allele that raises BMI-AP up-regulates the NM017526 transcript of LEPROT but down-regulates AK023598 transcript from the same gene in adult tissues. This observation was consistent across the different eQTL studies and tissues, suggesting that this variant may regulate the alternative splicing of a cassette exon in adult blood and subcutaneous and omental adipose tissue. In addition, the LEPR/LEPROT variant was associated with methylation levels in the LEPR intron during mother’s pregnancy and at specific developmental stages of the offspring. Together, this functional analysis suggests that distinct molecular mechanisms in different tissues are involved in the expression regulation of these genes at different developmental stages.

LEPROT and the LEPR downstream mechanisms involved on the regulation of BMI are likely to be developmental stage dependent. In humans, loss-of-function mutations in the LEPR markedly increase weight of infants after birth that persists through adulthood (25). Some LEPR isoforms may function in leptin clearance or buffering (soluble LEPR). In our eQTL data, the G allele that raises BMI-AP up-regulates the NM017526 transcript of LEPROT but down-regulates AK023598 transcript from the same gene in adult tissues. This observation was consistent across the different eQTL studies and tissues, suggesting that this variant may regulate the alternative splicing of a cassette exon in adult blood and subcutaneous and omental adipose tissue. In addition, the LEPR/LEPROT variant was associated with methylation levels in the LEPR intron during mother’s pregnancy and at specific developmental stages of the offspring. Together, this functional analysis suggests that distinct molecular mechanisms in different tissues are involved in the expression regulation of these genes at different developmental stages.

The variant at LEPR/LEPROT colocalized with causal variants regulating the expression of LEPR and LEPROT in different tissues. LEPROT and the LEPR genes share the same promoter but encode distinct transcripts (13). LEPROT is cotranscribed with the LEPR, and both are expressed in multiple tissues with different functionalities. LEPR is widely distributed in peripheral tissues, shows signaling capability, and is thought to transport leptin across the blood-brain barrier (25). Some LEPR isoforms may function in leptin clearance or buffering (soluble LEPR). In our eQTL data, the G allele that raises BMI-AP up-regulates the NM017526 transcript of LEPROT but down-regulates AK023598 transcript from the same gene in adult tissues. This observation was consistent across the different eQTL studies and tissues, suggesting that this variant may regulate the alternative splicing of a cassette exon in adult blood and subcutaneous and omental adipose tissue. In addition, the LEPR/LEPROT variant was associated with methylation levels in the LEPR intron during mother’s pregnancy and at specific developmental stages of the offspring. Together, this functional analysis suggests that distinct molecular mechanisms in different tissues are involved in the expression regulation of these genes at different developmental stages.

In a model of BMI development consisting of the superimposition of genetic and other environmental factors also warrants additional research (27).

In conclusion, this longitudinal GWAS study, based on derived traits from growth modeling, has uncovered a completely new variant in LEPR/LEPROT locus that specifically associates with BMI at the peak of adiposity in infancy. The present study identified two BMI developmental stages in infancy and later childhood with distinct genetic makeup. Our results support the notion that genetic determinants of adult BMI progressively start acting in later childhood but not necessarily before the AP in infancy (5, 6). This finding may corroborate a model of BMI development consisting of the superimposition of two biological processes with distinct genetic drivers (Fig. 5), which, in turn, suggests that interventions in childhood aiming to modify BMI and achieve long-lasting reductions in the risk of adult obesity need to take into account the developmental stage. We believe that the identification of genetic factors underpinning the BMI trajectory is a fundamental step toward understanding the etiology of obesity and may inform strategies to prevent and treat it.

Table 3. SNP heritability of the early growth traits. SNP heritability estimated with LD score using all common SNPs (MAF > 0.01) in stage 1 GWAS meta-analysis.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Estimated heritability</th>
<th>SE</th>
<th>95% CI</th>
<th>Mean χ²</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI-AP</td>
<td>0.29</td>
<td>0.13</td>
<td>0.46</td>
<td>1.03</td>
<td>4.7 × 10⁻⁴</td>
</tr>
<tr>
<td>BMI-AR</td>
<td>0.38</td>
<td>0.22</td>
<td>0.53</td>
<td>1.013</td>
<td>2.7 × 10⁻⁶</td>
</tr>
<tr>
<td>Age-AP</td>
<td>−0.03</td>
<td>−0.18</td>
<td>0.13</td>
<td>1.001</td>
<td>7.4 × 10⁻¹</td>
</tr>
<tr>
<td>Age-AR</td>
<td>0.36</td>
<td>0.20</td>
<td>0.52</td>
<td>1.007</td>
<td>1.1 × 10⁻⁵</td>
</tr>
<tr>
<td>PHV</td>
<td>0.11</td>
<td>−0.03</td>
<td>0.25</td>
<td>1.006</td>
<td>1.3 × 10⁻¹</td>
</tr>
<tr>
<td>PWV</td>
<td>0.32</td>
<td>0.18</td>
<td>0.45</td>
<td>1.011</td>
<td>2.5 × 10⁻⁶</td>
</tr>
</tbody>
</table>
Derivation of individual was used to estimate the growth parameter, PHV (in centimeters). Maximum-likelihood method for best fitting curves for each infant was included in the Reed1 model fitting. Parameters from the fitted curves are described in detail in a previous publication. Early growth traits were derived from sex-specific nonlinear random-effect model as described previously. Parametric Reed1 growth model was fitted in publication. The methods for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 24 months of age were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The following model was used to calculate the Age-AP and BMI-AP, and the analysis was restricted to singletons with BMI measures from 2 weeks to 18 months of age. The model is as follows:

\[
\log(BMI) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Age}^3 + \beta_4 \text{Sex} + u_0 + u_1 (\text{Age}) + \epsilon
\]

where BMI is expressed in kilograms per square meter and age in years. \(\beta_0, \beta_1, \beta_2, \beta_3, \) and \(\beta_4\) are the fixed-effects terms, \(u_0\) and \(u_1\) are the individual-level random effects, and \(\epsilon\) is the residual error. The Age-AP was calculated from the model as the age at maximum BMI between 0.25 and 1.25 years according to preliminary research. The following model was used to measure the age and BMI-AR in childhood:

\[
\log(BMI) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Age}^3 + \beta_4 \text{Sex} + u_0 + u_1 (\text{Age}) + \epsilon
\]

where BMI is expressed in kilograms per square meter and age in years. \(\beta_0, \beta_1, \beta_2, \beta_3, \) and \(\beta_4\) are the fixed-effects terms, \(u_0\) and \(u_1\) are the individual-level random effects, and \(\epsilon\) is the residual error. The Age-AR was calculated as the age at minimum BMI between 2.5 and 8.5 years according to preliminary research. The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month). The methods used for growth modeling and derivation of growth parameters from the fitted curves are described in detail in a previous publication. Parametric Reed1 growth model was fitted in sex-stratified nonlinear random-effect model as described previously. Term-born singletons (defined as ≥37 completed weeks of gestation) with at least three height or weight measurements from birth to 13 years (fig. S1) were included in the Reed1 model fitting. Maximum-likelihood method for best fitting curves for each individual was used to estimate the growth parameter, PHV (in centimeters per month), and PWV (in kilograms per month).
Stage 1 genome-wide association analyses and meta-analyses

According to the availability of dense enough data for growth modeling, a total of up to 7215, 6222, 6219, and 6051 children were used to analyze PHV/PWV, Age-AP, BMI-AP, and Age-AR/BMI-AR, respectively (fig. S2). We only included children who were born between 37 and 41 completed weeks of gestation (i.e., term born) from singleton pregnancies and children who had more than three growth measurements available within the age range in question. Gestational age (GA) was either defined from the date of the last menstrual period or ultrasound scans depending on the study. All six early growth traits except for Age-AP and Age-AR were naturally log-transformed to reduce skewness, and all traits were converted to z-scores before association testing to facilitate the comparison of results across the studies. We tested the directly genotyped and imputed variants for association with each of the six early growth traits in a linear regression model, assuming an additive genetic effect. The regression models were adjusted for sex and principal components (PCs) derived from the genome-wide data to adjust for potential population substructure (the necessary number of PCs included varied by study). GA is a marker of multiple factors influencing pregnancy that may influence the child growth trajectory. Regression of all phenotypes on GA adjusting for sex produced significant associations, apart from BMI-AP and Age-AR, which showed significant associations with sex only. On the basis of this observation, we adjusted all GWAS analyses for GA and sex apart from BMI-AP and Age-AR, which were adjusted for sex only. The risk of introducing collider bias was dismissed because gestational effects occur before birth, and the loci found did not overlap with GA signals in the GWAS catalog or PhenoScanner. The genome-wide association analyses (i.e., stage 1) were performed using either SNPTEST or MACH2QTL in each cohort, and data exchange facilities were provided by the AIMS server (40). All stage 1 study beta estimates and their SEs were meta-analyzed using the inverse-variance fixed-effects method in the METAL software (41). SNPs with poor imputation quality (e.g., \( R^2 < 0.3 \) for MACH and “proper info” score < 0.4 for IMPUTE) and/or an HWE \( P \) value cutoff as given in table S1.

Selection of SNPs for stage 2 follow-up

All loci reaching \( P < 1 \times 10^{-7} \) from stage 1 GWAS of each early growth trait were selected for follow-up in stage 2. These included the two SNPs associated with Age-AR in the FTO locus (rs1421085) and in the intergenic region between RANBP3L and SLC1A3 (rs2956578), and the SNP associated with BMI-AP in LEPR/LEPROT (rs9436303). Four further SNPs (one SNP associated with BMI-AP near PCSK1 (rs10515235), one SNP associated with Age-AR in TFAP2B (rs2817419), and two SNPs associated with BMI-AR near GNPD2 (rs10938397) and in DLG2 (rs2055816)) were selected for follow-up on the basis of showing an association with an early growth trait at \( P < 1 \times 10^{-5} \) and being in/near genes with established links to adiposity and metabolic phenotypes except for DLG2, a possible candidate gene involved in glucose metabolism (43). In addition, one locus with a plausible association (\( P = 5.91 \times 10^{-7} \)) with PWV, near TMEM18 (rs2860323), was also selected for follow-up based on previous reports showing an association with severe early-onset obesity (12) and its association with BMI in adulthood (44) and childhood (6) (table S3). No loci for PHV or Age-AP passed the \( P \) value threshold or other selection criteria used for follow-up. Table S3 shows the SNP selection criteria and proxies used in more detail.

Stage 2 follow-up of lead SNPs

For follow-up of lead signals selected from stage 1, we used data from up to 16,550 children of European descent from 12 additional population-based studies (up to 11 studies for each early growth trait), namely, the ALSPAC (United Kingdom), Cambridge Baby Growth Study (United Kingdom), Children’s Hospital of Philadelphia (United States), Copenhagen Prospective Study on Children (Denmark), Danish National Birth Cohort (Denmark), Étude des Déterminants pré- et postnatals du développement et de la santé de l’ENfant (EDEN; France), The Exeter Family Study of Childhood Health (United Kingdom), INFancia y Medio Ambiente Project (Spain), Lifestyle-Immune System–Allergy Study [LISA (R, Germany)], Northern Finland Birth Cohort Study 1986 (Finland), The Physical Activity and Nutrition in Children (Finland), and Southampton Women’s Survey (United Kingdom). We used de novo SNP genotyped or imputed data for the eight SNPs (or proxies of \( r^2 > 0.8 \)) selected from stage 1 and tested their association in a total of 5367, 16,550, 12,256, and 12,192 children of European ancestry with PWV, BMI-AP, Age-AR, and BMI-AR, respectively (Fig. 2). Direct genotyping was performed in some follow-up studies by KBiosciences Ltd. (Hoddesdon, United Kingdom) using their own novel system of fluorescence-based competitive allele-specific polymerase chain reaction (KASPar). The call rates for all genotyped SNPs were >95%. Study characteristics, genotyping platform, imputation and association test software used, as well as sample and genotyping and imputation quality control steps in each stage 1 study are given in table S2. We used the same methods as in stage 1 for sample selection, genotyping quality control, association testing, and meta-analysis.

Combined analysis of stage 1 and stage 2 samples

All stage 1 and 2 results were meta-analyzed using the inverse-variance fixed-effects method in either METAL (41) or R (version 3.2.0; www.r-project.org/). In these combined analyses, loci reaching \( P < 5 \times 10^{-8} \) were considered as genome-wide significant, and loci reaching \( P < 5 \times 10^{-6} \) were considered as a suggestive association. Heterogeneity between studies was tested by Cochran’s Q tests, and the proportion of variance due to heterogeneity was assessed using \( I^2 \) index for each individual SNP at each stage.
Estimation of genetic variance explained
The variance explained ($h^2$) by each SNP was calculated using the risk allele frequency ($f$) and beta ($\beta$) from the meta-analyses using the formula $h^2 = \beta^2 (1 - f) 2f$.

Stage 1 (discovery stage) GWAS FDR
We applied Efron’s method (45) to the $P$ values of all common SNPs (MAF > 1%) in stage 1 GWAS to estimate the FDR of the prioritized SNPs.

Bias-reducing correction for winner’s curse
We applied the method of Zhong and Prentice (46) to the eight signals prioritized in our discovery GWAS and reported bias-reduced estimates for the following discovery cutoffs $P < 10^{-5}$, $P < 10^{-6}$, and $P < 10^{-7}$.

GWAS multiple testing correction for six phenotypes
The standard joint meta-analysis genome-wide significance cutoff ($\alpha = 5 \times 10^{-8}$) was corrected using Bonferroni method for the number of independent tests. First, we applied the method of Li and Ji (47) to the phenotypic correlation matrix of the NFBC1966 ($N > 2500$) to estimate the number of “independent” phenotypes ($m = 5$). The corrected cutoff was estimated with Bonferroni method applied to five independent tests ($\alpha' = \alpha / 5 = 10^{-8}$).

Analysis of the phenotypic effects of the lead GWAS SNPs in published genetic studies
The phenotypic implication of the lead GWAS SNP on 778 phenotypes was obtained from the Gene ATLAS (9) PheWAS conducted on 452,264 white British individuals from UK Biobank available at http://geneatlas.roslin.ed.ac.uk. In addition, we looked up the lead GWAS SNP or a proxy on published large-scale GWAS datasets using PhenoScanner (10) available at www.phenoscanner.medschl.cam.ac.uk/phenoscanner. Only proxy SNPs from the 1000 genomes panel in high LD ($R^2 > 0.8$) with the lead GWAS SNP were considered. We then searched for all SNPs with phenotypic associations in the same chromosomal region of our lead GWAS signals using the GWAS catalog (48) available at www.ebi.ac.uk/gwas/. We reported associations obtained from these analyses with significance cutoff $P < 5 \times 10^{-8}$. Last, we searched for non-GWAS genetic studies, i.e., family, pedigree, and clinical studies with PubMed queries available at www.ncbi.nlm.nih.gov/pubmed/.

Conditional analyses
We conducted conditional analysis of the lead and proxy SNP expected dosages using a linear regression model adjusted for sex and GA in 3459 children from the NFBC1966 study. We considered two models to assess the effect of both SNPs. First, the early growth trait is regressed on the lead SNP adjusting for the study covariates sex and GA. Second, the proxy SNP is added to the previous model. The lead and proxy SNP effects are considered independent if the effect size estimate of the lead SNP in model 2 did not vary more than 20% of the effect estimate of model 1, and the corresponding $P$ value reached a nominal significance ($\alpha = 0.05$).

Variance effect prediction analysis
We obtained information about the putative effect of the lead GWAS SNPs using VEP tool (49) available at www.ensembl.org/Homo_sapiens/Tools/VEP. The analysis included pathogenicity, splicing, and conservation predictions as well as regulatory annotations.

Overlap of the genetic makeup of early growth traits with adult and childhood phenotypes
To gain insights into the potential overlap in the genetic makeup of early growth traits with adult and childhood phenotypes, we searched databases and the literature for the phenotypic implications of our four GWAS SNPs. First, we retrieved from the Gene Atlas (9) PheWAS in the UK Biobank data all phenotypic associations ($P < 5 \times 10^{-8}$) with our four GWAS SNPs (table S4). Second, we retrieved from the PhenoScanner (10) database all SNPs in the literature with phenotypic associations ($P < 5 \times 10^{-8}$) and in high LD ($R^2 > 0.8$) with our four GWAS lead variants (table S5). Third, we systematically searched in the GWAS catalog (48) database all SNPs with phenotypic associations ($P < 5 \times 10^{-8}$) in the chromosomal regions of our four GWAS lead variants.

Bayesian colocalization
Colocalization analyses were performed using our stage 1 GWAS results with multivariate eQTL results from GTEx data (www.gtexportal.org/home/datasets) (15). For each GTEx tissue (n = 48 tissues), we first identified all genes with significant cis-eQTLs at <5% of FDR. For each such gene, we retrieved the GWAS summary statistics for each of the three traits (BMI-AP, BMI-AR, and Age-AR), for all SNPs in common between the GWAS and the eQTL data [typically everything within 1 megabase (Mb) of the gene transcription start site]. If the GWAS locus contained one or more eQTL variant at $P < 10^{-5}$, then we implemented the computational procedure outlined in the coloc package in R (https://github.com/chr1swallace/coloc/blob/master/R/coloc-package.R) (50) with default parameters and using the MAFs of European ancestry individuals from 1000 Genomes study.

Expression quantitative trait locus
We searched for cis-eQTLs in liver, skin, whole blood, subcutaneous fat, and omental fat ex vivo tissue made available by the MuTHER (51), KORA (52), DeCode (53), Lee Kaplan (54), and BIOS (55) studies. The association analyses were performed with the GWAS lead SNP following the procedure described previously (56). The analysis of eQTLs was limited to genes in cis within a ±1-Mb window of the lead SNP. For each GWAS lead SNP, we separately reported the top eQTL in the locus and the coincident cis-eQTLs with significance cutoffs of $P < 1 \times 10^{-5}$ and an FDR of <5%, respectively. For studies where gene expression was measured using a microarray technology, the microarray probes were annotated with information accessed on ProbeDB (available at www.ncbi.nlm.nih.gov/probe/). If the probe IDs were not available in the ProbeDB, then we aligned the probe sequence to HG38 with blast algorithm available at https://blast.ncbi.nlm.nih.gov/Blast.cgi and then annotated the transcripts overlapping the genomic coordinates using consensual information in GenBank, RefSeq, ENCODE (Encyclopedia of DNA Elements), and UCSC (University of California, Santa Cruz) databases.

Methylation quantitative trait locus
We searched for cis-methylation QTL in blood at five different life stages using mQTLdb (www.mqtldb.org/). Methylation QTL data were generated as previously described (16) using ALSPAC study data. Only GWAS SNPs that colocalized with eQTL data were looked up.

Genetic correlations using LD score regression analyses
We used the LD hub (57) available at http://ldsc.broadinstitute.org to quantify the genetic correlation between each of the six early
growth traits and a selection of 49 disease/trait of interest from 33 GWAS studies in the following precompiled categories: education, anthropometric traits, lipids, glycemic traits, bone mineral density, neurological/psychiatric diseases, and other traits (including adiponectin, coronary artery disease, type 2 diabetes, and menarche). We carried out the LD score regression analyses for blood pressure traits using the Python scripts provided on the developer’s website at https://github.com/bulik/lsc. Before running the LD score regression analyses, each summary statistics file was reformatted using the munge_sumstats.py Python script, which filtered the SNPs to HapMap 3 SNPs as recommended on the developer’s website to minimize any bias from poor imputation quality. SNPs were also excluded if an MAF of <0.01, ambiguous strand, duplicate rsID, and reported sample size are less than 60% of the total available. If the sample size for each SNP was available, then we used the –N-col to specify the relevant sample size column in the GWAS summary statistics file, and when no sample size column was available, we used the maximum sample size reported in the GWAS meta-analysis. After the GWAS summary statistics files were reformatted, we then used the ldsc.py Python script to run the LD score regression analyses between each of the six early growth traits and systolic blood pressure and diastolic blood pressure. The precompiled European LD scores calculated from 1000 Genomes data available on the developer’s website were used for LD score regression.

Adult BMI GRS
We calculated a weighted GRSs of adult BMI with the 97 SNPs associated with BMI at genome-wide levels of significance in the GIANT consortium (18) using the R package gtx and following the procedure described in (58). Briefly, a risk score estimating the pleiotropic effect of adult BMI variants on each early growth trait was inferred from summary statistics obtained from the stage 1 GWAS meta-analyses. Risk score models with evidence of heterogeneity (Phet < 0.05, only BMI-AP) were refitted using a downwards elimination of SNPs with largest effect size until the model is not heterogeneous (Phet > 0.05). In addition, we estimated evidence of horizontal pleiotropy between adult BMI and each early growth trait with the package MR-PRESSO (59).

Pathway enrichment analysis
To explore the pathways associated with early growth traits, we applied MAGENTA (version 2) (19) to the stage 1 GWAS results. Briefly, each gene in the genome was mapped to a single SNP with the lowest P value within a 110-kb upstream or 40-kb downstream window of the gene. The corresponding P value, representing each gene, was corrected for confounding factors such as gene size, LD patterns, SNP density, and other genetic factors. The adjusted P values were ranked, and the observed number of genes in a given pathway above a specified P value threshold (75th and 95th percentiles used) was calculated. This number was compared with that from repeating the process based on 10,000 randomly permuted pathways of identical size. In doing so, an empirical gene set enrichment association (GSEA) P value for each pathway was computed. In our study, individual pathways with an FDR of <0.05 and nominal GSEA P < 0.05 were deemed significant, and, unless otherwise stated, results for the 95th percentile cutoff analysis were reported.

SNP heritability
We estimated the SNP heritability, the proportion of variance explained by common SNPs (MAF > 1%), with LD score as implemented in LD hub and using our six stage 1 GWAS meta-analyses on early growth traits (see the next paragraph for detailed information on postprocessing of GWAS data for LD score analysis). LD score regression estimates were obtained using a linear model with intercept, which aims at correcting for systematic confounders in GWAS summary statistics such as population stratification. In addition, we provided SumHer (21) SNP heritability estimates due to the current development on the (mainly downward) bias of LD score regression estimates. SumHer estimates were obtained using a regression model including an intercept. The estimates of heritability explained by common SNPs is the ratio of the SNP heritability obtained from LD score regression and the overall heritability obtained from family and twin studies.

Post hoc power analysis
We conducted a post hoc power analysis to determine the effect size in SD units we are powered to detect (power, 80%). The following experimental setup is considered to parameterize the null hypothesis: stage 1 meta-analysis sample size (n = 6222), the smallest minimum allele frequency observed among the four lead GWAS SNPs (MAF = 0.22; most conservative), imputation quality R² = 0.8, significance level P < 5 × 10⁻⁸, and genotypes assumed to be in HWE. Analysis was conducted as previously described (60). Briefly, the noncentrality parameter (NCP) gives the expected value of the test statistic under the null hypothesis parameterized above. The power to detect an effect size b > NCP is the probability of obtaining an effect larger or equal to NCP under the alternative hypothesis parameterized by the normal distribution with mean b and SD set equal to the SE of NCP.

SUPPLEMENTARY MATERIALS
Supplemental material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/9/eaaw3095/DC1
Table S1. Study characteristics, exclusions, genotyping, quality control, and imputation of stage 1 studies.
Table S2. Study characteristics, exclusions, genotyping, and quality control in stage 2 studies.
Table S3. The SNP selection criteria used for selecting loci from stage 1 GWAS meta-analysis data and proxies used for follow-up in stage 2.
Table S4. The association of the four genome-wide significant SNPs or proxies in high LD (R² > 0.8) with other phenotypes in published GWASs retrieved from PhenoScanner database.
Table S5. The association of the four genome-wide significant SNPs with other phenotypes in the Gene Atlas PheWAS on the UK Biobank data.
Table S6. Conditional analysis in the NFBC1966 data (N = 2585) of the BMI–AP association with the lead GWAS SNP rs9436303 adjusting for the early-onset obesity SNP rs11208659.
Table S7. Variant effect prediction of the four genome-wide significant SNPs.
Table S8. Biological and functional mechanisms of the nearest genes to the four genome-wide significant SNPs.
Table S9. Top eQTLs (FDR < 1%) in five ex vivo tissues in high LD (R² > 0.8) with the lead GWAS SNPs in LEPR/LEPROT locus.
Table S10. Direct lookup on eQTL (P < 0.001) data in five ex vivo tissues of the lead GWAS SNPs in LEPR/LEPROT locus.
Table S11. Direct lookup of the lead GWAS SNPs in LEPR/LEPROT and TFFAP2B locus on methylation QTL (FDR < 0.1%) in blood drawn at five different life stages: mother's pregnancy (~29.2 years, SD = 4.4 years) and middle age (~47.5 years, SD = 4.5 years), and offspring's birth (0 years), childhood (~7.5 years, SD = 0.15 years), and adolescence (~17.1 years, SD = 1.0 years).
Table S12. Cross-trait genetic correlations between five early growth traits and 80 other GWAS phenotypes from LD score regression analyses.
Table S13. The directional consistency between phenotypic and genetic correlations for the same trait.
Table S14. Lookup of the GIANT consortium BMI-associated SNPs on the stage 1 GWAS meta-analyses of the six early growth traits.
Table S15. The GRS of adult BMI using SNP weights from the GIANT consortium applied to the stage 1 GWAS meta-analyses.
childhood and adult traits. LEPR, TFAP2B, and PPARG have been associated with early growth traits that have been previously linked with adult BMI.

2. LEPR, PPARG, and PPARG2 associate with BMI at age 18 years if their mother smoked during pregnancy.

3. Genome-wide SNP and GTEx cis-eQTL data used in the colocalization analysis of the early growth–associated loci.

5. Tissue-specific PPs of colocalization of TFAP2B.


10. Adult BMI QR analysis of early growth traits.

Note 1. Literature search for epidemiological associations between early growth traits and childhood and adult traits.

Note 2. Cohort description (see also tables S1 and S2 for genotyping details and figs. S1 to S3).

Note 3. Funding and acknowledgments by the study.

References (61–104)

REFERENCES AND NOTES


13 of 17

Downloaded from http://advances.sciencemag.org/ on September 6, 2019
 contributions to weight, height, and BMI from birth to 19 years of age: an international replication study of over 12,000 twin pairs. PLOS ONE 7, e0103123 (2012).


Downloaded from advances.sciencemag.org on September 6, 2019


Acknowledgments: This publication is the work of the authors, and M.-R.J., M.I.M., S.F.A.G., K.K.O., V.W.V.J., and P.O. will serve as guarantors for the contents. We thank SpiroMeta Consortium and deCODE genetics for providing summary statistics of the lung function GWAS and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.

Funding: All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. There is no bearing with other listed disclosures.

All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.

Funding: All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.

Funding: All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.

Funding: All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.

Funding: All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.

Funding: All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.

Funding: All the authors acknowledge the following sponsors for support: UK Medical Research Council and Wellcome Trust (grant reference 102215/2/13/2); NIH grant RO1 HD065646; Danish National Research Foundation; NIH Genes, Environment and Health Initiative (GEI U01HG004423); NIH GEI (U01HG004438); Lundbeck Foundation (R19-A2059); Danish Medical Research Council (09-065592); French Ministry of Research; INSERM; South West NHS Research and Development; Exeter NHS Research and Development; Nethrlands Organization for Health Research and Development (VIDI 016. 136. 361); European Union’s Horizon 2020 research and innovation programme under grant agreement nos. 633595 (DynaHEALTH), 733206 (LIFECYCLE), and 721567 (CAPICE); European Research Council (ERC) Horizon 2020 research and innovation programme under grant agreement no. 633595; National Public Health Institute, Helsinki, Finland; World Cancer Research Fund (WCRF UK) and the eQTL study respectively. Cohort-specific acknowledgments are given in note S3.
GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI


Sci Adv 5 (9), eaaw3095.
DOI: 10.1126/sciadv.aaw3095