Genome-wide association study reveals a dynamic role of common genetic variation in infant and early childhood growth

Øyvind Helgeland[1,2], Marc Vaudel[1], Petur B. Juliusson[3,4,5], Oddgeir Lingaas Holmen[6,7], Julius Juodakis[8], Jonas Bacelis[8,9], Bo Jacobsson[2,8], Haakon Lindekleiv[10], Kristian Hveem[6,11], Rolv Terje Lie[1,12], Gun Peggy Knudsen[2], Camilla Stoltenberg[12,13], Per Magnus[14,15], Jørn V. Sagen[1,3,16], Anders Molven[1,17,18], Stefan Johansson[1,19,20] and Pål R. Njølstad[1,4,20]**

[1]KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway.

[2]Department of Genetics and Bioinformatics, Health Data and Digitalization, Norwegian Institute of Public Health, Oslo, Norway.

[3]Department of Clinical Science, University of Bergen, Bergen, Norway.

[4]Department of Pediatrics and Adolescents, Haukeland University Hospital, Bergen, Norway.

[5]Department of Health Registries, Norwegian Institute of Public Health, Bergen, Norway.

[6]KG Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

[7]HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim.

[8]Department of Gynecology and Obstetrics, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

[9]Department of Gynecology and Obstetrics, Sahlgrenska University Hospital, Gothenburg, Sweden

[10]University of Tromsø, Tromsø, Norway.

[11]HUNT Research Center, Levanger, Norway.

[12]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway.

[13]Norwegian Institute of Public Health, Oslo, Norway.

[14]Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway.

[15]Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway.

[16]Hormone Laboratory, Haukeland University Hospital, Bergen, Norway.

[17]Department of Clinical Medicine, University of Bergen, Bergen, Norway.

[18]Department of Pathology, Haukeland University Hospital, Bergen, Norway.

[19]Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway.

[20]Joint senior authors: S. Johansson, P.R. Njølstad.

1

[*]Address correspondence to: Professor Pål Rasmus Njølstad, M.D., Ph.D., K.G. Jebsen Center for Diabetes Research, Hospital for Children and Adolescents, Haukeland University Hospital, Haukelandsbakken 15, N-5021 Bergen, Norway (pal.njolstad@uib.no); or Professor Stefan Johansson, Ph.D., K.G. Jebsen Center for Diabetes Research, Department of Medical Genetics, Haukeland University Hospital, Haukelandsbakken 15, N-5021 Bergen, Norway (stefan.johansson@uib.no)

Abstract, 234 words; Text, 1,787 words; Figures, 4; Tables, 1; Supplementary material

Infant and childhood growth are dynamic processes characterized by drastic changes in fat mass and body mass index (BMI) at distinct developmental stages. To elucidate how genetic variation influences these processes, we performed the first genome-wide association study (GWAS) of BMI measurements at 12 time points from birth to eight years of age (9,286 children, 74,105 measurements) in the Norwegian Mother and Child Cohort Study (MoBa) with replication in 5,235 children (41,502 measurements). We identified five loci associated with BMI at distinct developmental stages with different patterns of association. Notably, we identified a novel transient effect in the leptin receptor (*LEPR*) locus, with no effect at birth, increasing effect on BMI in infancy, peaking at 6-12 months (rs2767486, $P_{6m} = 2.0 \times 10^{-21}$, $\beta_{6m} = 0.16$) and little effect after age five. A similar transient effect was found near the leptin gene (*LEP*), peaking at 1.5 years of age (rs10487505, $P_{1.5y} = 1.3 \times 10^{-8}$, $\beta_{1.5y} = 0.079$). Both signals are protein quantitative trait loci (pQTLs) for soluble *LEPR* and *LEP* in plasma in adults and independent from signals associated with other adult traits mapped to the respective genes, suggesting novel key roles of common variation in the leptin signaling pathway for healthy infant growth. Hence, our longitudinal analysis uncovers a complex and dynamic influence of common variation on BMI during infant and early childhood growth, dominated by the *LEP-LEPR* axis in infancy.

BMI patterns in infancy and childhood follow well-characterized trajectories: a rapid increase soon after birth until approximately nine months, the adiposity peak, followed by a gradual decline until approximately 4-6 years of age and then the adiposity rebound, when BMI starts to increase again until the end of puberty[1]. Recently, a study revealed that the most powerful predictor of obesity in adolescence is an increase in BMI between two and six years of age[2], but the underlying cause for this remains unknown. To explore how common genetic variation influences these processes, we performed the first genome-wide association study (GWAS) of

3

early growth in the population-based birth cohort, the Norwegian Mother and Child Cohort Study (MoBa)[3] (Supplementary Table 1). A total of 17,474 children in MoBa were genotyped in discovery and replication combined. The children's BMI was measured at birth, 6 weeks, 3, 6, 8 months, and 1, 1.5, 2, 3, 5, 7, and 8 years of age (Fig. 1a). We performed genotype quality control (QC), imputation using the Haplotype Reference Consortium (HRC), and phenotype QC, leaving 9,286 and 5,235 samples for the discovery and replication cohorts, respectively, all of Norwegian ancestry.

We conducted separate linear regression analyses of standardized BMI for each time point using an additive genetic model. The lead SNPs at independent loci reaching $P < 10^{-7}$ at one or more time points in the discovery sample were taken forward for replication (Table 1). This revealed a highly dynamic pattern of association during early growth. SNPs in five independent loci reached genome-wide significance, presenting peak association at different time points: (1) a novel, intronic SNP rs2767486 in the *LEPR* locus peaking at six months; (2) an intronic SNP, rs13035244, near *ADCY3* peaking at one year; (3) an intronic SNP rs6842303 near *LCORL* peaking at 1.5 year; (4) an intergenic SNP rs10487505 near *LEP* peaking at 1.5 year; and (5) an intronic SNP rs9922708 near *FTO* peaking at seven years (Fig. 1b, c, and Supplementary Fig. 1, 2).

The strongest association with BMI was found for rs2767486 at six months ($P_{6m} = 2.0 \times 10^{-21}$, $\beta_{6m} = 0.16$) in the *LEPR/LEPROT* locus. The locus strongly associated with BMI from three months of age, with effects peaking at 6-12 months, and waning from age three with little effect at eight years (Fig. 1c, 2, and Supplementary Fig. 3a). We found no evidence of association at birth for rs2767486 or nearby markers in our data or in recent large publicly available GWASs of birth weight[4] and adult BMI[5,6]. Thus, this locus most likely affects BMI development primarily during infancy. Conditioning on rs2767486 revealed a putative additional signal in the *LEPR*

4

locus, rs17127815 (P_{6m} = 7.5 × 10^{-5} after conditioning), that mirrored the association pattern of the main signal (Supplementary Fig. 3b).

LEPR encodes the leptin receptor, which functions as a receptor for the adipose cell-specific hormone leptin. High leptin levels suppress hunger by interacting with the long form of the leptin receptor (OB-RL) in the hypothalamus[7]. The soluble form of leptin receptor (sOB-R), which is produced through ectodomain shedding of OB-RL in peripheral tissues, can bind leptin in circulation, and thereby reduce its effect on the central nervous system[8]. The *LEPR* locus has previously been implicated in monogenic morbid obesity[9,10], severe childhood obesity[11], age of menarche[12], age of voice breaking[13], levels of fibrinogen[14] and C-reactive protein[15], several blood cell count traits[16,17], and plasma sOB-R levels[17,18]. To test whether any of the established variants for these traits explains the observed association with BMI in infancy, we repeated the analysis conditioning on the top SNPs reported in these studies. The association with infant BMI remained unaffected by conditioning on these SNPs, except for rs2767485 (Supplementary Fig. 4a), the strongest pQTL for sOB-R-plasma levels in adults[18]. Intriguingly, this SNP is located only 12.2 kbp upstream our top SNP rs2767486, with strong LD (r^2 = 0.9) between the BMI-raising and the sOB-R-increasing alleles.

The strong association between variants in the *LEPR* locus and infant BMI suggests an important role of leptin signaling in early growth. The genome-wide significant association with infant BMI for rs10487505 located 20 kbp upstream of *LEP* is therefore noteworthy. This SNP is a known pQTL for circulating leptin levels in adults[19]. The leptin-increasing allele from Kilpeläinen et al.[19] is associated with lower infant BMI in our data. The effect presents a rise-and-fall pattern, rising during the 3-12 months period when the *LEPR* signal is at its plateau, reaching its peak at 1.5 years ($P_{1.5y}$ = 1.3 × 10^{-8}, $\beta_{1.5y}$ = 0.08) before waning (Fig. 1c, Supplementary Fig. 5c, 6). Children homozygous for the alleles associating with higher sOB-R

5

and lower leptin levels exhibited higher mean standardized BMI (+0.65) than children homozygous for the opposite alleles (Fig. 3).

We identified a novel association with BMI in the *LCORL* locus for rs6842303, presenting a similar rise-and-fall pattern with peak effect at 1.5 years ($P_{1.5y} = 7.5 \times 10^{-9}$, $\beta_{1.5y} = 0.09$) (Fig. 1c, Supplementary Fig. 5b, 6). Previously, this marker has been associated with related traits such as birth weight, birth length, infant length, and adult height. Interestingly, rs6842303 has also been associated with peak height velocity in infancy[20], but no association was reported in the largest adult BMI GWASs to date[5,6]. This supports our finding of a transient effect of *LCORL* in early growth.

The second strongest signal was found at the *ADCY3* locus. Biallelic mutations in *ADCY3* have recently been found to cause severe syndromic obesity[21,22]. *ADCY3* is known to interact with *MC4R*, and rare mutations in *MC4R* account for 3-5% of severe obesity[23]. The lead *ADCY3* SNP, rs13035244, showed no association at birth, gradually became genome-wide significant with a peak effect between one and 1.5 years ($P_{1y} = 7.9 \times 10^{-13}$, $\beta_{1y} = 0.10$), and then stabilized during the course of childhood (Fig. 1c, Supplementary Fig. 5a, 6). This result is in agreement with a previous study of growth trajectories in children from one to 17 years of age[24].

In contrast to the rise-and-fall pattern reported here for signals in the *LEPR*, *ADCY3*, *LEP*, and *LCORL* loci, the *FTO* risk allele displayed a trend towards a slightly negative effect around adiposity peak before gradually turning positive from three years of age, reaching genome-wide significance at seven years ($P_{7y} = 2.8 \times 10^{-12}$, $\beta_{7y} = 0.12$), in agreement with Sovio et al.[25]. (Fig. 1c, Supplementary Fig. 5d, 6)

Previous studies have suggested a tight genetic overlap between child and adult BMI, but the details of this relationship across the first years of life remain elusive[24,26]. We used LD score regression[27] in LD Hub[28] to quantify the shared genetic contribution between BMI at each of the

6

12 time points and other traits (Fig. 4a,b and Supplementary Fig. 7). Our data show that although adult BMI and other adult obesity traits normally associated with poor metabolic control were positively correlated with childhood BMI from age 5-8 years, this correlation was much weaker below the age of three years. Notably, the genetic correlation with a range of non-anthropometric traits varied substantially at infant age (Supplementary Fig. 8). Polygenic risk score analyses across all time points for markers associated with birth weight[4], childhood BMI[26], and adult BMI[5,6] revealed similar patterns (Fig. 4c). We also used LD score regression to estimate the SNP-based heritability of BMI measurements across infancy and childhood. The LD score regression-based heritability estimates varied with age, with relatively modest levels at birth and during the adiposity rebound, and high levels when adiposity is high, *i.e.* around adiposity peak and from seven years of age onwards (Fig. 4a). This finding is supported by twin-studies that also show high heritability estimates for BMI in infancy, lower levels around four years of age, followed by higher estimates in later childhood[29]. Collectively, these results further indicate that the genetic mechanisms underlying BMI change from infancy to adulthood.

To our knowledge, we report the first GWAS with dense measurements of BMI during the first year of life. The few GWASs published on BMI in infancy and childhood mainly involve children above five years of age, *i.e.* during adiposity rebound[24,26]. These studies point toward a strong genetic correlation for BMI around adiposity rebound and adulthood. Our results confirm a strong overlap of the genetics of BMI from five to eight years and adulthood, however, this association is much less pronounced during infancy. Infant weight and height have considerable heritable components[30]. Our results suggest that there are distinct molecular mechanisms that dynamically and specifically influence weight gain in infancy, partly acting through leptin signaling.

Leptin has an important role in fetal growth, and is positively correlated with birth weight[31]. Leptin levels are high at birth and decrease quickly, whereas sOB-R levels are low at birth and

7

increase rapidly during the first postnatal days[32]. This pattern is hypothesized to be an important mechanism for suppressing leptin-induced energy expenditure during the first neonatal days. The sOB-R level remains very high during the first two years of life and then declines[33], mirroring the association of *LEPR* with infant BMI observed in our study (Fig. 1c). An effect of genetic variant(s) on the level of sOB-R in infancy is therefore a possible causal mechanism underlying the association with BMI. An interaction between the *LEPR*- and *LEP*-associated variants with increased BMI in individuals who carry both the sOB-R-raising and leptin-lowering alleles would further support a mechanism where sOB-R in circulation sequesters leptin, reducing its membrane receptor activation, hence promoting energy intake during infancy. The SNPs associated with increased BMI during infancy near *LEPR* and *LEP* are not known to affect adult BMI. In fact, they are not in LD with any marker associated with adult diseases, and might thus promote healthy weight gain during infancy, a notion further supported at the genome level by LD score regression. This result is further supported by a recent independent study[34] suggesting that SNPs in the *LEPR*/*LEPROT* locus are associated with BMI at the adiposity peak.

In summary, in our first, large GWAS performed in the Mother Child Cohort of Norway capitalizing on a wealth of phenotypes, the longitudinal analysis uncovers a complex and dynamic influence of common genetic variation on BMI during infant and early childhood growth, dominated by the *LEP-LEPR* axis in infancy. Improved understanding of infant weight biology is important as childhood obesity as well as undernutrition and premature births are worldwide challenges. Our study provides novel knowledge of time-resolved genetic determinants for infant and early childhood growth, suggesting that weight management intervention should be tailored to developmental stage and genetic profile of the patients. For instance, homeostatic increase in the level of sOB-R during infancy might have a positive effect

8

on weight gain without being associated with adult overweight, offering a potential drug target

for ensuring weight gain in infant care.

Methods

Study population

The Norwegian Mother and Child Cohort Study (MoBa) is an open-ended cohort study that recruited pregnant women in Norway from 1999 to 2008. Approximately 114,000 children, 95,000 mothers, and 75,000 fathers of predominantly Norwegian ancestry were enrolled in the study from 50 hospitals all across Norway[3]. Anthropometric measurements of the children were carried out at hospitals (at birth) and during routine visits by trained nurses at 6 weeks, 3, 6, 8 months, and 1, 1.5, 2, 3, 5, 7, and 8 years of age. Parents later transcribed these measurements to questionnaires. In 2012, the project Better Health By Harvesting Biobanks (HARVEST) randomly selected 11,490 umbilical cord blood DNA samples from MoBa's biobank for genotyping, excluding samples matching any of the following criteria: (1) stillborn, (2) deceased, (3) twins, (4) non-existing Medical Birth Registry (MBR) data, (5) missing anthropometric measurements at birth in MBR, (6) pregnancies where the mother did not answer the first questionnaire (as a proxy for higher fallout rate), and (7) missing parental DNA samples. In 2016, HARVEST randomly selected a second set of samples, 5,984, using the same criteria.

Genotyping

For the discovery sample, genotyping was performed using Illumina's HumanCoreExome-12 v.1.1 and HumanCoreExome-24 v.1.0 arrays for 6,938 and 4,552 samples, respectively, at the Genomics Core Facility located at the Norwegian University of Science and Technology, Trondheim, Norway. The replication sample was genotyped using Illumina's Global Screening Array v.1.0 for all 5,984 samples at the Erasmus University Medical Center in Rotterdam, Netherlands. We used the Genome Reference Consortium Human Build 37 (GRCh37) reference genome for all annotations and included autosomal markers only for this study.

10

Genotypes were called in Illumina Genome Studio (v.2011.1 for discovery and v.2.0.3 for replication). Cluster positions were identified from samples with call rate ≥ 0.98 and GenCall score ≥ 0.15. We excluded variants with low call rates, signal intensity, quality scores, heterozygote excess and deviation from Hardy-Weinberg equilibrium (HWE) based on the following QC parameters: call rate < 98%, cluster separation < 0.4, 10% GC-score < 0.3, AA T Dev > 0.025, HWE P-value < 10^{-6}. Samples were excluded based on call rate < 98% and heterozygosity excess > 4 SD. Study participants with non-Norwegian ancestry were excluded after merging with samples from the HapMap project (ver. 3). Sample pairs with PI_HAT > 0.1 in identical-by-descent (IBD) calculations were resolved by removing a random sample in each pair. After genotype calling and QC, 9,286 (80.8%) from the discovery sample set, and 5,235 (87.5%) from the replication sample remained eligible for analysis.

Pre-phasing and imputation

Prior to imputation, insertions and deletions were removed to make the dataset congruent with Haplotype Reference Consortium (HRC) v.1.1 imputation panel using HRC Imputation preparation tool by Will Rayner version 4.2.5 (see URLs): insertions and deletions were excluded. Allele, marker position, and strand orientation were updated to match the reference panel. A total of 384,855 and 568,275 markers remained eligible for phasing and imputation for the discovery and replication set, respectively. Pre-phasing was conducted locally using Shapeit v2.790[35]. Imputation was performed at the Sanger Imputation Server (see URLs) with positional Burrows-Wheeler transform[36] and HRC version 1.1 as reference panel.

Phenotypes

Age, height, and weight values were extracted from Medical Birth Registry (MBR) of Norway for birth, and from the study questionnaires for remaining time points. Pregnancy duration in days was extracted from MBR and pregnancies with duration < 37 × 7 days were excluded (515 pregnancies). Height and weight values were inspected at each age and those provided in centimeter or gram instead of meter and kilogram, respectively, were converted. Extreme outliers, typically an error in handwritten text parsing or a consequence of incorrect units, were excluded (47 length and 8 weight measurements). A value x was considered as extreme outlier if $x > m + 2 \times (perc_{99} - m)$ or $x < m - 2 \times (m - perc_1)$, where m represents the median and $perc_1$ and $perc_{99}$ the 1st and 99th percentiles, respectively.

Subsequently, height and weight curves were inspected for extreme outliers by monitoring the variation of height and weight over time as follows: (i) the height and weight ratio between consecutive ages were calculated at each time point but the last: $r_i = x_{i+1}/x_i$ where r_i is the ratio at time point i and x_i is height or weight at i; (ii) the ratios were scaled after logarithm base 2 transformation, $r'_i = f(\log_2(r_i))$, using the function f of equation 1:

$$f(x_{s,i}) = \frac{x_{s,i} - m_{s,i}}{F_{s,i}^{-1}(\Phi(z)) - m_{s,i}}, z = \begin{cases} 1 & if \ x_i \geq m_{s,i} \\ -1 & otherwise \end{cases} \qquad (1)$$

Where $x_{s,i}$ is the value for an individual of sex s at time point i, $m_{s,i}$ is the median, $F_{s,i}^{-1}$ the empirical quantile function of the values at i of individuals of sex s presenting at least three values before age two (exclusive) and at least two values after age two (inclusive), and Φ the distribution function of the standard normal distribution; (iii) the height or weight of an individual at time point i, presenting surrounding scaled ratios r'_{i-1} and r'_i was considered as

12

outlier and excluded if $r'_{i-1} > 1$ and $r'_i \leftarrow 1$ or if $r'_{i-1} \leftarrow 1$ and $r'_i > 1$, corresponding to peaks or gaps in the curve, respectively.

If for an individual of sex s, two consecutive height values, h_i and h_{i+1} presented a decrease in height, *i.e.* $h_{i+1} < h_i$, this was considered an artefact and corrected as follows.

If the individual presented three or more other height measurements, h_j with $j \neq i$ and $j \neq i+1$, for each j the corresponding height at i and $i+1$ was estimated by interpolating the height curve using the ratios as in equation 2:

$$x_{i,j} = \widehat{r_{i,j}} \times x_j \qquad (2)$$

Where $x_{i,j}$ is the value at i interpolated from j, x_j is the value at j, and $\widehat{r_{i,j}} = \prod_{j}^{i} \hat{r}_k$ if $j < i$ and

$\widehat{r_{i,j}} = \dfrac{1}{\prod_{i}^{j} \hat{r}_k}$ if $j > i$, with $\hat{r}_k$ the median of the ratios r at time point k for the individuals of sex s

presenting at least three values before age two (exclusive) and at least two values after age two (inclusive). If, for all j, $h_i > h_{i,j}$, h_i was considered as outlier and excluded. Similarly, if, for all j, $h_{i+1} < h_{i,j}$, h_{i+1} was considered as outlier and excluded.

Alternatively, if the individual presented two or fewer other height measurements, and $h_i > h_{high}$, h_i was considered as outlier and removed, with h_{high} defined as in equation 3:

$$h_{high} = m_{s,i} + \Phi^{-1}(0.99) \times \left(F_{s,i}^{-1}(\Phi(1)) - m_{s,i} \right) \qquad (3)$$

Where $m_{s,i}$ is the median and $F_{s,i}^{-1}$ the empirical quantile function of the heights at i of individuals of sex s presenting at least three values before age two (exclusive) and at least two

13

values after age two (inclusive), Φ and Φ^{-1} the distribution and quantile functions of the standard normal distribution, respectively. Similarly, if the individual presented two or less other height measurements, and $h_{i+1} < h_{low}$, h_{i+1} was considered as outlier and removed, with h_{low} defined as in equation 4:

$$h_{low} = m_{s,i} - \Phi^{-1}(0.99) \times \left(m_{s,i} - F_{s,i}^{-1}(\Phi(-1)) \right) \qquad (4)$$

If h_i and h_{i+1} were not considered as outliers, h_{i_0} and h_{i+1_0} were defined as the median of $h_{i,j}$ as defined in equation 2, for all $j \neq i$ and $j \neq i+1$, respectively. Starting from $h_{i_k} = h_i$, $h_{i+1_k} = h_{i+1}$, h_i and h_{i+1} were iteratively decreased or increased, respectively, until $h_{i+1} \geq h_i$ as described in equation 5 and 6.

$$h_{i_{k+1}} = \begin{cases} h_{i_0} + 0.9 \times (h_{i_k} - h_{i_0}) & if \left| h_{i_k} - h_{i_0} \right| > \left| h_{i+1_k} - h_{i+1_0} \right| \\ h_{i_k} & otherwise \end{cases} \qquad (5)$$

$$h_{i+1_{k+1}} = \begin{cases} h_{i+1_k} & if \left| h_{i_k} - h_{i_0} \right| > \left| h_{i+1_k} - h_{i+1_0} \right| \\ h_{i+1_0} + 0.9 \times (h_{i+1_k} - h_{i+1_0}) & otherwise \end{cases} \qquad (6)$$

Subsequently, height and weight missing values were imputed from the individual height and weight curves at all ages for individuals presenting at least three values before age two (exclusive) and at least two values after age two (inclusive), and until age two (exclusive), for individuals presenting at least three values before age two (exclusive). A missing value at i was imputed to $x_i = median(x_{i,j})$, with $x_{i,j}$ as defined in equation 2. Importantly, missing values were imputed only if at least two non-imputed values were present at both earlier and later ages. Upon imputation of missing values, outlier removal and height decrease correction was

14

conducted as described previously, and the new missing values were imputed using the same rules.

Finally, the body mass index (BMI) was computed where both height and weight values were available. At each time point, BMI values were scaled prior to association as described in equation 1. These scaled values are referred to as *standardized BMI* in the text.

The analysis of the phenotypes was conducted in R version 3.5.1 (2018-07-02) -- "Feather Spray" (https://www.R-project.org).

Statistical analyses

Genome-wide analyses were performed using SNPTEST v.2.5.2 using dosages of alternate allele with an additive linear model using sex, batch, and ten principal components as covariates. LD score regression was performed with LD Hub v.1.9.0 using LDSC v.1.0.0[27] using all markers remaining after performing pruning recommended by the LD Hub[28] authors.

Figures

All figures in the manuscript were generated in R version 3.5.1 (2018-07-02) -- "Feather Spray" (https://www.R-project.org). In addition to the system packages, the following packages were used: ggplot2 version 3.0.0, scico version 1.0.0, gtable version 0.2.0, ggrepel version 0.8.0, and ggdendro version 0.1-20.

URLs

HRC or 1000G Imputation preparation and checking:

http://www.well.ox.ac.uk/~wrayner/tools; Sanger Imputation Service,

https://imputation.sanger.ac.uk;

Ethics

The study was approved by the Regional Committee for Medical and Health Research Ethics in Norway (#2012/67).

Author contributions

Ø.H. and M.V. performed the analyses.

O.L., T.Z., J.J, J.B., B.J., H.L., K.H., R.T.L., G.P.K., C.S., and P.M.M. contributed to sample acquisition and genotyping.

J.J. and J.B. assisted with genotype quality control.

Ø.H., M.V., S.J., and P.R.N. wrote the manuscript with contributions from all authors.

P.B.J, J.V.S., and A.M. critically revised the manuscript for important intellectual content.

S.J. and P.R.N. designed and directed the study.

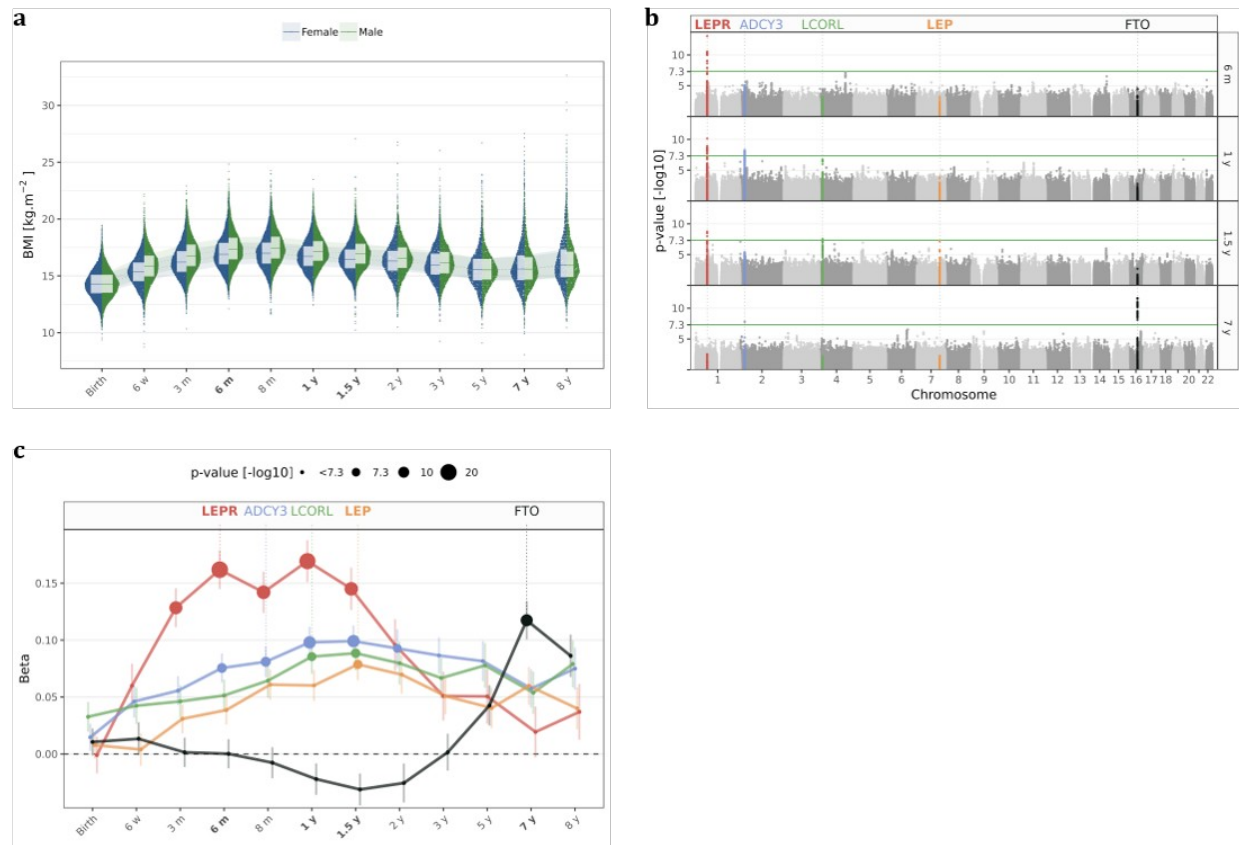P.R.N. secured funding and initiated the study.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgements

17

Figure legends

Fig. 1 | GWAS of BMI from 12 timepoints in the Mother Child Cohort of Norway. a, BMI values in kg/m^2 for all samples used for association are plotted at each time point, with the normalized density of points restricting the jitter along the x-axis to the left for females (blue) and to the right for males (green). Ribbons and box plots, showing the median and quartiles, are plotted in background and foreground, respectively. **b**, Manhattan plots showing association results for the discovery sample (N=9,286) at six time points: 6 months, 1 year, 1.5 year, and 7 years. *LEPR*, *ADCY3*, *LCORL*, *LEP*, and *FTO* loci are highlighted in red, blue, green, orange, and black, respectively. **c**, Regression betas are plotted in sd units at each time point for rs2767486, rs13035244, rs6842303, rs10487505, and rs9922708, the lead SNPs of the *LEPR*, *ADCY3*, *LCORL*, *LEP*, and *FTO* loci, respectively, in the same colors as 1b. Results are presented for the meta-analysis of discovery and replication sample. The size of the points is proportional to -log10(P-value of the association). Error bars represent 1 standard error of mean (SEM) on each side of the point.

18
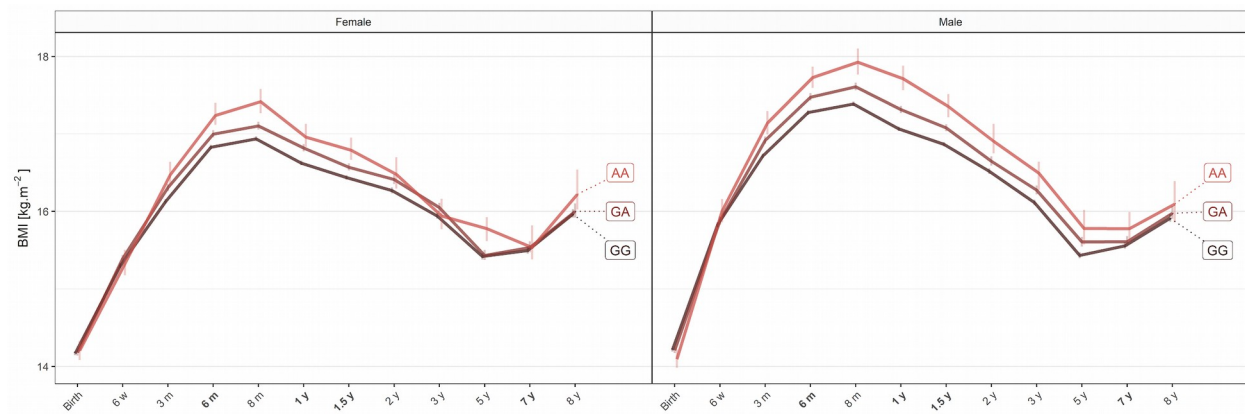
Fig. 2 | BMI trajectories stratified by genotype of the lead SNP of the *LEPR* locus (rs2767486). The median of the BMI in kg/m^2 at each time point for all samples (discovery + replication) stratified by genotype GG, GA, or AA in black, brown, and red, respectively, for females (left) and males (right). Error bars represent 1 SEM on each side of the point. Trajectories for all five GWA-significant SNPs identified in the study are depicted in Supplementary Figure 6.
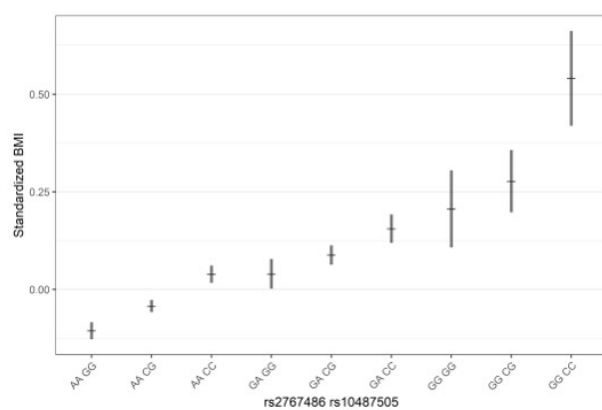
Fig. 3 | Standardized BMI at 1.5 years of the lead SNPs in the *LEPR* and *LEP* loci stratified by the combined genotypes of rs2767486 and rs10487505, respectively. The mean of the standardized BMI of all samples (discovery + replication) at 1.5 years is plotted after stratifying by genotype. Error bars represent 1 SEM on each side of the point.
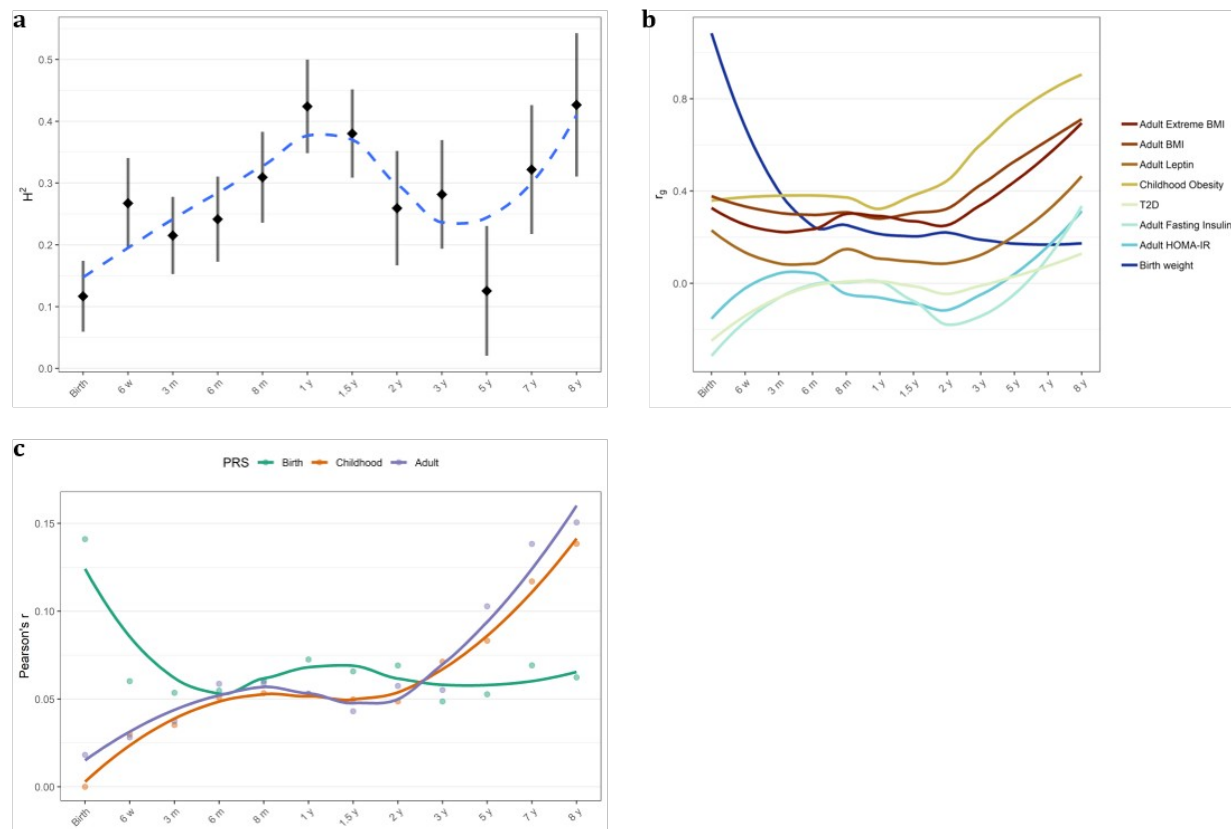
Fig. 4 | Evolution of heritability, correlation with other traits, and risk scores at all time points. a. H^2 estimates based on LD score regression for BMI are plotted at each age (black) along with locally estimated scatterplot smoothing (LOESS) local regression (in blue). Error bars represent 1 SEM on each side of the point. **b**, LOESS local regression of the LD score regression coefficient (r_g) between our BMI association results at each age and phenotypes from LD Hub (i) adult extreme BMI, (ii) adult BMI, (iii) leptin not adjusted for BMI, (iv) childhood obesity, (v) type 2 diabetes, T2D, (vi) fasting insulin main effect, (vii) homeostatic model assessment for insulin resistance, HOMA-IR. **c**, Pearson's r of the correlation between polygenic risk scores (PRS) and BMI at each age along with LOESS local regression. Correlation with PRS for birth weight, childhood BMI, and adult BMI are displayed in green, orange, and purple, respectively.

rsid	Chr	Position	Nearest Gene	Age at Peak	EA/ Non-EA	EAF	Discovery β (SE)	Discovery p-value	Replication β (SE)	Replication p-value	Meta-analysis β (SE)	Meta-analysis p-value	Het.I^2
rs2767486	1	65991203	*LEPR*	6 m	G/A	0.16	0.158 (0.021)	8.8E-14	0.168 (0.029)	4.4E-09	0.162 (0.017)	**2.0E-21**	0.0
rs13035244	2	25134009	*ADCY3*	1 y	C/T	0.44	0.099 (0.017)	5.3E-09	0.095 (0.023)	3.5E-05	0.098 (0.014)	**7.9E-13**	0.0
rs9922708	16	53831146	*FTO*	7 y	T/C	0.44	0.147 (0.021)	2.4E-12	0.064 (0.028)	2.4E-02	0.117 (0.017)	**2.8E-12**	82.4
rs6842303	4	17854055	*LCORL*	1.5 y	T/G	0.28	0.105 (0.019)	2.8E-08	0.057 (0.026)	2.9E-02	0.089 (0.015)	**7.5E-09**	56.2
rs10484505	7	127860163	*LEP*	1.5 y	C/G	0.49	0.094 (0.017)	5.4E-08	0.051 (0.023)	2.6E-02	0.079 (0.014)	**1.3E-08**	54.1
rs9469637	6	33895805	*GRM4*	Birth	G/A	0.99	0.496 (0.091)	5.4E-08	0.140 (0.127)	2.7E-01	0.375 (0.074)	4.1E-07	80.6
rs316344	6	2540316	*SERPINB1*	5 y	A/C	0.34	0.137 (0.023)	1.4E-09	0.003 (0.031)	9.3E-01	0.091 (0.018)	6.8E-07	91.7
rs72829508	10	98109442	*OPALIN*	3 y	A/T	0.97	0.290 (0.054)	8.7E-08	0.063 (0.078)	4.2E-01	0.217 (0.045)	1.1E-06	82.4
rs117320430	11	50242657	*OR4C12*	3 m	T/C	0.02	0.305 (0.054)	2.1E-08	-0.009 (0.07)	9.0E-01	0.188 (0.043)	1.3E-05	92.0
rs113084611	4	145622200	*HHIP*	6 m	A/G	0.05	0.197 (0.037)	7.3E-08	-0.006 (0.047)	9.0E-01	0.122 (0.029)	2.8E-05	91.3
rs112170166	2	730313	*TMEM18*	1.5 y	C/A	0.98	0.370 (0.069)	8.1E-08	-0.034 (0.088)	7.0E-01	0.217 (0.054)	6.3E-05	92.3
rs4880573	10	2707992	*PFKP*	8 y	T/A	0.13	0.192 (0.034)	2.6E-08	-0.062 (0.044)	1.6E-01	0.096 (0.027)	4.0E-04	95.1

Table 1 | Summary statics for the signals that met criteria for replication. For each SNP, the table lists (i) the rsid; (ii) the genomic coordinates in build GRCh37; (iii) the nearest gene; (iv) the age at peak, *i.e.* lowest P-value; (v) the BMI-increasing and non-increasing alleles, EA and Non-EA, respectively; (vi) the BMI-increasing allele frequency (EAF); (vii) the regression beta (β), standard error (SE), and associated P-value for discovery, replication, and meta-analysis; and (viii) heterogeneity I^2.

References

1   Rolland-Cachera, M. F. *et al.* Adiposity rebound in children: a simple indicator for predicting obesity. *Am J Clin Nutr* **39**, 129-135, doi:10.1093/ajcn/39.1.129 (1984).

2   Geserick, M. *et al.* Acceleration of BMI in Early Childhood and Risk of Sustained Obesity. *N Engl J Med* **379**, 1303-1312, doi:10.1056/NEJMoa1803527 (2018).

3   Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* **45**, 382-388, doi:10.1093/ije/dyw029 (2016).

4   Horikoshi, M. *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538**, 248-252, doi:10.1038/nature19806 (2016).

5   Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206, doi:10.1038/nature14177 (2015).

6   Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649, doi:10.1093/hmg/ddy271 (2018).

7   Sainz, N., Barrenetxe, J., Moreno-Aliaga, M. J. & Martinez, J. A. Leptin resistance and diet-induced obesity: central and peripheral actions of leptin. *Metabolism* **64**, 35-46, doi:10.1016/j.metabol.2014.10.015 (2015).

8   Schaab, M. & Kratzsch, J. The soluble leptin receptor. *Best Practice & Research Clinical Endocrinology & Metabolism* **29**, 661-670, doi:https://doi.org/10.1016/j.beem.2015.08.002 (2015).

9   Farooqi, I. S. & O'Rahilly, S. Genetic factors in human obesity. *Obes Rev* **8 Suppl 1**, 37-40, doi:10.1111/j.1467-789X.2007.00315.x (2007).

10   Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398-401, doi:10.1038/32911 (1998).

11   Wheeler, E. *et al.* Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet* **45**, 513-517, doi:10.1038/ng.2607 (2013).

12   Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* **49**, 834-841, doi:10.1038/ng.3841 (2017).

13   Day, F. R. *et al.* Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat Commun* **6**, 8842, doi:10.1038/ncomms9842 (2015).

14   Sabater-Lleal, M. *et al.* Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. *Circulation* **128**, 1310-1324, doi:10.1161/CIRCULATIONAHA.113.002251 (2013).

15   Elliott, P. *et al.* Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA* **302**, 37-48, doi:10.1001/jama.2009.954 (2009).

16   Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e1419, doi:10.1016/j.cell.2016.10.042 (2016).

17   Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* **8**, 14357, doi:10.1038/ncomms14357 (2017).

18   Sun, Q. *et al.* Genome-wide association study identifies polymorphisms in LEPR as determinants of plasma soluble leptin receptor levels. *Hum Mol Genet* **19**, 1846-1855, doi:10.1093/hmg/ddq056 (2010).

19   Kilpelainen, T. O. *et al.* Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. *Nat Commun* **7**, 10494, doi:10.1038/ncomms10494 (2016).

20   Sovio, U. *et al.* Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet* **5**, e1000409, doi:10.1371/journal.pgen.1000409 (2009).

21   Saeed, S. *et al.* Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nat Genet* **50**, 175-179, doi:10.1038/s41588-017-0023-6 (2018).

22   Grarup, N. *et al.* Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nat Genet* **50**, 172-174, doi:10.1038/s41588-017-0022-7 (2018).

23   Siljee, J. E. *et al.* Subcellular localization of MC4R with ADCY3 at neuronal primary cilia underlies a common pathway for genetic predisposition to obesity. *Nat Genet* **50**, 180-185, doi:10.1038/s41588-017-0020-9 (2018).

24   Warrington, N. M. *et al.* A genome-wide association study of body mass index across early life and childhood. *Int J Epidemiol* **44**, 700-712, doi:10.1093/ije/dyv077 (2015).

25   Sovio, U. *et al.* Association between common variation at the FTO locus and changes in body mass index from infancy to late childhood: the complex nature of genetic association through growth and development. *PLoS Genet* **7**, e1001307, doi:10.1371/journal.pgen.1001307 (2011).

26   Felix, J. F. *et al.* Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum Mol Genet* **25**, 389-403, doi:10.1093/hmg/ddv472 (2016).

27   Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).

28   Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279, doi:10.1093/bioinformatics/btw613 (2017).

29   Silventoinen, K. *et al.* Genetic and environmental effects on body mass index from infancy to the onset of adulthood: an individual-based pooled analysis of 45 twin cohorts participating in the COllaborative project of Development of Anthropometrical measures in Twins (CODATwins) study. *Am J Clin Nutr* **104**, 371-379, doi:10.3945/ajcn.116.130252 (2016).

30   Mook-Kanamori, D. O. *et al.* Heritability estimates of body size in fetal life and early childhood. *PLoS One* **7**, e39901, doi:10.1371/journal.pone.0039901 (2012).

31   Vatten, L. J., Nilsen, S. T., Odegard, R. A., Romundstad, P. R. & Austgulen, R. Insulin-like growth factor I and leptin in umbilical cord plasma and infant birth size at term. *Pediatrics* **109**, 1131-1135 (2002).

32   Kratzsch, J. *et al.* Inverse changes in the serum levels of the soluble leptin receptor and leptin in neonates: relations to anthropometric data. *J Clin Endocrinol Metab* **90**, 2212-2217, doi:10.1210/jc.2004-1454 (2005).

33   Kratzsch, J. *et al.* Circulating soluble leptin receptor and free leptin index during childhood, puberty, and adolescence. *J Clin Endocrinol Metab* **87**, 4587-4594, doi:10.1210/jc.2002-020001 (2002).

34   De Silva, N. M. G. *et al.* Genetic architecture of early childhood growth phenotypes gives insights into their link with later obesity. *bioRxiv* (2017).

35   Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6, doi:10.1038/nmeth.2307 (2013).

36   Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* **30**, 1266-1272, doi:10.1093/bioinformatics/btu014 (2014).