



Published in final edited form as:

Int J Obes (Lond). 2017 January ; 41(1): 194–199. doi:10.1038/ijo.2016.184.

An epigenome-wide study of body mass index and DNA methylation in blood using participants from the Sister Study cohort

Lauren E. Wilson¹, Sophia Harlid^{1,2}, Zongli Xu¹, Dale P. Sandler¹, and Jack A. Taylor¹

¹Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina, United States of America

²Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden

Abstract

Background/Objectives—The relationship between obesity and chronic disease risk is well-established; the underlying biological mechanisms driving this risk increase may include obesity-related epigenetic modifications. To explore this hypothesis, we conducted a genome-wide analysis of DNA methylation and body mass index (BMI) using data from a subset of women in the Sister Study.

Subjects/Methods—The Sister Study is a cohort of 50,884 U.S. women who had a sister with breast cancer but were free of breast cancer themselves at enrollment. Study participants completed examinations which included measurements of height and weight, and provided blood samples. Blood DNA methylation data generated with the Illumina Infinium HumanMethylation27 BeadChip array covering 27,589 CpG sites was available for 871 women from a prior study of breast cancer and DNA methylation. To identify differentially methylated CpG sites associated with body mass index, we analyzed this methylation data using robust linear regression with adjustment for age and case status. For those CpGs passing the false discovery rate significance level, we examined the association in a replication set comprised of a non-overlapping group of 187 women from the Sister Study who had DNA methylation data generated using the Infinium HumanMethylation450 BeadChip array. Analysis of this expanded 450K array identified additional BMI-associated sites which were investigated with targeted pyrosequencing.

Results—Four CpG sites reached genome-wide significance (FDR $q < 0.05$) in the discovery set and associations for all four were significant at strict Bonferroni correction in the replication set. An additional 23 sites passed FDR in the replication set and 5 were replicated by pyrosequencing

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Author Contact: Jack A. Taylor, P.O. Box 12233, Mail Drop A3-05, Research Triangle Park, NC 27709, Tel (919) 541-4631, Fax (301) 480-3290, taylor@niehs.nih.gov.

Conflict of Interest

The authors declare no conflict of interest.

Authorship

All authors contributed to the design of the study, the generation, analysis, and interpretation of results, and the drafting and revision of the manuscript. All authors have reviewed and approved the final manuscript. LEW and JAT confirm they had full access to the data and final responsibility for the decision to submit for publication.

in the discovery set. Several of the genes identified including *ANGPT4*, *RORC*, *SOCS3*, *FSD2*, *XYLT1*, *ABCG1*, *STK39*, *ASB2*, and *CRHR2* have been linked to obesity and obesity-related chronic diseases.

Conclusions—Our findings support the hypothesis that obesity-related epigenetic differences are detectable in blood and may be related to risk of chronic disease.

INTRODUCTION

Obesity is a well-established risk factor for chronic diseases including cancer, metabolic disease, and Type 2 diabetes. High body mass index has some degree of heritability, but is also influenced by lifestyle and environment (1-4). High body mass index and excess adiposity likely contribute to disease risk through a variety of biological pathways, which may include epigenetic processes such as alterations in DNA methylation patterns (5, 6).

There is a growing body of evidence supporting links between DNA methylation and obesity. A recently published study of adipose tissue collected from a small number of monozygotic twins found evidence of differential methylation between lean and overweight/obese twins at multiple genes(7). Other studies have identified differentially methylated CpGs in adipose tissue between lean and obese individuals, including genes which were also differentially expressed in the adipose tissue (8-10). Because adipose tissue is difficult to obtain in large population studies, the few existing large epigenome-wide association studies (EWAS) have used blood DNA to identify potential methylation differences related to body mass index. These studies have identified more than 40 CpG sites showing an association between methylation levels and body mass index in mixed cohorts of men and women (10, 11); obesity-related differences in DNA methylation may help identify key biological pathways involved in disease pathogenesis. Using a nationwide cohort study, we have previously reported the association between obesity and risk of breast cancer (12). Here, we extend that work using a subset of the subjects to examine the association between body-mass index and blood DNA methylation.

MATERIALS AND METHODS

Study Population

Study participants were drawn from the Sister Study cohort, a prospective volunteer cohort of women from the United States and Puerto Rico. From 2003-2009, women were recruited into the Sister Study if they had a sister with breast cancer but had not been diagnosed with breast cancer themselves. They complete annual, biennial, and triennial updates on cancer and various other lifestyle, exposure, and health factors. Participants have a home visit from a trained examiner at study baseline which includes blood collection and measurements of height and weight. The study was approved by the Institutional Review Board of the National Institute of Environmental Health Sciences, NIH and the Copernicus Group Institutional Review Board. Inclusion in this analysis is covered by the informed consent form all included cohort participants agreed to and signed. Due to ethical restrictions (participant confidentiality), data are available upon request. Permission to access Sister Study data may be obtained at www.sisterstudystars.org or by contacting the authors.

DNA methylation array analysis

We examined the relationship between methylation and body mass index using two DNA methylation datasets available within the Sister Study. The “discovery set” was comprised of 871 white, non-Hispanic women with methylation array data available on 27,589 CpG (cytosine-phosphate-guanine) sites from a nested case-cohort study of breast cancer. A second, smaller “replication set” was comprised of 187 white, non-Hispanic women with methylation array data on 485,512 CpG sites from a nested case-control study of diethylstilbestrol (DES) exposure. Study populations and details of methods for methylation measurements have been previously described (12-14).

Statistical Analysis

To examine the association between body mass index and DNA methylation in the discovery and replication sets, we used robust linear regression modeling. To correct for multiple testing the false discovery rate (FDR) was set at $q < 0.05$ as previously described (12, 14); specific CpGs were examined in the replication set for the same direction of association and an unadjusted $p < 0.05$. Data pre-processing, normalization methods and quality control measures are detailed in the supplementary material of the previously published papers (12, 14). Briefly, methylation intensity values were background-corrected using the Robust Multichip Average (RMA) method (15) and quantile-normalized across arrays. Methylation array plates included laboratory controls with known methylation levels to assess precision of measurement, and duplicate samples to assess reproducibility of results within the assay. Each methylation array included probes to assess bisulfite conversion efficiency and negative control probes to measure background fluorescent intensity. Samples with poor bisulfite conversion efficiency ($< 3,800$) or having $>5\%$ of probes with unreliable measures (detection $P > 0.05$) were excluded ($N=3$ probes in the 27k discovery set and $N=491$ probes in the 450k replication set). The methylation outcome was estimated using beta values, which are calculated using fluorescence intensities for unmethylated (U) and methylated (M) alleles as $M/(M+U+100)$. Beta values range from 0 (completely unmethylated) to 1 (100% methylated). In the 450K discovery set, we also excluded CpG probes with single nucleotide polymorphisms (SNPs) present at target sites or that mapped to multiple genomic regions ($N=48,158$), Illumina-designed SNP probes ($N=65$), and CpG sites on the X and Y chromosomes ($N=10,257$). In the 27K discovery data set we examined 27,575 probes and in the 450K replication data set we examined 426,606 probes. Singular value decomposition (SVD) analysis of the raw dataset revealed that the top principal components derived from the methylation beta value matrix were highly correlated with plate, bisulfite conversion intensities and age. We adjusted for these factors together with breast cancer status in all linear regression association analyses. All association tests were also adjusted for the proportions of different types of white blood cells estimated using a method described by Houseman et al (16, 17).

Pyrosequencing Validation

Although the sample size in the replication set was much smaller ($N=187$), the 450K array used for this set provides methylation data at approximately 320,000 additional CpG sites not covered by the 27K array used in the discovery set. We selected five of the top BMI-

associated CpGs identified from 450K analysis of the replication set for independent validation by pyrosequencing in samples from the discovery set. Pyrosequencing assays for 5 CpGs cg17501210 (RPS6KA2), cg06500161 (ABCG1), cg07728579 (FSD2), cg11775828 (STK39), and cg13134297 (CRHR2) were designed using Pyromark Assay Design version 2.0.2.15 (Qiagen, Valencia, CA). Primer sequences are detailed in Supplemental Table 1. Reaction mixtures (25 μ L) containing 100ng of bisulfite converted DNA, 5 pmol of each primer (forward and reverse) polymerase chain reaction (PCR) buffer (Invitrogen), 3 mM MgCl₂, 1 mM dNTP, and 0.8 units of taq polymerase (Invitrogen), were heated to 95°C for 15 min, followed by 45 PCR cycles (95°C for 20 sec, 55°C for 20 sec and 72°C for 20 sec) with a final extension at 72°C for 5 min. After PCR, the biotin-labeled PCR product was hybridized to streptavidin-coated sepharose beads (GE Healthcare, Madison, WI) and denatured in 0.2 M sodium hydroxide to provide a single-stranded sequencing template. Pyrosequencing primers (0.3 μ mol/L) were annealed to the single-stranded template and the pyrosequencing was carried out using PyroMark Q96 MD System (Qiagen) according to the manufacturer's instructions. The percentage methylation was quantified using the Pyro Q-CpG Software (Qiagen). Associations between methylation levels and BMI were examined using linear regression analysis of the methylation percentages and BMI with adjustment for age and breast cancer status.

RESULTS

Body Mass Index in the Sister Study DNA methylation datasets

Basic characteristics of Sister Study participants with methylation array data are summarized in Table 1. As the discovery set includes a substantial number of women who later developed breast cancer, characteristics are stratified by breast cancer case status. The mean BMI in both the discovery subset (27.1 kg/m², SD 5.7 kg/m²) and replication subset (26.7 kg/m², SD 5.6 kg/m²) was slightly above the CDC's "healthy BMI" upper boundary. In both the discovery and replication set, at least 50% of participants fell into the overweight (BMI 25.0-29.9) or obese (BMI 30+) categories. The replication set had a larger proportion of women in the healthy BMI category (49%) than the discovery set (42%). 7% of the discovery set and 6% of the replication set reported they were current smokers, and slightly more women in the discovery set had at least one pregnancy (87%) compared to the replication set (81%).

Analysis of BMI and DNA Methylation in discovery set

In an analysis limited to women in the discovery set who had not developed breast cancer (N=571), BMI was associated with methylation at two CpG sites located in the genes *LGALS3BP* (q-value=0.02) and *ANGPT4* (q-value=0.05) (Table 2). When the analysis was expanded to include women who developed breast cancer during study follow-up with adjustment for case status (N=871), the same two sites remained associated (cg14870271 and cg03218374) and two additional associated CpG sites were identified in the genes *RORC* and *SOCS3* (cg18149207 at q-value=0.02 and cg27637521 at q-value=0.02). All four of these CpG sites are present on the Infinium450K array used in the replication study.

Analysis of BMI and DNA Methylation in replication set

The association between BMI and DNA methylation at these four CpGs was examined in the replication sample (N=187); p-values for all four associations passed a strict Bonferroni correction for multiple testing ($p < 0.0125$) as shown in Table 2. The CpG sites located in *LGALS3BP* and *RORC* showed an approximately 0.1% increase in methylation with each increasing unit (1 kg/m²) of BMI, and the site in *ANGPT4* showed 0.2% increase in methylation per 1 unit of BMI. These differences would correspond to approximately 1% higher methylation at these sites between individuals at normal weight (BMI=20) versus obese individuals (BMI=30). The CpG located in *SOCS3* showed an association in the opposite direction, with 0.05% decrease in methylation with each increasing unit of BMI.

A 450K array-wide analysis of DNA methylation and BMI in the replication set women revealed associations ($q < 0.05$) for 23 CpG sites (Table 3). Three of these 23 sites were covered on the 27K array; one of the three, cg21282997 located in the gene *IL18RAP*, showed an association with BMI at $p = 0.02$ in the discovery set. The remaining 20 CpG sites were not covered on the 27K array; from these 20 we selected five of the top CpG sites for pyrosequencing analysis to determine methylation status in participants from the discovery set (Table 3).

Results of pyrosequencing validation

Pyrosequencing analysis of 871 DNA samples from the discovery set validated the associations between body mass index and methylation at the five selected CpG sites (Table 4). Point estimates for the associations between BMI and methylation at all five CpGs tested (cg17501210; cg06500161; cg07728579; cg11775828; and cg13134297) generated using pyrosequencing passed a Bonferroni-corrected p-value threshold of 0.01 and both the magnitude and direction of effect estimates were similar between array and pyrosequencing data. Pyrosequencing also provided information on CpGs flanking the sites of interest for cg17501210 (*RPS6KA2*) and cg06500161 (*ABCG1*), and these flanking CpGs had findings similar to that of the target CpG (Table 4).

Replication of findings from published studies of BMI and genome-wide DNA methylation

Recent epigenome-wide association studies have used the Infinium HumanMethylation450 BeadChip to investigate body mass index and peripheral blood DNA methylation data. We examined whether differentially methylated CpGs reported in those studies were differentially methylated in our replication set data (Supplemental Table 2). A study from the Cardiogenics Consortium identified 5 CpGs associated with BMI in their discovery cohort of men and women and replicated three of these associations (18). Our data provides additional support for their observed associations in a women-only cohort with BMI at two CpG sites: one in *KLF13* (cg07814318 at $p = 0.0003$) and one in *KIAA0664/CLUH* (cg09664445 at $p = 0.04$). A second study utilizing the Atherosclerosis Risk in Communities cohort and multiple replication cohorts identified 76 CpGs that were differentially associated with body mass index and replicated 37 sites (10). 18 of these associations were recently replicated in an Arab population (19) and we found that 27 of these associations replicated in our dataset, including the site cg06500161 in *ABCG1* identified in our discovery set. We

provide the first replication for associations with BMI at the CpG sites cg09554443 in *CD247*, cg04986899 in *XYLT1*, cg05242915 on chromosome 19, and cg03562528 in *ASB2*.

DISCUSSION

Using 27K, 450K, and pyrosequencing data on 1,066 women from the Sister Study cohort, we identified and replicated eight previously unreported associations between body mass index and CpG site DNA methylation and confirm a previously reported and replicated association at cg06500161 in *ABCG1*. Our results also provide support for association at 27 previously reported additional CpG sites and provide the first replication for associations with BMI at four of these sites.

We identified and confirmed methylation differences associated with body mass index at CpG sites in the genes *LGALS3BP*, *ANGPT4*, *RORC*, *SOCS3*, *RSPK6A2*, *FSD2*, *ABCG1*, *STK39*, and *CRHR2*. Several of the genes with differentially methylated sites in our study have been linked to obesity and obesity-related chronic diseases. The decreased methylation with increasing BMI observed at cg27637521 is of particular interest as it is 25 bp from the transcriptional start site of *SOCS3* (suppressor of cytokine signaling 3). The two flanking CpGs in *SOCS3* (cg10508317 and cg10279487) also show decreased methylation with increasing BMI (p 0.003). Animal models indicate that *SOCS3* expression inhibits leptin signaling and is likely involved in the decreased leptin sensitivity observed in obese individuals (20-24). *SOCS3* is a negative regulator of cytokine signaling; expression is induced by various cytokines, including IL-6, IL-10, and interferon (IFN)-gamma (25-27). Genetic polymorphisms in and near *SOCS3* are associated with obesity in human population studies (28-30). In cohort studies of Indian Asians and Europeans, decreased DNA methylation in blood at another CpG (cg18181703) in the coding region of *SOCS3* has been linked to risk of Type 2 (T2) diabetes. (31).

We also identified a differentially methylated region in the gene *CRHR2* (corticotrophin-releasing hormone receptor 2); the CpG site we identified (cg13134297) is in the first intron of this gene. As with *SOCS3*, examination of flanking CpGs on the 450K array (cg23068772 and cg22007110) shows associations with BMI (p 0.002) with the same directionality. *CRHR2* is involved in corticotrophin-releasing hormone signaling and likely plays a role in coordinating endocrine and autonomic responses to stress (32). *STK39* (Serine Threonine Kinase 39) also plays a role in cellular responses to stress; we identified and confirmed increasing methylation with increasing body mass index at this gene at cg11775828. Flanking CpGs on the 450K array (cg21899461 and cg03331300) were also associated with BMI at p<0.05. Polymorphisms in this gene have been linked to risk of hypertension (33-35). Epigenetic silencing of *STK39* through DNA hypermethylation in B-cell lymphoma appears to promote cancer progression by inhibiting apoptosis in cells with DNA damage (36).

The differentially methylated CpGs we identified in the genes *RORC* and *FSD2* may also be promising targets for future studies of obesity and development of cardiovascular and metabolic diseases. *RORC* (RAR-Related Orphan Receptor C) regulates production of the inflammatory cytokine IL-17 by T helper-17 (Th17) cells; Th17 cell activation and cytokine

production appears to play a critical role in the pathogenesis of diabetes (37-41). Additionally, genetic variation in *RORC* has been linked to body fat composition levels in cattle (42). SNPs in *FSD2* (Fibronectin Type III And SPRY Domain Containing 2) have been linked to the presence and volume of carotid plaque in a Caribbean Hispanic population (43).

We provided the first replication of associations with BMI at the CpGs cg09554443 in *CD247*, cg04986899 in *XYLT1*, cg05242915 on chromosome 19, and cg03562528 in *ASB2*. *CD247* (T-cell surface glycoprotein CD3 zeta chain) plays an important role in antigen recognition in the immune response and shows altered expression patterns in peripheral blood in individuals with Type 2 Diabetes (44). *XYLT1* (Xylosyltransferase 1) is necessary for biosynthesis of glycosaminoglycan chains. Genetic variations in this gene can increase risk of abdominal aortic aneurysm (45). *ASB2* (Ankyrin repeat and SOCS box protein 2) is differentially methylated in atherosclerotic aorta tissue compared to healthy aortic tissue (46).

The observed association between adult BMI and methylation at cg03218374 in the Angiopoetin-4 (*ANGPT4*) gene is consistent with an existing study that links methylation at this same CpG in cord blood to infant birth weight (47), indicating that this association between body size and methylation is consistent in both infants and adults. This CpG is located in the promoter region of *ANGPT4*, 18bp upstream of the transcription start site. *ANGPT4* plays an important role in vascular angiogenesis, and may help to mediate lipid breakdown (48). We have also confirmed the association between BMI and methylation at *ABCG1* (cg06500161) observed in previous studies (10, 11). Methylation of multiple CpG sites in *ABCG1* is linked to individual variability in blood lipid levels (49). In our pyrosequencing validation analysis of *RPS6KA2* and *ABCG1*, our primers covered additional CpG sites (2 sites in *RPS6KA2* and 1 in *ABCG1*) that were not covered on the Infinium450K array. These additional CpG sites exhibited the same associations with BMI as the adjacent sites on the array (Table 4). Both sites in *ABCG1*, 1 bp apart, and all three sites in *RPS6KA2* (spanning a 9 bp region) show increasing methylation with increasing body mass index.

We have identified and validated associations between body mass index and methylation at nine CpG sites using a combination of array-based and pyrosequencing methylation measures, and provided further confirmation for an additional 27 associations reported in previous studies. A potential limitation of our study is the inclusion of DES-exposed women in our replication set, which might potentially affect our results if DES also affects methylation patterns. However, previous analysis of these participants' DES exposure status and the DNA methylation array data showed no detectable associations between DES exposure and DNA methylation (14). While this study's cross-sectional design limits our ability to determine whether these methylation differences are caused by obesity, it has the advantage of large number of participants with both DNA methylation data and examiner-measured height and weight.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We wish to thank Dr. Sarah Reese and Dr. Yong-Moon Park for their critical review of this manuscript. We would also like to thank the women who volunteered to participate in the Sister Study. We thank the National Institutes of Health (NIH) Center for Inherited Disease Research and the National Institute of Environmental Health Sciences (NIEHS) Molecular Genetics Core and Microarray Core for their technical support. This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01 ES044005, Z01 ES044032, and Z01 ES049033) which provided funding for design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, and approval of the manuscript.

References

1. Bouchard C, Tremblay A. Genetic influences on the response of body fat and fat distribution to positive and negative energy balances in human identical twins. *J Nutr.* May; 1997 127(5 Suppl): 943S–7S. PubMed PMID: 9164270. Epub 1997/05/01. eng. [PubMed: 9164270]
2. Albuquerque D, Stice E, Rodriguez-Lopez R, Manco L, Nobrega C. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. *Molecular genetics and genomics : MGG.* Aug; 2015 290(4):1191–221. PubMed PMID: 25749980. Epub 2015/03/10. eng. [PubMed: 25749980]
3. Hemani G, Yang J, Vinkhuyzen A, Powell JE, Willemsen G, Hottenga JJ, et al. Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. *Am J Hum Genet.* Nov 7; 2013 93(5):865–75. PubMed PMID: 24183453. Pubmed Central PMCID: PMC3965855. Epub 2013/11/05. eng. [PubMed: 24183453]
4. Arabshahi S, Ibiebele TI, Hughes MC, Lahmann PH, Williams GM, van der Pols JC. Dietary patterns and weight change: 15-year longitudinal study in Australian adults. *European journal of nutrition.* Feb 26.2016 PubMed PMID: 26919993. Epub 2016/02/28. Eng.
5. Martinez JA, Milagro FI, Claycombe KJ, Schalinske KL. Epigenetics in adipose tissue, obesity, weight loss, and diabetes. *Advances in nutrition (Bethesda, Md).* Jan; 2014 5(1):71–81. PubMed PMID: 24425725. Pubmed Central PMCID: PMC3884103. Epub 2014/01/16. eng.
6. Garcia-Jimenez C, Gutierrez-Salmeron M, Chocarro-Calvo A, Garcia-Martinez JM, Castano A, De la Vieja A. From obesity to diabetes and cancer: epidemiological links and role of therapies. *Br J Cancer.* Feb 23.2016 PubMed PMID: 26908326. Epub 2016/02/26. Eng.
7. Pietilainen KH, Ismail K, Jarvinen E, Heinonen S, Tummars M, Bollepalli S, et al. DNA methylation and gene expression patterns in adipose tissue differ significantly within young adult monozygotic BMI-discordant twin pairs. *International journal of obesity (2005).* Oct 26.2015 PubMed PMID: 26499446. Epub 2015/10/27. Eng.
8. Arner P, Sinha I, Thorell A, Ryden M, Dahlman-Wright K, Dahlman I. The epigenetic signature of subcutaneous fat cells is linked to altered expression of genes implicated in lipid metabolism in obese women. *Clinical epigenetics.* 2015; 7(1):93. PubMed PMID: 26351548. Pubmed Central PMCID: PMC4562340. Epub 2015/09/10. eng. [PubMed: 26351548]
9. Gillberg L, Perfilyev A, Brons C, Thomasen M, Grunnet LG, Volkov P, et al. Adipose tissue transcriptomics and epigenomics in low birthweight men and controls: role of high-fat overfeeding. *Diabetologia.* Jan 11.2016 PubMed PMID: 26750116. Epub 2016/01/12. Eng.
10. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, 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.* Aug 1; 2015 24(15):4464–79. PubMed PMID: 25935004. Pubmed Central PMCID: PMC4492394. Epub 2015/05/03. eng. [PubMed: 25935004]
11. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference.

- Obesity (Silver Spring, Md). Jul; 2015 23(7):1493–501. PubMed PMID: 26110892. Pubmed Central PMCID: PMC4482015. Epub 2015/06/26. eng.
12. Xu Z, Bolick SC, DeRoo LA, Weinberg CR, Sandler DP, Taylor JA. Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *J Natl Cancer Inst.* May 15; 2013 105(10):694–700. PubMed PMID: 23578854. Pubmed Central PMCID: PMC3653821. Epub 2013/04/13. eng. [PubMed: 23578854]
 13. Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the sister study. *Environmental health perspectives.* Jul; 2014 122(7):673–8. PubMed PMID: 24704585. Pubmed Central PMCID: PMC4080519. Epub 2014/04/08. eng. [PubMed: 24704585]
 14. Harlid S, Xu Z, Panduri V, D'Aloisio AA, DeRoo LA, Sandler DP, et al. In utero exposure to diethylstilbestrol and blood DNA methylation in women ages 40-59 years from the sister study. *PLoS One.* 2015; 10(3):e0118757. PubMed PMID: 25751399. Pubmed Central PMCID: PMC4353728. Epub 2015/03/10. eng. [PubMed: 25751399]
 15. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England).* Apr; 2003 4(2):249–64. PubMed PMID: 12925520. Epub 2003/08/20. eng.
 16. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics.* 2012; 13:86. PubMed PMID: 22568884. Pubmed Central PMCID: PMC3532182. Epub 2012/05/10. eng. [PubMed: 22568884]
 17. Accomando WP, Wiencke JK, Houseman EA, Butler RA, Zheng S, Nelson HH, et al. Decreased NK cells in patients with head and neck cancer determined in archival DNA. *Clinical cancer research : an official journal of the American Association for Cancer Research.* Nov 15; 2012 18(22):6147–54. PubMed PMID: 23014525. Pubmed Central PMCID: PMC3500449. Epub 2012/09/28. eng. [PubMed: 23014525]
 18. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet.* Jun 7; 2014 383(9933):1990–8. PubMed PMID: 24630777. Epub 2014/03/19. eng. [PubMed: 24630777]
 19. Al Muftah WA, Al-Shafai M, Zaghlool SB, Visconti A, Tsai P-C, Kumar P, et al. Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clinical epigenetics.* 2016 01/28 12/10/received 01/18/accepted;8:13. PubMed PMID: PMC4730771.
 20. Bjørnbæk C, Elmquist JK, Frantz JD, Shoelson SE, Flier JS. Identification of SOCS-3 as a Potential Mediator of Central Leptin Resistance. *Molecular Cell.* 1998; 1(4):619–25. 3//. [PubMed: 9660946]
 21. Pedroso JAB, Buonfiglio DC, Cardinali LI, Furigo IC, Ramos-Lobo AM, Tirapegui J, et al. Inactivation of SOCS3 in leptin receptor-expressing cells protects mice from diet-induced insulin resistance but does not prevent obesity. *Molecular Metabolism.* 2014; 3(6):608–18. 9//. [PubMed: 25161884]
 22. Howard JK, Flier JS. Attenuation of leptin and insulin signaling by SOCS proteins. *Trends in Endocrinology & Metabolism.* 2006; 17(9):365–71. 11//. [PubMed: 17010638]
 23. Mori H, Hanada R, Hanada T, Aki D, Mashima R, Nishinakamura H, et al. Socs3 deficiency in the brain elevates leptin sensitivity and confers resistance to diet-induced obesity. *Nature Medicine.* 2004; 10(7):739–43.
 24. Howard JK, Cave BJ, Oksanen LJ, Tzamelis I, Bjørnbæk C, Flier JS. Enhanced leptin sensitivity and attenuation of diet-induced obesity in mice with haploinsufficiency of Socs3. *Nature Medicine.* 2004; 10(7):734–8.
 25. Alexander WS, Hilton DJ. The role of suppressors of cytokine signaling (SOCS) proteins in regulation of the immune response. *Annual review of immunology.* 2004; 22:503–29. PubMed PMID: 15032587. Epub 2004/03/23. eng.
 26. Cohney SJ, Sanden D, Cacalano NA, Yoshimura A, Mui A, Migone TS, et al. SOCS-3 is tyrosine phosphorylated in response to interleukin-2 and suppresses STAT5 phosphorylation and lymphocyte proliferation. *Molecular and cellular biology.* Jul; 1999 19(7):4980–8. PubMed PMID: 10373548. Pubmed Central PMCID: PMC84319. Epub 1999/06/22. eng. [PubMed: 10373548]

27. Dagvadorj J, Naiki Y, Tumurkhuu G, Shadat Mohammad Noman A, Iftakhar-E-Khuda I, Komatsu T, et al. Tumor necrosis factor- α augments lipopolysaccharide-induced suppressor of cytokine signalling 3 (SOCS-3) protein expression by preventing the degradation. *Immunology*. 2010; 129(1):97–104. [PubMed: 20050332]
28. Li P, Tiwari HK, Lin WY, Allison DB, Chung WK, Leibel RL, et al. Genetic association analysis of 30 genes related to obesity in a European American population. *International journal of obesity* (2005). May; 2014 38(5):724–9. PubMed PMID: 23900445. Pubmed Central PMCID: PMC3909018. Epub 2013/08/01. eng. [PubMed: 23900445]
29. Tang W, Zou JJ, Chen XF, Zheng JY, Zeng HZ, Liu ZM, et al. Association of two polymorphisms within and near SOCS3 gene with obesity in three nationalities in Xinjiang province of China. *Acta pharmacologica Sinica*. Nov; 2011 32(11):1381–6. PubMed PMID: 21892197. Pubmed Central PMCID: PMC4002737. Epub 2011/09/06. eng. [PubMed: 21892197]
30. Talbert ME, Langefeld CD, Ziegler J, Mychaleckyj JC, Haffner SM, Norris JM, et al. Polymorphisms near SOCS3 are associated with obesity and glucose homeostasis traits in Hispanic Americans from the Insulin Resistance Atherosclerosis Family Study. *Hum Genet*. Mar; 2009 125(2):153–62. PubMed PMID: 19083014. Pubmed Central PMCID: PMC2804661. Epub 2008/12/17. eng. [PubMed: 19083014]
31. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *The lancet Diabetes & endocrinology*. Jul; 2015 3(7): 526–34. PubMed PMID: 26095709. Epub 2015/06/23. eng. [PubMed: 26095709]
32. Grammatopoulos DK. Insights into mechanisms of corticotropin-releasing hormone receptor signal transduction. *British journal of pharmacology*. May; 2012 166(1):85–97. PubMed PMID: 21883143. Pubmed Central PMCID: PMC3415640. Epub 2011/09/03. eng. [PubMed: 21883143]
33. Persu A, Evenepoel L, Jin Y, Mendola A, Staessen JA, Viskula M. 6C.05: STK39 AND WNK1 ARE POTENTIAL HYPERTENSION SUSCEPTIBILITY GENES IN THE BELHYPGEN COHORT. *Journal of hypertension*. Jun.2015 33(Suppl 1):e80. PubMed PMID: 26102938. Epub 2015/06/24. eng.
34. Maatta KM, Nikkari ST, Lahteela KH, Palmroos PB, Kunnas TA. A functional variant in the serine-threonine kinase coding gene is associated with hypertension: a case-control study in a Finnish population, the Tampere adult population cardiovascular risk study. *Journal of hypertension*. Mar; 2013 31(3):516–20. discussion 20. PubMed PMID: 23235358. Epub 2012/12/14. eng. [PubMed: 23235358]
35. Li B, Yang M, Liu JW. Expression of STK39 in peripheral blood of hypertension patients and the relationship between its genetic polymorphism and blood pressure. *Genetics and molecular research : GMR*. 2015; 14(4):16461–8. PubMed PMID: 26662444. Epub 2015/12/15. eng. [PubMed: 26662444]
36. Balatoni CE, Dawson DW, Suh J, Sherman MH, Sanders G, Hong JS, et al. Epigenetic silencing of Stk39 in B-cell lymphoma inhibits apoptosis from genotoxic stress. *The American journal of pathology*. Oct; 2009 175(4):1653–61. PubMed PMID: 19717643. Pubmed Central PMCID: PMC2751561. Epub 2009/09/01. eng. [PubMed: 19717643]
37. Endo Y, Asou HK, Matsugae N, Hirahara K, Shinoda K, Tumes DJ, et al. Obesity Drives Th17 Cell Differentiation by Inducing the Lipid Metabolic Kinase, ACC1. *Cell reports*. Aug 11; 2015 12(6): 1042–55. PubMed PMID: 26235623. Epub 2015/08/04. eng. [PubMed: 26235623]
38. Ip B, Cilfone NA, Belkina AC, DeFuria J, Jagannathan-Bogdan M, Zhu M, et al. Th17 cytokines differentiate obesity from obesity-associated type 2 diabetes and promote TNF α production. *Obesity (Silver Spring, Md)*. Jan; 2016 24(1):102–12. PubMed PMID: 26576827. Pubmed Central PMCID: PMC4688084. Epub 2015/11/19. eng.
39. Reinert-Hartwall L, Honkanen J, Salo HM, Nieminen JK, Luopajarvi K, Harkonen T, et al. Th1/Th17 plasticity is a marker of advanced beta cell autoimmunity and impaired glucose tolerance in humans. *J Immunol*. Jan 1; 2015 194(1):68–75. PubMed PMID: 25480564. Pubmed Central PMCID: PMC4273995. Epub 2014/12/07. eng. [PubMed: 25480564]
40. Romanowski M, Domanski L, Pawlik A, Osekowska B, Dziedziejko V, Safranow K, et al. Interleukin-17 gene polymorphisms in patients with post-transplant diabetes mellitus. *European*

- review for medical and pharmacological sciences. Sep; 2015 19(17):3152–6. PubMed PMID: 26400516. Epub 2015/09/25. eng. [PubMed: 26400516]
41. Bellemore SM, Nikoopour E, Krougly O, Lee-Chan E, Fouser LA, Singh B. Pathogenic T helper type 17 cells contribute to type 1 diabetes independently of interleukin-22. *Clinical and experimental immunology*. Mar; 2016 183(3):380–8. PubMed PMID: 26496462. Pubmed Central PMCID: PMC4750601. Epub 2015/10/27. eng. [PubMed: 26496462]
 42. Barendse W, Bunch RJ, Kijas JW, Thomas MB. The effect of genetic variation of the retinoic acid receptor-related orphan receptor C gene on fatness in cattle. *Genetics*. Feb; 2007 175(2):843–53. PubMed PMID: 17151246. Pubmed Central PMCID: PMC1800623. Epub 2006/12/08. eng. [PubMed: 17151246]
 43. Dong C, Beecham A, Slifer S, Wang L, Blanton SH, Wright CB, et al. Genomewide linkage and peakwise association analyses of carotid plaque in Caribbean Hispanics. *Stroke; a journal of cerebral circulation*. Dec; 2010 41(12):2750–6. PubMed PMID: 20966410. Pubmed Central PMCID: PMC3004531. Epub 2010/10/23. eng.
 44. Eldor R, Klieger Y, Sade-Feldman M, Vaknin I, Varfolomeev I, Fuchs C, et al. CD247, a novel T cell-derived diagnostic and prognostic biomarker for detecting disease progression and severity in patients with type 2 diabetes. *Diabetes care*. Jan; 2015 38(1):113–8. PubMed PMID: 25368105. Epub 2014/11/05. eng. [PubMed: 25368105]
 45. Gotting C, Prante C, Schillinger M, Exner M, Domanovits H, Raith M, et al. Xylosyltransferase I variants and their impact on abdominal aortic aneurysms. *Clinica chimica acta. international journal of clinical chemistry*. May; 2008 391(1-2):41–5. PubMed PMID: 18294457. Epub 2008/02/26. eng. [PubMed: 18294457]
 46. Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S, Condom E, et al. DNA methylation map of human atherosclerosis. *Circulation Cardiovascular genetics*. Oct; 2014 7(5):692–700. PubMed PMID: 25091541. Epub 2014/08/06. eng. [PubMed: 25091541]
 47. Turan N, Ghalwash MF, Katari S, Coutifaris C, Obradovic Z, Sapienza C. DNA methylation differences at growth related genes correlate with birth weight: a molecular signature linked to developmental origins of adult disease? *BMC medical genomics*. 2012; 5:10. PubMed PMID: 22498030. Pubmed Central PMCID: PMC3359247. Epub 2012/04/14. eng. [PubMed: 22498030]
 48. Gray NE, Lam LN, Yang K, Zhou AY, Koliwad S, Wang JC. Angiotensin-like 4 (Angptl4) protein is a physiological mediator of intracellular lipolysis in murine adipocytes. *J Biol Chem*. Mar 9; 2012 287(11):8444–56. PubMed PMID: 22267746. Pubmed Central PMCID: PMC3318686. Epub 2012/01/24. eng. [PubMed: 22267746]
 49. Pfeiffer L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, et al. DNA methylation of lipid-related genes affects blood lipid levels. *Circulation Cardiovascular genetics*. Apr; 2015 8(2):334–42. PubMed PMID: 25583993. Epub 2015/01/15. eng. [PubMed: 25583993]

Characteristics of participants in the discovery set (N=871) and replication set (N=187) stratified by breast cancer case status

Table 1

Characteristics	Discovery Set (N=871)		Replication Set (N=187)			
	All N (%)	Controls (N=573) N (%)	Developed breast cancer (N=298) N (%)	All N (%)	Controls (N=181) N (%)	Developed breast cancer (N=6) N (%)
Mean Age (SD)	55.2 (9.1)	54.7 (9.0)	56.1 (9.2)	50.5 (4.8)	50.4 (4.8)	53.3 (3.8)
Mean BMI (SD)	27.1 (5.7)	27.0 (5.5)	27.4 (5.9)	26.7 (5.6)	26.7 (5.8)	28.2 (9.0)
Categorized BMI						
<18.5 (Underweight)	9 (1)	4 (1)	5 (2)	0	0	0
18.5 – 24.9 (Normal)	362 (42)	241 (42)	121 (41)	92 (49)	88 (49)	4 (67)
25-29.9 (Overweight)	256 (29)	173 (30)	83 (28)	49 (27)	48 (27)	1 (17)
30+ (Obese)	244 (28)	155 (27)	89 (30)	45 (24)	44 (24)	1 (17)
Smoking Status						
Never	458 (53)	294 (51)	164 (55)	103 (55)	99 (55)	4 (67)
Social	20 (2)	15 (3)	5 (2)	5 (3)	5 (3)	0
Former	331 (38)	222 (39)	109 (37)	67 (36)	66 (37)	1 (17)
Current	62 (7)	42 (7)	20 (7)	11 (6)	10 (6)	1 (17)
Ever pregnant						
No	109 (13)	71 (12)	38 (13)	36 (19)	36 (20)	0
Yes	761 (87)	501 (88)	260 (87)	150 (81)	144 (80)	6 (100)

CpG sites with associations between body mass index and methylation passing the false discovery rate ($q < 0.05$) in the discovery set. Associations for these CpGs ($p < 0.05$) were also observed in the replication set.

Table 2

CpG Site	Participants free of breast cancer					
	Gene	Discovery set free of breast cancer (N=573)	Methylation change: normal to obese *	Replication set free of breast cancer (N=181)		
		Coefficient	P-Value	Coefficient	P-Value	
cg14870271	LGALS3BP	0.001623	1.61E-06	1.6%	0.001	0.01343
cg03218374	ANGPT4	0.001464	5.42E-06	1.4%	0.002	4.95E-06

All participants with adjustment for breast case status						
CpG Site	Gene	Discovery Set (N=871)		Replication Set (N=187)		
		Coefficient	P-Value	Methylation change: normal to obese *	Coefficient	P-Value
cg14870271	LGALS3BP	0.001251	2.26E-06	1.2%	0.001	0.010
cg18149207	RORC	0.000885	3.06E-06	0.8%	0.001	6.2E-04
cg27637521	SOC3	-0.00024	2.83E-06	-0.2%	-0.0005	0.001
cg03218374	ANGPT4	0.001141	8.74E-06	1.1%	0.002	1.11E-06

* Change in methylation associated with an increase in BMI from BMI=20 (classified as normal weight) to BMI=30 (classified as obese)

Table 3

CpG sites where associations between methylation and body mass index pass the false discovery rate on the 450K methylation array (N=181). Three of these CpGs were also covered on the 27k array (N=871). Of the nineteen CpGs not covered on the 27k discovery set array, 5 CpGs were selected for pyrosequencing analysis.

CpG Probe	Gene	Methylation Coefficient	Mean	FDR	P-value from discovery set array	Pyrosequencing
cg09349128 [*]	-	-0.003	0.35	1.95E-06	-	
cg17501210	RPS6KA2	-0.003	0.78	8.57E-06	-	selected
cg05176551	-	0.002	0.85	0.003	-	
cg07918509	ICAM2	-0.001	0.90	0.003	0.47	
cg06500161 [*]	ABCG1	0.002	0.62	0.006	-	selected
cg20117675	DHX16	-0.003	0.79	0.011	-	
cg07728579	FSD2	0.002	0.80	0.011	-	selected
cg12992827 [*]	-	-0.003	0.77	0.011	-	
cg26161057	AHDC1	-0.001	0.87	0.011	-	
cg11775828	STK39	0.003	0.43	0.013	-	selected
cg21282997	IL18RAP	0.003	0.36	0.014	0.02	
cg23679085	AP2M1	<0.001	0.03	0.016	0.63	
cg26899718	SMPD3	0.001	0.74	0.016	-	
cg08035749	-	0.001	0.87	0.018	-	
cg12760041	KANSL1L	<0.001	0.02	0.018	-	
cg17901584 ^{*a}	DHCR24	-0.004	0.53	0.019	-	
cg13134297	CRHR2	-0.002	0.24	0.025	-	selected
cg19818308	-	0.002	0.56	0.025	-	
cg25215047	LINC00371	<0.001	0.95	0.025	-	
cg17214023	-	-0.002	0.56	0.030	-	
cg11345976	LCLAT1	0.003	0.70	0.032	-	
cg23285465	FAM207A	<0.001	0.01	0.032	-	
cg07622521	ATOH8	-0.001	0.16	0.048	-	

^{*} Also reported in Demerath et al., 2015

^a Also reported in Ollikainen et al., 2015

Table 4

Top CpG hits associated with body mass index on the 450K array (N=181) which were not included on 27K array. Associations with BMI were confirmed by pyrosequencing using bisulfite-converted DNA left from the participants included on the 27k array (N=871)

CpG Site	Gene	450K Array (N=181)		Pyrosequencing confirmation	
		Coefficient	P-Value (or q)	Controls only	Breast cancer cases and controls
cg17501210	RPS6KA2	-0.003	4.39E-11	-0.003	<0.0001
cg06500161	ABCG1	0.002	7.75E-08	0.001	<0.0001
cg07728579	FSD2	0.002	2.66E-07	0.0006	0.0019
cg11775828	STK39	0.003	3.48E-07	0.001	<0.0001
cg13134297	CRHR2	-0.002	1.09E-06	-0.0005	0.0024
N/A *	RPS6KA2 (1bp upstream of cg17501210)	-	-	-0.004	<0.0001
N/A *	RPS6KA2 (2 bp downstream of cg17501210)	-	-	-0.003	<0.0001
N/A *	ABCG1 (1 bp downstream of cg06500161)	-	-	0.003	<0.0001

* Sites were not included on 450K BeadChip array and were captured only in the pyrosequencing assay