



Genome-wide DNA methylation in saliva and body size of adolescent girls

Aim: We performed an epigenome-wide association study within the Finnish Health in Teens cohort to identify differential DNA methylation and its association with BMI in adolescents. **Materials & methods:** Differential DNA methylation analyses of 3.1 million CpG sites were performed in saliva samples from 50 lean and 50 heavy adolescent girls by genome-wide targeted bisulfite-sequencing. **Results:** We identified 100 CpG sites with p-values < 0.000524, seven regions by 'bumphunting' and five CpG islands that differed significantly between the two groups. The ten CpG sites and regions most strongly associated with BMI substantially overlapped with obesity- and insulin-related genes, including *MC2R*, *IGFBPL1*, *IP6K1* and *IGF2BP1*. **Conclusion:** Our findings suggest an association between the saliva methylome and BMI in adolescence.

First draft submitted: 21 April 2016; Accepted for publication: 17 August 2016; Published online: 20 October 2016

Keywords: adolescents • BMI • DNA methylation • epigenome-wide association study • girls • insulin • melanocortin • obesity • saliva • targeted bisulfite-sequencing

The worldwide prevalence of obesity has more than doubled since 1980, and in 2013 there were 2.1 billion overweight and obese people worldwide [1]. Obesity rates have increased in most populations and in all age groups worldwide, and Finland is no exception [2,3]. Overweight and obese children are likely to stay obese into adulthood, and they tend to develop conditions such as Type 2 diabetes and cardiovascular diseases more frequently and at a younger age [4]. Heritability of obesity has been estimated at about 50–70% [5], and more than 90 genetic loci have been associated with obesity-related traits [6]. However, these genetic loci account for only 2.7% of variation in BMI [6]. This suggests that nongenetic factors, such as epigenetics, may be important determinants of obesity. DNA methylation (DNAm) varies by genotype, sex, age [7,8], cell type [9] and environmental influences, such as smoking [10]. DNAm is more plastic *in utero* and early life, potentially in response to environ-

mental factors [11,12]. DNAm is cell type specific and can vary in samples of mixed cell types, such as saliva and blood samples [13,14]. However, most sites in the genome have a common methylation profile across cell types [15–17].

Associations between epigenetic marks and adiposity have been shown in animal models [18,19] and humans [14,20], although it is not always clear if these epigenetic changes preceded obesity, or *vice versa* [14]. Differential DNAm between lean and heavy individuals in metabolically important genes has also been identified in several different cell types [14], and consistent findings have been shown for altered methylation in the *IGF2/H19* imprinted region [21,22]. Genome-wide studies of peripheral blood cells have shown extensive, but small DNAm alterations in genes related to obesity, immune response, cell differentiation and regulation of transcription [14,23–24]. Tissue-specific differential DNAm has been found in leptin,

Trine B Rounge^{*1,2}, Christian M Page^{3,4}, Maija Lepistö⁵, Pekka Ellonen⁵, Bettina K Andreassen² & Elisabete Weiderpass^{1,2,6,7}

¹Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland

²Department of Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway

³Department of Neurology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

⁴Department of Noncommunicable Diseases, Norwegian Institute of Public Health, Oslo, Norway

⁵Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

⁶Department of Medical Epidemiology & Biostatistics, Karolinska Institutet, Stockholm, Sweden

⁷Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway

*Author for correspondence: trine.rounge@krefregisteret.no

pro-opiomelanocortin, glucose transporter-4 and insulin [20,25,26]. A large epigenome-wide association study (EWAS) on whole blood from European cohorts showed obesity-associated DNAm in *HIF3A*, which was replicated in other studies and tissues [27,28].

In adolescents, five gene regions showed differential DNAm between those that lost weight in an intervention study and those that did not [29]. One study reported that DNAm at the *RXRA* locus at birth accounted for 26% of the variation in obesity at age 9 years [30]. Moreover, DNAm in cord blood has been associated with body size and composition in childhood, although few CpG sites were significant after correcting for multiple testing [31].

We performed an EWAS study of saliva DNAm in lean and heavy adolescent girls from the Finnish Health in Teens (Fin-HIT) cohort with the aim to identify differential DNAm and its association with body size. In addition, we aimed to show the usefulness of saliva as a source of DNA in EWAS studies.

Materials & methods

Participants

The Fin-HIT study included about 11,000 adolescents aged 9–12 years from schools throughout Finland. Within the framework of the Fin-HIT study, all adolescents received a measuring tape and written detailed instructions, including pictures, on how to measure and report height, weight and waist circumference at home with the assistance of an adult. We used this information to calculate BMI (weight in kg/height in m²). For the present analysis, we randomly selected 100 Fin-HIT cohort members aged 11 years: 50 from the tenth percentile of BMI (<15 kg/m²) and 50 from

the 90th percentile of BMI (>21.9 kg/m²), referred to as lean and heavy girls, respectively (Table 1). The Fin-HIT participants reported their menarcheal status in the baseline electronic questionnaires, answered at school on tablets. Of the girls included in this study 60 had reached puberty (menarche = yes), 35 had not (menarche = no) and five did not answer the question (missing). The menarche status was included as a covariable in the model to adjust for weight differences due to puberty.

The Coordinating Ethics Committees of the Hospital Districts of Helsinki and Uusimaa approved the study, and informed consent was obtained from all adolescents and one of their legal guardians.

Procedures

All Fin-HIT cohort members provided saliva samples at baseline using the Oragene DNA Self-Collection Kit (OG-500, DNAgenotek). DNA was extracted from these samples using an automated protocol and chemagic DNA Saliva Kit (PerkinElmer, MA, USA). Three µg of DNA were used as an input in the targeted bisulfite sequencing (TBS) protocol. The SureSelect^{XT} Human Methyl-Seq (Agilent, CA, USA) protocol, including DNA shearing, ligation of methylated adapters, capturing by hybridization, bisulfite conversion using Zymo Research’s EZ DNA Methylation-Gold Kit (Zymo, CA, USA), PCR amplification, indexing and sample pooling, was used according to the manufacturer’s recommendations. The predesigned target system captured a 84 Mb target and 3.7 million CpG sites. This includes CpG islands (20 Mb), cancer and tissue specific differential methylated regions (10 Mb), Gencode promoters (37 Mb), CpG island

Table 1. Overview of the characteristics of the 100 11-year-old girls from the Finnish Health in Teens study and corresponding saliva samples.

Phenotype data	Mean (min/max)	Lean mean (n = 50) (min/max)	Heavy (n = 50) (min/max)	p-value
BMI (kg/m ²)		14.4 (12.8/15)	24.1 (21.9/32.1)	<2.2 × 10 ⁻¹⁶
Waist circumference (cm)		58.2 (52/64)	79.6 (70/93)	<2.2 × 10 ⁻¹⁶
Menarche		0.64	0.614	1
Sequencing data	Mean	Min	Max	
High-quality sequencing pairs	42 mill	19 mill	80 mill	
Cytosine position analyzed	1647 mill	49 mill	2922 mill	
CpG at 10× coverage	3.1 mill	2.3 mill	6.7 mill	
CpG at 50× coverage	1.2 mill	0.1 mill	2.2 mill	
CpG at 100× coverage	0.4 mill	3481	1.0 mill	
Methylated CpG autosomal (%)	49.2	43.5	60.8	
Methylated CpG mitochondria (%)	2.3	0.63	20.6	

Max: Maximum; Mill: Million; Min: Minimum.

shores and shelves, DNase I hypersensitive sites, Ref-seq genes and Ensembl regulatory features (48 Mb) (Supplementary Information 1). The libraries were sequenced on a HiSeq (Illumina, CA, USA) with a 2×93^{101} bp length and an average depth of 13 gigabases per sample. TBS produces DNAm values for all CpGs within its target, yielding a consecutive high-resolution view of the DNAm over entire regions.

Bioinformatics & statistical analyses

A custom script produced β -methylation values from raw sequencing data by combining open-source softwares. First, low-quality sequences and adaptors were removed using the Nsoni clip (version 0.115) [32]. After removal of duplicate sequences, the bisulfite-converted sequence reads were then mapped to the human genome (hg19) using Bismark (version 0.10) [33]. This software transformed the sequences into a C–T and G–A version and each of the transformed sequences were aligned to equivalently preconverted reference genome using four parallel instances of the short read aligner Bowtie2 (version 2.0.5) [34]. The Bismark methylation extractor and custom formatting scripts were used to calculate β -methylation values for CpG sites. The first seven bases of each sequence were ignored, as a strong bias toward nonmethylation was caused by the insertion of unmethylated cytosines during end-repair in the sequencing library preparation. CpG sites at which more than 25% of the samples had less than $10 \times$ coverage were discarded. Cytosines in a non-CpG context and mitochondrial DNA is only methylated to a small extent, thus hypermethylation can be interpreted as less than optimal efficiency of the of bisulfite conversation reaction. Methylation of non-CpG cytosines and mitochondrial CpG were highly correlated among the samples ($r^2 = 0.99$) (Supplementary Information 2). We used the average level of mitochondrial DNAm in the sample as a measurement of bisulfite conversation efficiency in the model.

Differential DNAm at CpG sites between lean and heavy girls was identified using a logistic regression model adjusting for puberty status and mitochondrial DNAm. The accumulation of the samples at the extreme end of the waist circumference or BMI range makes linear regression unsuitable. Thus, a case–control design with logistic regression was found most appropriate to identify differentially methylated CpG sites. After adjustment for multiple testing with a false discovery rate, no individual CpG site was significant.

To identify differentially methylated regions by combining near-by CpG sites, we used two different approaches: Bumphunter and Fisher's method [35,36]. Bumphunter was used to identify differentially methylated regions independent of gene annotation, as

implemented in R by Aryee *et al.* [37]. We used the same covariates as in the logistic regression model described above, and set a cut-off for the first screening step at the top 0.1% of signals. The maximum allowed distance between CpG sites within the same regions was set to 1500 bp, and the number of bootstraps to 500. Since this approach is independent of gene annotation, and can give results that are difficult to interpret in a biological setting, we also tested each CpG island for any enrichment of p-values. This was done by aggregating the p-values from all CpG sites within each island, and applying an extension of Fisher's method [36], to summarize the individual CpG-wise p-values in to one p-value for each CpG Island. The correlation between the p-values was assumed to be the same as the correlation between the methylation values, estimated using the Spearman's rank correlation coefficient. The qq-man application in R was used to produce Manhattan plots and QQ-plots [38].

Pathways enriched for differential DNAm were evaluated using the improved gene set enrichment analysis for genome-wide association studies (i-GSEA4GWAS v1.1) [39]. Predefined biological pathways and processes, including Biocarta, KEGG, Reactome pathways and gene ontology gene sets were assessed for enrichment. CpG sites up to 10 kb upstream and downstream of a given gene were mapped to the gene and the maximum $-\log(p\text{-value})$ of all the CpG sites mapped to that gene was used to represent the gene. i-GSEA4GWAS uses site label permutation. We report the p-value and false discovery rate for the statistically significant gene sets < 0.001 and the associations for each CpG site and island for the outcome variable BMI.

Results

The difference in mean BMI between lean and heavy girls was 9.7 kg/m^2 , and the difference in waist circumference was 21.4 cm. Menarche was reported by 64% of lean girls and 78% of heavy girls (Table 1).

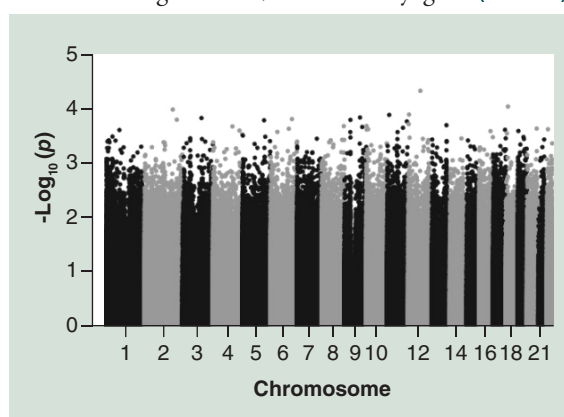


Figure 1. Manhattan plot of CpG sites with differential DNA methylation in lean and heavy adolescents.

We sequenced 3.1 million CpG sites with the minimum 10× coverage, and 1.2 million of these sites had more than 50× coverage. The mean DNAm on autosomal CpG sites for all samples was 49.2%, and we could not identify a statistically significant difference between the global autosomal DNAm levels between the groups. As expected, most mitochondrial CpGs were hypomethylated.

The genome-wide differential DNAm analyses between lean and heavy girls, adjusted for menarcheal status and bisulfite conversion, resulted in a top list of 100 CpG sites with p-values below 5.28×10^{-4} (Figure 1), but none that were genome-wide significant. More than 300 CpG had a p-value lower than 10^{-3} . The two differential CpG sites (chr12: 78227663; p-value 4.90×10^{-5} and chr18: 13868825; p-value 9.53×10^{-5}) most strongly associated with BMI were located within the *NAV3* intronic region and within the CpG island upstream of the obesity-related *MC2R* gene (Table 2). For both sites a significant hypermethylation in heavy relative to lean girls was observed (Figure 2A & B). The difference in DNAm varied across the groups in at least three closely related sites between 13868800 and 13868800 on chromosome 18 (*MC2R* region), with position 13868825 (cg27031837) showing the strongest association (Figure 2D). No notable difference in DNAm was observed surrounding chromosome 12 position 78227663 (Figure 2C). The top ten list of strongest associated CpG sites included four sites that were located in or nearby obesity-related genes (*MC2R*, *TMOD1*, *SLC35D3* and *IGFBPL1*), and two sites (*MICAL2* and *NRP2*) were located in or nearby genes of interest in obesity research. Four CpG sites over-

lapped with targets on the 450K array (cg27031837, cg24371251, cg02203665 and cg01154445).

Significant differences in DNAm were observed in genomic regions and CpG islands associated with BMI. Seven regions, independent of annotation, were identified with the software bumphunter to have a significant aggregation of closely located differential DNAm sites (Table 3). The region with the strongest association with BMI was located in a gene-dense region upstream of the diabetes-related *IP6KI* gene. ‘Bumphunting’ also identified a region including 18 CpG sites within the CpG island upstream of the *MPL* gene to be significantly different methylated in lean and heavy girls. The annotation specific analyses using the extended Fishers method identified five CpG islands to have genome-wide significant differential DNAm associated with BMI (Table 3). Notably, the CpG islands upstream of the *IGF2BP1* gene was among the top five hits (Table 3). The TBS allowed a detailed view of correlation between differential DNAm within genomic locations. A consecutive mean difference in DNAm throughout the regions was observed in proximity of *NXN*, *MPL*, *ABLIM2* and *IGF2BD1* (Figure 3C, E, F & J, respectively). A few colocalized sites were driving the DNAm difference over the regions close to *MYO1E/LDHAL6B* and *TRPV5* (Figure 3B & D, respectively). No clear correlation in differential DNAm could be observed in proximity of *AMIGO3*, CpG island 24 (6512) on chromosome 5, *PCDH19* and *GRK4* (Figure 3A, G, H & I, respectively).

Gene set enrichment analysis identified pathways enriched for genes in a 10-kb range of CpG sites with

Table 2. The ten DNA methylation sites most strongly associated with BMI.

Chr	Pos	p-value adjusted	Heavy relative to lean	n	Mean coverage	Close genes	Obesity related [†]	Ref.
12	78227663	4.90×10^{-5}	↑	93	34	<i>NAV3</i>		
18	13868825	9.53×10^{-5}	↑	93	45	<i>MC2R</i> , CpGi:37, cg27031837	Yes	[40]
2	180728399	1.08×10^{-4}	↑	94	111	<i>ZNF385B/mir1258</i> , cg24371251		
12	5153954	1.33×10^{-4}	↑	94	77	<i>KCNA5</i> , CpGi:125		
11	12114874	1.36×10^{-4}	↓	94	255	<i>MICAL2</i>	Potentially	[41]
9	100262507	1.51×10^{-4}	↑	94	115	<i>TMOD1</i> , cg02203665	Yes	[42]
3	122720613	1.54×10^{-4}	↑	94	101	<i>SEMA5B</i>		
6	137244696	1.62×10^{-4}	↑	93	49	<i>SLC35D3</i> , CpGi:270	Yes	[43]
2	206548172	1.66×10^{-4}	↓	93	41	<i>NRP2</i> , cg01154445	Potentially	[44]
9	38487226	1.67×10^{-4}	↑	93	63	<i>IGFBPL1/FAM95C</i>	Yes	[45]

↑ Hyper- and ↓ hypo-methylation in heavy relative to lean girls. ↑↓ No consistent directionality.
[†]Examples are referred.
 Chr: Chromosome; Pos: Position.

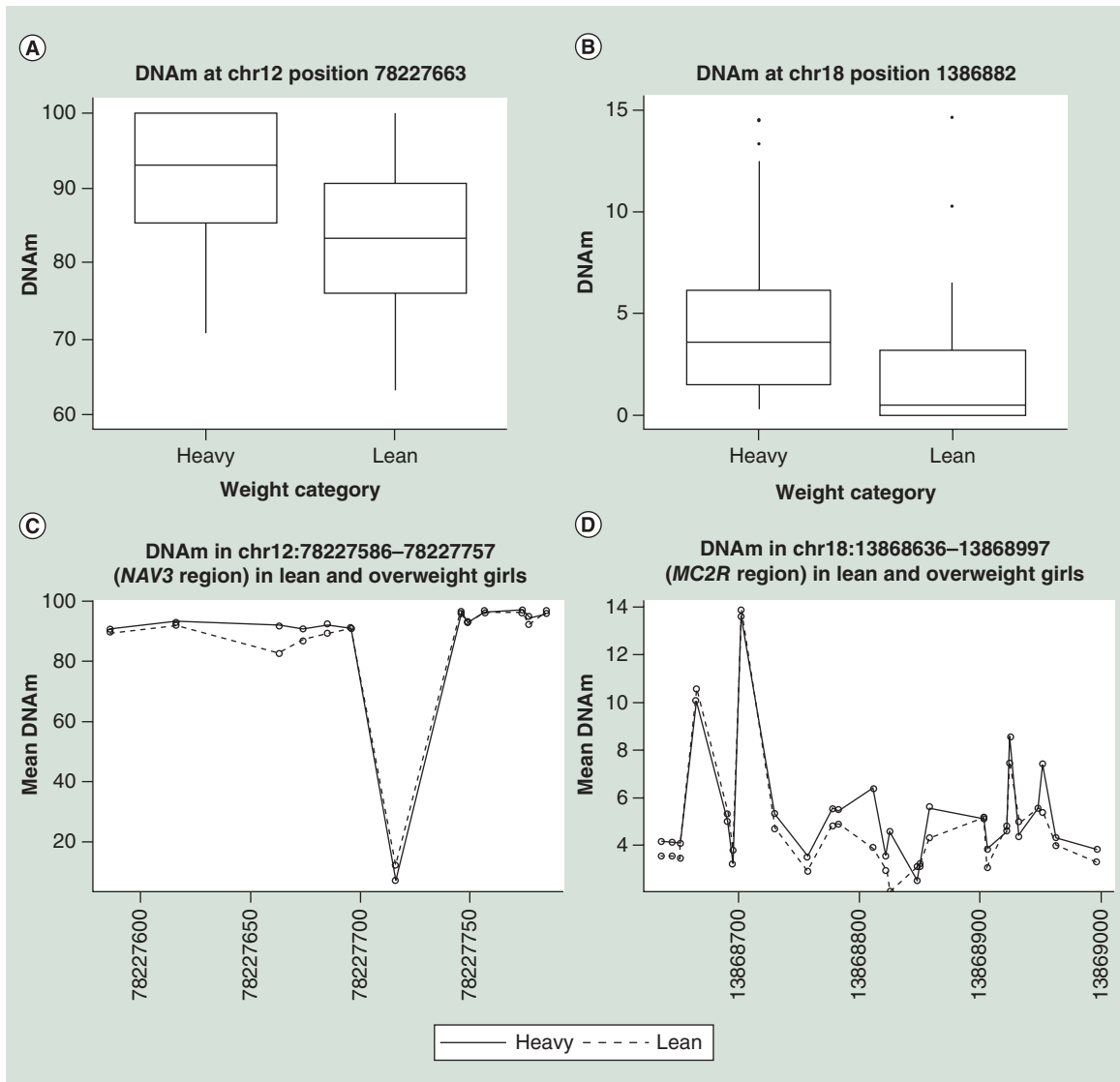


Figure 2. Detailed overview of DNA methylation of the two associated CpG sites with the strongest association with BMI and the surrounding regions. A boxplot of DNAm values for the lean and heavy group for: (A) position chr12:78227663 and (B) chr18:13868825. The targeted bisulfite sequencing also captures DNAm of the surrounding regions, here showing difference in mean DNAm in the lean group (dashed line) and heavy group (solid line) for: (C) chr12:78227586–78227757 and (D) chr18:13868730–13868997. chr: Chromosome; DNAm: DNA methylation.

differential DNAm. Pathways such as metabolism of androgen and estrogen, starch and sucrose, and porphyrin and chlorophyll were enriched with nominally significant CpG sites (Table 4). All these gene sets contained the genes *UGT1A1–10*.

Discussion

We identified genomic sites, regions and pathways that differed in saliva DNAm between 11-year-old lean and heavy Finnish girls by investigating 3.1 million CpG sites. Many of the sites and regions were collocated with known obesity-related genes, predominantly in the insulin-melanocotin pathway. Considering the limited

sample size of 50 girls in each group, we identified a number of noteworthy CpG sites and regions. Collocation of the ten CpG sites most strongly associated with BMI and obesity-related gene regions [40–45], as well as the fact that our results replicate previous findings [31], strengthening the probability that the observed associations were true associations. We observed differential DNAm at chr12:78227663 nearby the gene *NAV3*, which, to our knowledge, has not been previously associated with body size. It is plausible that differential DNAm at chr12:78227663 may affect the expression of *SYT1*, which is located 500 kb downstream of *NAV3*. *SYT1* is a regulator of exocytosis and endocytosis of

Table 3. The differential DNA methylation regions most strongly associated with BMI identified by Bumphunter or CpG island analyses.

Method	Chr	Start	End	Mean coverage	p-value	Adjusted p-value	Heavy relative to lean	CpGs	CpGi ID/close genes
BH	chr3	49755876	49756190	44	7.33×10^{-05}	0.020	↑	36	CpGi:136/AMIGO03, RNFI23, GMPPB, IP6K1
BH	chr15	59498717	59499229	27	1.04×10^{-04}	0.026	↕	19	LDHAL6B, MYO1E, MIR2116
BH	chr17	821592	821596	63	1.20×10^{-04}	0.036	↑	2	NXN, gc16094767, rs62070219, rs62070218
BH	chr7	142625882	142626550	28	2.17×10^{-04}	0.067	↑	8	TRPV5
BH	chr1	43815128	43815324	92	4.27×10^{-04}	0.082	↑	18	CpGi:76/MPL
BH	chr18	77586199	77586240	43	4.74×10^{-04}	0.093	↑	6	CpGi:73
BH	chr12	70793386	70793517	90	3.21×10^{-04}	0.101	↑	3	KCNMB4, intron
CpGi	chr4	8087171	8087220	24	6.73×10^{-06}	0.17	↕	5	5609/ABLIM2
CpGi	chr5	3324770	3324967	34	1.07×10^{-04}	0.78	↑	15	6512
CpGi	chrX	99666133	99666378	27	3.26×10^{-04}	0.78	↓	27	13196/PCDH19
CpGi	chr4	3043200	3043540	29	7.35×10^{-04}	0.78	↑	27	5506/GRK4
CpGi	chr17	47091975	47092197	49	7.05×10^{-04}	0.78	↑	17	22117/IGF2BD1

↑ Mainly hyper- and ↓ hypo-methylation in heavy relative to lean girls. ↕ No consistent directionality.
 BH: Bumphunter analyses; Chr: Chromosome; CpGi: Extended Fisher analyses of CpG island.

insulin-containing vesicles [46] and is associated with eating behavior in pigs [47].

The *MC2R* CpG island harbored the second strongest signal from a CpG site. The melanocortin system plays an important role in the pathogenesis of obesity and in metabolic syndrome [40]. Mutations in the *MC4R* gene are the most common cause of monogenic obesity in humans, and play a role in regulating energy expenditure and energy intake through the control of satiety. *MC3R* has a more subtle role in energy homeostasis [48], whereas *MC2R* is involved in the regulation of steroidogenesis and has recently been indicated to have a direct effect on adipose insulin sensitivity, endocrine function and thermogenesis [40]. Our findings support a role for *MC2R* in key adipose functions and further suggest that all melanocortin receptors play crucial roles in the regulation of energy balance [48]. Our data show that DNAm differs in at least three CpG sites within the CpG island upstream of *MC2R*, thus these entire regions should be investigated further. Most of the genes with the strongest associations (*MICAL2*, *TMOD1*, *SLC35D3*, *NRP2*, *IGFBPL1*) have been linked to obesity either directly or as a receptor of an obesity-related gene. *IGF* has consistently been identified as a differentially methylated region associated with BMI regardless of cell type [21,22], which we have now replicated in saliva.

TBS enables DNAm association analyses of single CpG sites, but also enables assessment of aggregated or correlated methylation within genes and regulatory regions, or regional assessment independent of annotation. With increased statistical power from aggregation of DNAm independently of annotation using ‘Bumphunter’ the strongest association was found upstream of the insulin-stimulated gene *IP6K1* [49]. We replicated the findings of Relton *et al.*, which showed associations between core blood *MPL* differential DNAm and childhood obesity [31], using ‘Bumphunter’. Since our results based on saliva DNA replicated associations identified in core-blood DNAm [31], we might speculate that the differential DNAm found in saliva cells may already be present *in utero*. A detailed look at this region (Figure 3E) showed that lean adolescents had lower DNAm throughout a 200 bp region. A recent study indicated that a variant in the *GRK4* gene and high intake of sodium increase the risk of obesity [50]. Our CpG island association results, also benefitting from increased statistical power due to aggregation of methylation at CpG islands, indicate that DNAm in the upstream CpG island *GRK4* (Figure 3I) might also be of interest. The CpG island nearby *IGF2BP1* also demonstrated an association between the insulin pathway and DNAm in saliva. This finding replicates previous associations between *IGF2BP1* DNAm and

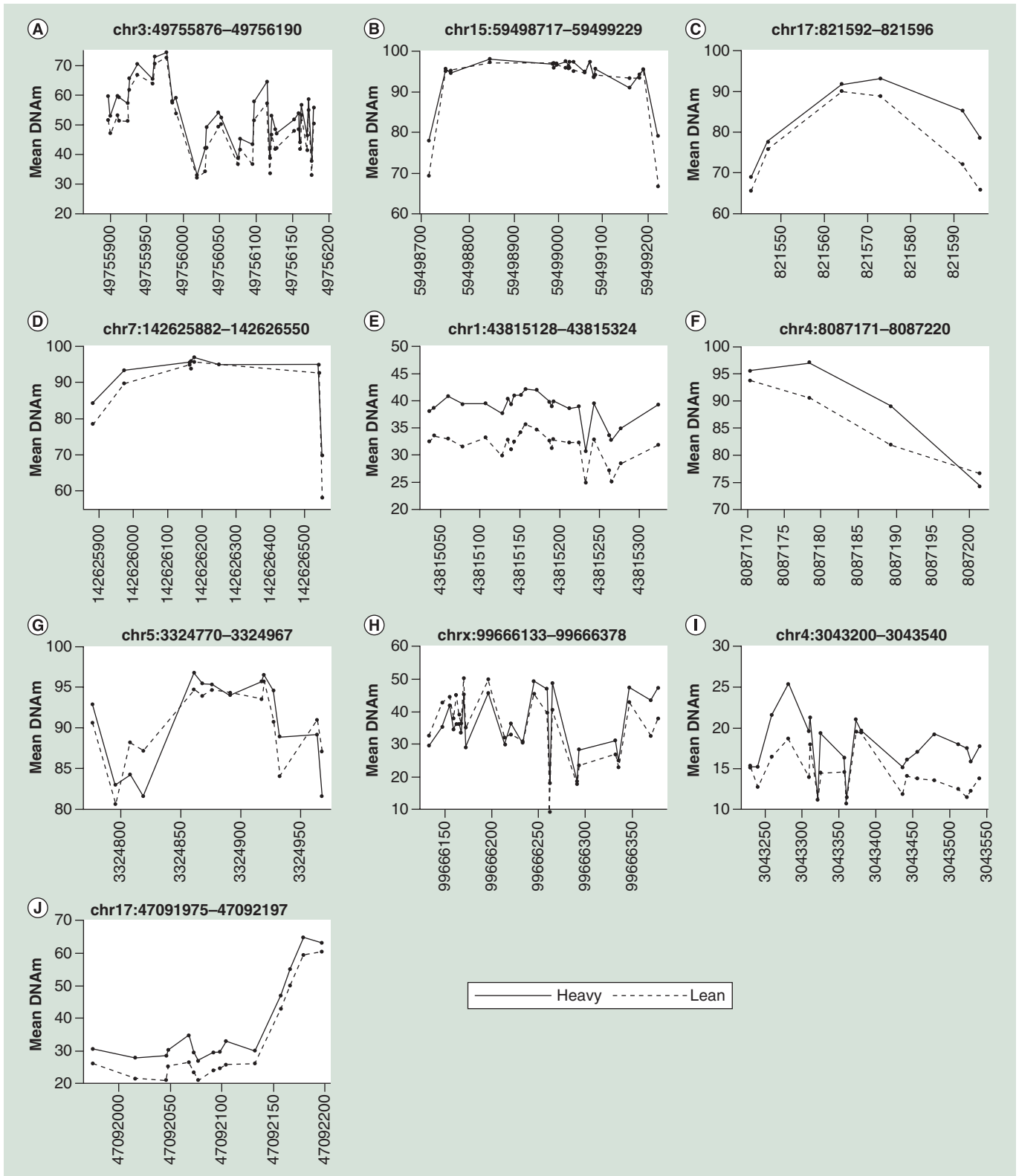


Figure 3. The mean CpG DNA methylation in lean (dashed line) and heavy (solid line) throughout the five regions with the strongest associations with BMI identified by Bumphunter and CpGi analyses. The difference in DNA methylation means are due to a consecutive difference in the same direction (C), (E), (F) and (J), differences with intermediate non-different CpGs in (D) and (I), DNA methylation difference in a few CpGs illustrated by (B), or alternating methylation levels illustrated by (A), (G) and (H). chr: Chromosome; DNAm: DNA methylation.

Table 4. Gene set enrichment analyses of CpG sites and islands associated to BMI and waist circumference.

Pathway	p-value	FDR	Significant/selected/all genes
HSA00150 androgen and estrogen metabolism	0.0010	0.060	40/40/54
HSA00860 porphyrin and chlorophyll metabolism	<0.001	0.084	31/32/41
Starch and sucrose metabolism	0.0020	0.141	32/32/44

Distance between gene and CpG site or island were 10 kb.
FDR: False discovery rate.

body mass [51], as well as a recent study on monozygotic twins, which indicated that the differences in DNAm are due to nonshared environmental factors that influence body weight [52].

Three pathways were enriched for differential DNAm, and previous studies have also shown obesity-related epigenetic changes in some of these pathways [53,54]. However, the enrichment in all three pathways was driven mainly by DNAm differences in the genes *UGT1A1-10*; thus further studies on these genes are warranted.

Several studies have identified associations between BMI and DNAm in the *H19/IGF2* region on chromosome 11 [21,22] and more recently in the *ABCG1* gene region on chromosome 21 [27,55]. None of these gene regions had signals among our top 100 findings, although two signals downstream of *H19* (chromosome 11, 2020577 and 2020553) were nominal significant (both with p-value = 0.002). We also identified a signal downstream of the *ABCG1* CpG island (p-value = 0.003). Lack of power and difference in tissue type (blood vs saliva) may explain why we did not replicate these findings.

All individuals in this study were girls born in the same year, removing potential confounding by age [8] and sex. 11-year-olds are also less likely to have a history of potential DNAm confounders such as dieting, medication use and smoking than older girls and adults [56,57]. Menarche was used as an indicator of puberty onset and likely effects DNAm levels [58,59]; hence, menarche status was included in the analytical models. This adjustment might conceal associations with regions such as the *CYP19A1* promoter CpG island, which has previously been found to be associated with puberty [58]. The study design used the extremes of the BMI curve. This design was chosen to compensate for the relatively small sample size but may mask similarities in DNAm among lean and heavy girls that differ in normal-weight adolescents. Although BMI is only an indicator of body fat in adolescents [60], it is likely that the substantial difference in BMI between the tenth and 90th percentiles indicates a clear distinction in body size between these groups.

Gene expression in saliva and DNAm may be influenced by factors such as diet. However, evaluating the potential role of reverse causation is an issue outside the scope of this study. Varying cell-type composition is a challenge in analyses of the saliva methylome, as it is in analyses of the blood methylome. There is a good reference cell-type composition for the blood methylome. A similar good reference does not exist for the saliva methylome. This, in combination with a lack of knowledge on the CpG sites outside the 450K array target [61], made a reference-free cell-type adjustment approach [62] impossible. A correlation between BMI in adolescents and cell-type composition in saliva is unlikely, thus we believe the results without this adjustment are valid.

Most recent EWAS studies were based on the Infinium human Methylation 450K or the older 27K array. However, our study used a TBS approach, which covered 84 Mb of the genome and enabled the detection of differential DNAm with high genomic resolution. While the 450K covers about approximately 1.7%, the TBS covered about 11.1% of the CpG sites in the human genome. Considering that most of the identified regions of interest presented here are not covered by 450K, this difference in resolution is important. It will be of great clinical interest to confirm the saliva DNAm associations to body size and the directionality of these findings. Further investigations of the epigenetic mechanisms of body mass regulation for each of the associations are warranted.

Conclusion & future perspective

In summary, the saliva methylome EWAS of 50 lean and 50 heavy Finnish girls identified differentially methylated CpG sites. The two strongest signals were located near *NAV3* and *MC2R*. The substantial overlap with obesity- and insulin-related genes, including *MC2R*, *IGFBPL1*, *IP6K1* and *IGF2BP1*, and the replication of *MPL*, indicate true associations between the saliva methylome and BMI. Saliva is a readily available source of DNA that can shed light on the etiology of DNAm and obesity. In the future, saliva DNAm could

be used as a biomarker to identify persons at risk of developing obesity.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/epi-2016-0045

Acknowledgements

The authors would like to thank the participants of the Fin-HIT study and the steering committee: S Simola, S Von Kremer, J Skand, C Sarkkola and S Raju, for coordinating samples and data management; and J Erikson, E Roos and A-E Lehesjoki for supporting the study. The authors further thank the Institute for Molecular Medicine Finland (FIMM) biobank for sample handling and DNA extraction, and the FIMM technology center for sequencing and providing computational resources.

Financial & competing interests disclosure

This work was supported by the Folkhälsan Research Foundation: The Academy of Finland (grant number 250704); The Life and Health Medical Fund (grant number 1-23-28);

The Swedish Cultural Foundation in Finland (grant number 15/0897); The Signe and Ane Gyllenberg Foundation (grant number 37-1977-43); and The Yrjö Jahnsson Foundation (grant number 11486). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Writing assistance was provided by Trudy Perdrix-Thoma at Professional Standards Editing, Inc. and was funded by the above-mentioned grants.

Ethical conduct of research

The Coordinating Ethics Committees of the Hospital Districts of Helsinki and Uusimaa approved the study, and informed consent was obtained from all adolescents and one of their legal guardians.

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Executive summary

- An epigenome-wide association study between lean and heavy 11-year-old girls identified differential DNA methylation in 100 CpG sites with p-values below 5.28×10^{-4} .
- The two differential DNA methylation sites with the strongest association with BMI were located within the *NAV3* and *MC2R* genes.
- Our finding support a role for *MC2R* in key adipose functions.
- 'Bumphunting' identified seven associated regions; the region with the strongest association with BMI was located close to the diabetes-related *IP6K1* gene.
- Associations between the *MPL* gene region and childhood obesity, previously shown in blood, were replicated in saliva.
- The CpG islands upstream of *IGF2BP1* were among those that were most significantly associated with body size.
- Gene set enrichment analysis identified diet-related pathways associated with body size.
- Associations between body size and methylation of CpG sites close to obesity- and insulin-related genes indicate an effect of the saliva methylome on BMI or *vice versa*.

References

Papers of special note have been highlighted as:

• of interest; •• of considerable interest

- 1 Ng M, Fleming T, Robinson M *et al*. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 6736, 1–16 (2014).
- 2 Vuorela N, Saha M-T, Salo M. Prevalence of overweight and obesity in 5- and 12-year-old Finnish children in 1986 and 2006. *Acta Paediatr.* 98, 507–512 (2009).
- 3 Vuorela N, Saha MT, Salo MK. Change in prevalence of overweight and obesity in Finnish children – comparison between 1974 and 2001. *Acta Paediatr.* 100, 109–115 (2011).
- 4 Park MH, Sovio U, Viner RM, Hardy RJ, Kinra S. Overweight in childhood, adolescence and adulthood and cardiovascular risk in later life: pooled analysis of three British birth cohorts. *PLoS ONE* 8, 3–8 (2013).
- 5 Hebebrand J, Hinney A. Environmental and genetic risk factors in obesity. *Child Adolesc. Psychiatr. Clin. N. Am.* 18, 83–94 (2009).
- 6 Locke AE, Kahali B, Berndt SI *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–206 (2015).
- 7 Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Iris E. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS ONE* 4, 21–23 (2009).
- 8 Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 14, R115 (2013).
- 9 Meissner A, Mikkelsen TS, Gu H *et al*. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770 (2010).
- 10 Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation:

- 27K discovery and replication. *Am. J. Hum. Genet.* 88, 450–457 (2011).
- 11 Wang D, Liu X, Zhou Y *et al.* Individual variation and longitudinal pattern of genome-wide DNA methylation from birth to the first two years of life. *Epigenetics* 7, 594–605 (2012).
- 12 Alisch RS, Barwick BG, Chopra P *et al.* Age-associated DNA methylation in pediatric populations age-associated DNA methylation in pediatric populations. *Genome Res.* 22, 623–632 (2012).
- 13 Jacoby M, Gohrbandt S, Clause V, Brons NH, Muller CP. Interindividual variability and co-regulation of DNA methylation differ among blood cell populations. *Epigenetics* 7, 1421–1434 (2012).
- 14 van Dijk SJ, Molloy PL, Varinli H, Morrison JL, Muhlhauser BS. Epigenetics and human obesity. *Int. J. Obes. (Lond.)* 39, 1–13 (2014).
- **Provides overview of global methylation, gene specific and genome-wide methylation associated to human obesity from 46 recent studies.**
- 15 Ziller MJ, Gu H, Müller F *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481 (2013).
- 16 Adalsteinsson BT, Gudnason H, Aspelund T *et al.* Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS ONE* 7, 1–9 (2012).
- 17 Talens RP, Boomsma DI, Tobi EW *et al.* Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J.* 24, 3135–3144 (2010).
- 18 Dolinoy DC. The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Mutr. Rev.* 29, 997–1003 (2008).
- 19 Öst A, Lempradl A, Casas E *et al.* Paternal diet defines offspring chromatin state and intergenerational obesity. *Cell* 159, 1352–1364 (2014).
- 20 Lavebratt C, Almgren M, Ekström TJ. Epigenetic regulation in obesity. *Int. J. Obes.* 36, 757–765 (2012).
- **Reviews the evidence of the epigenetic mechanisms in the regulation of obesity.**
- 21 Drake AJ, McPherson RC, Godfrey KM *et al.* An unbalanced maternal diet in pregnancy associates with offspring epigenetic changes in genes controlling glucocorticoid action and foetal growth. *Clin. Endocrinol. (Oxf.)* 77, 808–815 (2012).
- 22 Huang R-C, Galati JC, Burrows S *et al.* DNA methylation of the IGF2/H19 imprinting control region and adiposity distribution in young adults. *Clin. Epigenetics* 4, 21 (2012).
- **Shows associations between IGF2/H19 methylation subcutaneous fat in adolescents.**
- 23 Almén MS, Nilsson EK, Jacobsson JA *et al.* Genome-wide analysis reveals DNA methylation markers that vary with both age and obesity. *Gene* 548, 61–67 (2014).
- 24 Xu X, Su S, Barnes VA *et al.* A genome-wide methylation study on obesity: differential variability and differential methylation. *Epigenetics* 8, 522–533 (2013).
- 25 Plagemann A, Harder T, Brunn M *et al.* Hypothalamic proopiomelanocortin promoter methylation becomes altered by early overfeeding: an epigenetic model of obesity and the metabolic syndrome. *J. Physiol.* 587, 4963–4976 (2009).
- 26 Kuroda A, Rauch TA, Todorov I *et al.* Insulin gene expression is regulated by DNA methylation. *PLoS ONE* 4, e6953 (2009).
- 27 Demerath EW, Guan W, Grove ML *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change, and waist circumference in African American adults identifies multiple replicated loci. *Hum. Mol. Genet.* 24, 4464–4479 (2015).
- 28 Dick KJ, Nelson CP, Tsaprouni L *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* 6736, 1–9 (2014).
- **Large genome-wide analysis study of methylation at CpG sites in relation to BMI in European adults. The discovery sample set included blood-based methylation data from 479 individuals and the findings were replicated in over 2000 individuals.**
- 29 Moleres A, Campión J, Milagro FI *et al.* Differential DNA methylation patterns between high and low responders to a weight loss intervention in overweight or obese adolescents: the EVASYON study. *FASEB J.* 27, 2504–2512 (2013).
- 30 Godfrey KM, Sheppard A, Gluckman PD *et al.* Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes* 60, 1528–1534 (2011).
- 31 Relton CL, Groom A, St. Pourcain B *et al.* DNA methylation patterns in cord blood DNA and body size in childhood. *PLoS ONE* 7, e31821 (2012).
- **Study on DNA methylation levels at birth associated with body size later in childhood. They observed associations between DNA methylation in the MPL gene region and lean mass.**
- 32 GitHub: Victorian-Bioinformatics-Consortium. <https://github.com>
- 33 Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572 (2011).
- 34 Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012).
- 35 Jaffe AE, Murakami P, Lee H *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int. J. Epidemiol.* 41, 200–209 (2012).
- 36 Hou C-D. A simple approximation for the distribution of the weighted combination of non-independent or independent probabilities. *Stat. Probab. Lett.* 73, 179–187 (2005).
- 37 Aryee MJ, Jaffe AE, Corrada-Bravo H *et al.* Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30, 1363–1369 (2014)
- 38 Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv* doi:10.1101/005165 (2014) (Epub ahead of print).
- 39 Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene

- set enrichment analysis to genome-wide association study. *Nucleic Acids Res.* 38, 90–95 (2010).
- 40 Iwen KAH, Senyaman O, Schwartz A *et al.* Melanocortin crosstalk with adipose functions: ACTH directly induces insulin resistance, promotes a pro-inflammatory adipokine profile and stimulates UCP-1 in adipocytes. *J. Endocrinol.* 196, 465–472 (2004).
- 41 Li W, Jiao H, Wang K *et al.* A genome wide association study of plasma uric acid levels in obese cases and never-overweight controls. *Obesity* 21, 490–494 (2013).
- 42 Turcot V, Groom A, McConnell JC *et al.* Bioinformatic selection of putative epigenetically regulated loci associated with obesity using gene expression data. *Gene* 499, 99–107 (2012).
- 43 Zhang Z, Hao C, Li C *et al.* Mutation of *SLC35D3* causes metabolic syndrome by impairing dopamine signaling in striatal D1 neurons. *PLoS Genet.* 10, e1004124 (2014).
- 44 Schmidt AM, Moore KJ. The semaphorin 3E/PlexinD1 axis regulates macrophage inflammation in obesity. *Cell Metab.* 18, 461–462 (2013).
- 45 Hoeflich A, Russo VC. Physiology and pathophysiology of IGFBP-1 and IGFBP-2 – consensus and dissent on metabolic control and malignant potential. *Best Pract. Res. Clin. Endocrinol. Metab.* 29, 685–700 (2015).
- 46 Gauthier BR, Wollheim CB. Synaptotagmins bind calcium to release insulin. *Am. J. Physiol. Endocrinol. Metab.* 295, E1279–E1286 (2008).
- 47 Do DN, Strathe AB, Ostersen T, Jensen J, Mark T. Genome-wide association study reveals genetic architecture of eating behavior in pigs and its implications for humans obesity by comparative mapping. *PLoS ONE* 8, e71509 (2013).
- 48 Girardet C, Butler AA. Neural melanocortin receptors in obesity and related metabolic disorders. *Biochim. Biophys. Acta* 1842, 482–494 (2014).
- 49 Chakraborty A, Koldobskiy MA, Bello NT *et al.* Inositol pyrophosphates inhibit akt signaling, thereby regulating insulin sensitivity and weight gain. *Cell* 143, 897–910 (2010).
- 50 Lee M, Kim MK, Kim S, Park H, Park C, Park HK. Gender-based differences on the association between salt-sensitive genes and obesity in Korean children aged between 8 and 9 years. *PLoS ONE* 10, e0120111 (2015).
- 51 Hoyo C, Fortner K, Murtha AP *et al.* Association of cord blood methylation fractions at imprinted and birth weight. *Cancer Causes Control* 2, 635–645 (2012).
- 52 Cordova-Palomera A, Alemany S, Fatjo M, Nenadic I, Fan L, Gonza A. Birth weight, working memory and epigenetic signatures in *IGF2* and related genes: a MZ twin study. *PLoS ONE* 9, e103639 (2014).
- **Shows a link between DNA methylation levels of two CpG sites in the *IGF2BP1* gene and birth weight.**
- 53 Martínez JA, Milagro FI, Claycombe KJ, Schalinske KL. Epigenetics in adipose tissue, obesity, weight loss, and diabetes. *Adv. Nutr.* 5, 71–81 (2014).
- 54 Movérare-Skrtic S, Mellström D, Vandenput L, Ehrlich M, Ohlsson C. Peripheral blood leukocyte distribution and body mass index are associated with the methylation pattern of the androgen receptor promoter. *Endocrine* 35, 204–210 (2009).
- 55 Hidalgo B, Irvin MR, Sha J *et al.* Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the genetics of lipid lowering drugs and diet network study. *Diabetes* 63, 801–807 (2014).
- 56 Breton C V, Siegmund KD, Joubert BR *et al.* Prenatal tobacco smoke exposure is associated with childhood DNA CpG methylation. *PLoS ONE* 9, e99716 (2009).
- 57 Besingi W, Johansson Å. Smoke-related DNA methylation changes in the etiology of human disease. *Hum. Mol. Genet.* 23, 2290–2297 (2014).
- 58 Stueve TR, Wolff MS, Pajak A, Teitelbaum SL, Chen J. CYP19A1 promoter methylation in saliva associated with milestones of pubertal timing in urban girls. *BMC Pediatr.* 14, 78 (2009).
- 59 Lomniczi A, Loche A, Castellano JM *et al.* Epigenetic control of female puberty. *Nat. Neurosci.* 16, 281–289 (2013).
- 60 Freedman DS, Wang J, Maynard LM *et al.* Relation of BMI to fat and fat-free mass among children and adolescents. *Int. J. Obes.* 29, 1–8 (2004).
- 61 Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13, 86 (2012).
- 62 Zou J, Lippert C, Heckerman D, Aryee M, Listgarten J. Epigenome-wide association studies without the need for cell-type composition. *Nat. Methods* 11, 309–311 (2014).