


RESEARCH

Open Access



An epigenome-wide study of obesity in African American youth and young adults: novel findings, replication in neutrophils, and relationship with gene expression

Xiaoling Wang^{1*} , Yue Pan¹, Haidong Zhu¹, Guang Hao¹, Yisong Huang¹, Vernon Barnes¹, Huidong Shi², Harold Snieder³, James Pankow⁴, Kari North⁵, Megan Grove⁶, Weihua Guan⁴, Ellen Demerath⁴, Yanbin Dong¹ and Shaoyong Su¹

Abstract

Background: We conducted an epigenome-wide association study (EWAS) on obesity in healthy youth and young adults and further examined to what extent identified signals influenced gene expression and were independent of cell type composition and obesity-related cardio-metabolic risk factors. Genome-wide DNA methylation data from leukocytes were obtained from 700 African Americans aged 14–36. We also measured genome-wide DNA methylation data from neutrophils as well as genome-wide gene expression data from leukocytes in a subset of samples ($n = 188$).

Results: The EWAS identified 76 obesity-related CpG sites in leukocytes with $p < 1 \times 10^{-7}$. In silico replication in the ARIC study of 2097 African Americans aged 47–70 validated 54 CpG sites. Out of the 54 CpG sites, 29 associations with obesity were novel and 37 were replicated in neutrophils. Fifty one CpG sites were associated with at least one cardio-metabolic risk factor; however, the number reduced to 9 after adjustment for obesity. Sixteen CpG sites were associated with expression of 17 genes in *cis*, of which 5 genes displayed differential expression between obese cases and lean controls. We also replicated 71.5% of obesity-related CpG sites previously reported.

Conclusion: In this study of youth and young adults, we identified 29 novel CpG sites associated with obesity and replicated the majority of the CpG sites previously identified. We further demonstrated that the majority of the obesity-related CpG sites in leukocytes were not driven by cell composition or obesity-related cardio-metabolic risk factors. We also provided the direct link between DNA methylation-gene expression-obesity for 5 genes.

Keywords: DNA methylation, Youth, African Americans, Obesity, Leukocytes, Neutrophils

Background

Obesity is a complex disease resulting from interactions between genes and environmental factors that can be modified and/or mediated by epigenetic changes. DNA methylation is a pivotal and stable epigenetic mechanism, and DNA methylation levels at particular loci have been associated with obesity and its related cardio-metabolic traits [1]. Recently, four large-scale epigenome-wide association studies (EWAS) [2–5] in middle-aged and

older adults identified multiple DNA methylation loci in blood leukocytes that were associated with body mass index (BMI). However, the high prevalence of obesity comorbidities and use of medications in middle-aged or older populations may hide or bias obesity-related DNA methylation changes. For this reason, we investigated 700 healthy African American youth and young adults. With genome-wide DNA methylation data from leukocytes in the full data set and genome-wide DNA methylation data from purified neutrophils and genome-wide gene expression data from leukocytes in a subset of 188 subjects, the present study has three objectives. First, we conducted an EWAS of obesity in youth and young adults to identify

* Correspondence: xwang@augusta.edu

¹Department of Pediatrics, Georgia Prevention Institute, Medical College of Georgia, Augusta University, HS-1640, Augusta, GA 30912, USA
Full list of author information is available at the end of the article

new signals and further validate the findings with replication in an independent cohort of 2097 middle-aged African Americans from the Atherosclerosis Risk in Communities (ARIC) study [2]. Second, for these obesity-related DNA methylation loci, we checked whether the associations can be replicated in one purified cell type (neutrophils), whether the associations were driven by obesity-related cardio-metabolic traits, and whether the DNA methylation status of these loci was associated with gene expressions. Third, we examined to what extent we can replicate the findings previously reported in middle-aged and older adults. The identification of obesity-related DNA methylation changes in youth and young adult may provide new insights into the mechanisms linking obesity to associated clinical conditions at the early stages of the disease process and may provide new targets for early prevention.

Methods

Subjects

A total of 700 African American youths and young adults aged 14–36 were included in the current study. These subjects were participants from 3 existing cohorts (for details, see Additional file 1): the Epigenetic Basis of Obesity-Induced Cardiovascular Disease and Type 2 Diabetes (EpiGO) study [6] (96 obese [BMI percentile \geq 95%] vs. 92 lean controls [BMI percentile \leq 50%] aged 14–20), the LACHY study [7] (284 participants from the general population aged 14–18), and the BP stress study [8] (228 participants from the general population aged 18–36). All of these participants were free of chronic or acute disease and not on daily prescription medication for treatment of diseases. All participants were recruited from the southeastern USA. Height and weight were measured by standard methods using a wall-mounted stadiometer and a scale, respectively. BMI was calculated as weight/height², and BMI percentile was calculated according to their age, sex, height, and weight. These studies were approved by the Institutional Review Board of Augusta University and performed following the guidelines of the Declaration of Helsinki. Written informed consent was provided by all participants or by their parents if they were less than 18 years.

DNA extraction and genome-wide DNA methylation

For the participants from the EpiGO study, DNA was extracted from both peripheral leukocytes and neutrophils using the QIAamp DNA Mini Kit (QIAGEN). Peripheral neutrophils were obtained using the approach previously described [9]. For the other participants, DNA was extracted from stored buffy coats using the same kit.

Genome-wide DNA methylation levels of all these samples were analyzed by the Illumina Infinium Human Methylation 450K Beadchip (Illumina Inc.). The Minfi

package [10] and CPACOR (incorporating Control Probe Adjustment and reduction of global CORrelation) package [11] were used for initial quantification, data preprocessing, and quality control (QC). The key QC steps included the following: (1) Detectable probes were defined as the probes with detection p value $< 1 \times 10^{-16}$ in more than 95% samples; (2) Detectable samples were defined as the samples with more than 95% CpG sites having a detection p value $< 1 \times 10^{-16}$ and correct classification of gender based on the genome-wide DNA methylation data; (3) Probes on the X and Y chromosomes and the 65 SNP markers were excluded; (4) Illumina background correction and quantile normalization were applied to all intensity values, and beta value was further calculated and used as the index of CpG methylation levels; (5) A principal component analysis (PCA) of the control probe intensities (excluding negative control probes) was performed and the resulting PCs 1 to 30 were stored; (6) White blood cell sub-populations were estimated using the approach described by Houseman et al. [12]; and (7) A linear regression model was conducted for each CpG site with DNA methylation level as the dependent variable and the 30 PCs from the control probe intensities as well as the estimated cell compositions as the independent variables. The residuals were calculated and used as indices of DNA methylation levels in further analysis. The above steps were conducted for each cohort separately.

RNA extraction and genome-wide gene expression assays

For the participants of the EpiGO study, RNA samples were extracted from the peripheral leukocytes stored in the RNA cell protection reagents (QIAGEN Inc.) using the QIAamp RNA mini Kit (QIAGEN Inc.). RNA concentration and purity were evaluated on a NanoDrop spectrophotometer 2000 (Thermo Scientific Inc.). RNA integrity (RIN) was evaluated on a Bioanalyzer 2100 (Agilent Inc.). The RIN scores of all these samples were greater than 8, indicating the high quality of RNA.

Genome-wide gene expression data were obtained using the Illumina HumanHT-12 v4 Expression BeadChip (Illumina Inc.). This chip targets more than 48,000 probes that provide genome-wide coverage of well-characterized genes, gene candidates, and splice variants. The GenomeStudio Gene Expression Module (Illumina Inc.) was used for initial quantification, and the lumi package [13] was used for data preprocessing and QC. The key QC steps included the following: (1) Probes with detection p value < 0.05 in more than 50% of the samples were defined as “present”; (2) Log transformation and quartile normalization were applied to the gene expression data; and (3) A linear regression model was conducted for each probe with gene expression level as the dependent variable and batches (16 batches with 12 samples in each batch) as the independent variables. The residuals were calculated

and used as the indices of gene expression levels in further analysis.

Statistical analysis

The R package Limma [14] was used for the identification of differentially methylated CpG sites and differentially expressed genes related to the obesity phenotypes.

For DNA methylation analysis on peripheral leukocytes, the residual derived from the regression of each single CpG site methylation level (beta) on cell compositions and PCs from the control probes was used as dependent variable with age and sex as covariates. The test of interest was group (obese vs. lean) in the EpiGO study and BMI in the LACHY and BP stress cohort. Meta-analysis across the three cohorts was conducted using METAL [15] by converting the direction of effect and p value observed in each study into a signed Z score. This approach is very flexible and allows results to be combined when the β -coefficients and standard errors from individual studies are in different units. To account for multiple testing, a standard Bonferroni correction for the 473,788 CpG sites gives $p < 1 \times 10^{-07}$ as the significance threshold.

For DNA methylation analysis on peripheral neutrophils, the residual derived from the regression of each single CpG site methylation level on PCs from the control probes was used as dependent variable with age and sex as covariates. The test of interest was group (obese vs. lean).

For gene expression analysis on peripheral leukocytes, the residual derived from the regression of each probe gene expression level on batches was used as dependent variable with age and sex as covariates. The test of interest was group (obese vs. lean).

The differentially methylated CpG sites identified for obesity and BMI in peripheral leukocytes with a p value $< 1 \times 10^{-07}$ were carried forward for replication in the African American participants of the ARIC study [2] ($n = 2097$, 64% females, age range from 47 to 70). Leukocyte DNA and the same Illumina 450K platform for methylation analysis were used in ARIC study. Replication was defined as consistent direction of the β -coefficient and $FDR < 0.05$ in the BMI-related CpG sites from the ARIC study. The replicated CpG sites were taken forward in the following analyses: First, they were checked in obesity-related CpGs identified from neutrophils. Replication was defined as consistent direction of the β -coefficient and $FDR < 0.05$ in neutrophils. Second, their associations with seven cardio-metabolic traits including SBP, fasting glucose, fasting insulin, fasting triglycerides (TG), fasting total cholesterol (TC), fasting HDL-cholesterol (HDL), and fasting LDL-cholesterol (LDL) were tested in EpiGO and LACHY cohorts with and without the adjustment of obesity status/BMI. A p value < 0.05 was defined as significant. Third, *cis*-regulation of these CpG sites of gene expression (within \pm

250 kb of the CpG sites) was explored. Partial correlation of gene expression and DNA methylation was conducted with age, gender, and group (obese vs. lean) as covariates. A Bonferroni corrected p value < 0.05 according to the number of genes tested for each CpG site (range from 0 to 28) was defined as significant correlation. Obesity-related differentially expressed genes were defined as genes with their expression levels showing significant difference between obese cases and lean controls at p value < 0.05 .

Pathway enrichment analysis was conducted for the genome-wide DNA methylation obtained from leukocytes using gene set enrichment analysis (GSEA) [16]. GSEA was performed on an unfiltered, ranked list of genes (ranked by the p values without consideration of directions), and a running-sum statistic was used to determine the enrichment of an a priori defined gene sets (pathways) based on the gene ranks. Statistical significance of pathway enrichment scores were ascertained by permutation testing over size-matched random gene sets and multiple testing was controlled by the false discovery rate (FDR). A FDR of 5% was used. Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets were used as the reference gene sets. The CpG site showing the most significant p value within a gene was used to represent the DNA methylation level of the gene.

We also checked whether we could replicate previously identified obesity-related CpG sites ($n = 277$) in our current study of youth and young adult [2–5]. Replication was defined as a consistent direction of the β -coefficient and $FDR < 0.05$.

Results

Discovery meta-analysis in the youth and young adult

Table 1 lists the general characteristics of the subjects. We identified 76 CpG sites significantly associated ($p < 1 \times 10^{-07}$) with obesity in the meta-analysis of the three cohorts (Manhattan and QQ-plot, Fig. 1; Additional file 2: Table S1). Obesity was positively associated with the methylation of 65 CpG sites and negatively associated with the methylation of 11 CpG sites. The top CpG site (cg12170787, $p = 1.13 \times 10^{-16}$) locates in the intron 5 of the *SBNO2* gene (Strawberry Notch Homolog 2). Two more CpG sites (cg18608055, $p = 3.69 \times 10^{-10}$; cg07573872, $p = 1.12 \times 10^{-09}$) locating in the same region also passed the threshold for genome-wide significance. CpG site cg07573872 is located in a CpG island shore region and has been reported to be associated with BMI previously [2]. CpG sites cg18608055 and cg12170787 are 4524 bp and 4623 bp away from cg07573872 (Additional file 3: Figure S1). The *SBNO2* gene is expressed in peripheral leukocytes and is a key player in the IL-10-regulated anti-inflammatory signaling pathway [17].

Table 1 General characteristics of the subjects

General characteristics	EpiGO		LACHY	BP stress cohort
	Lean	Obese		
N	92	96	284	228
Female (%)	50.0	51.0	50.0	57.5
Age (years)	17.7 ± 1.7	17.7 ± 1.8	16.2 ± 1.3	27.8 ± 3.3
Age range (years)	14.0–20.9	14.1–21.0	13.8–19.0	18.9–35.6
BMI (kg/m ²)	18.8 ± 1.3	39.8 ± 6.8	24.1 ± 5.6	31.4 ± 8.6
BMI range (kg/m ²)	15.0–21.7	28.1–70.1	16.4–45.9	17.6–59.6
BMI percentile (%)	18.7 ± 11.0	98.8 ± 1.1	65.7 ± 28.3	–
BMI percentile range (%)	0.8–41.7	95.0–99.9	2.3–99.7	–

In silico replication with ARIC study

Of the 76 CpG sites significantly associated with obesity in our discovery meta-analysis, 54 replicated (FDR < 0.05) in the ARIC study (Table 2), of which 33 survived Bonferroni correction ($p < 6.58 \times 10^{-4}$). These 54 CpG sites annotated to 45 genes. In addition to *SBNO2* which exhibits 3 significant CpG sites, *SOCS3* (suppressor of cytokine signaling 3: cg18181703, cg04610187, and cg10508317), *CISH* (Cytokine-Inducible Src-Homology 2 –containing protein: cg21585138 and cg23005227), and *VMP1* (Vacuole Membrane Protein: cg16936953, cg12054453, cg24174557, cg18942579, and cg010409343) also have 2 or more significant CpG sites. The locations of these CpG sites within each gene and their correlations were provided in Additional file 3: Figure S1. Out of the 54 CpG sites associated with obesity, 29 CpG sites (annotated to 25 genes) were novel signals (highlighted in gray in Table 2).

Replication of obesity-related CpG sites identified from mixed cells in purified neutrophils

The 54 CpG sites were taken forward to check whether the associations with obesity could be validated in one purified cell type: neutrophils. The results were listed in

Additional file 2: Table S2. Thirty-seven CpG sites were found to be associated with obesity in neutrophils (FDR < 0.05 and consistent directions of effect), confirming that for the majority of the CpG sites (68.5%), the significant associations with obesity are not driven by potential cell compositions in leukocytes. For the 17 CpG sites that were not validated in neutrophils, it is difficult to determine whether these signals are driven by cell compositions or that neutrophils are not the biologically relevant cell type to target. The fact that 6 of the 17 CpG sites were hypermethylated (i.e., $\beta > 0.8$) in neutrophils compared with 0 in the 37 replicated CpG sites ($p = 1.2E-4$) indicated that at least for some of the obesity-related CpG sites which could not be validated in neutrophils, neutrophils might not be the relevant cell type to target.

Association with cardio-metabolic phenotypes

We examined the associations between the 54 replicated CpG sites and 7 cardiometabolic traits (SBP, fasting insulin, fasting glucose, fasting lipid panel including TG, TC, HDLC, and LDLC). Without the adjustment of obesity or BMI, we observed 148 significant associations ($p < 0.05$, Additional file 2: Table S3) with 39 for SBP, 45 for insulin, 1 for glucose, 14 for TG, 5 for TC, 25 for HDLC, and 19 for LDLC. CpGs that were significantly associated with higher SBP, insulin, glucose, TG, TC, and LDLC were also associated with higher BMI levels. For HDLC, CpGs were associated with lower BMI levels (Fig. 2). After the adjustment of obesity or BMI, only 11 associations remained significant (Additional file 2: Table S4). The results indicate that the majority of the associations of these CpG sites with cardiometabolic phenotypes are driven by obesity.

Gene expression analyses

Of the 54 CpG sites, 16 were significantly associated with expression of 17 unique genes in cis (Bonferroni corrected p value < 0.05) (Additional file 2: Table S5).

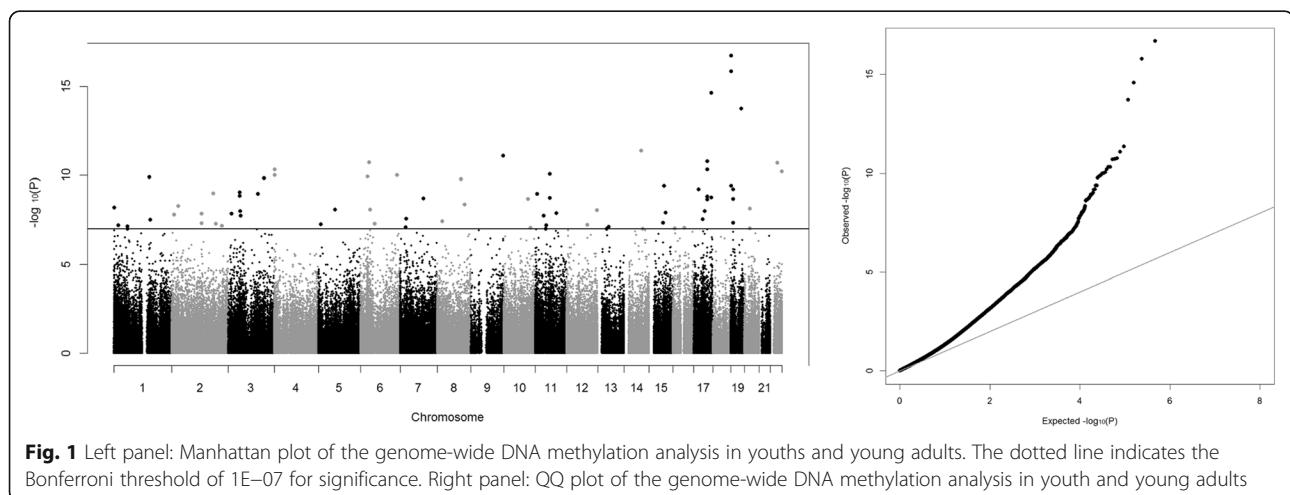


Table 2 Validated CpG sites in the ARIC study

TargetID	Closest gene	Chr.	Position	The youth and young adult discovery panel		ARIC		
				Directions*	P value	Beta**	P value	FDR
cg12170787	SBNO2	19	1130965	---	1.13E-16	-0.00347	1.30E-05	7.60E-05
cg18181703	SOCS3	17	76354621	---	1.94E-16	-0.00679	3.80E-07	4.13E-06
cg18608055	SBNO2	19	1130866	---	3.69E-16	-0.00460	5.80E-04	1.34E-03
cg26470501	BCL3	19	45252955	---	4.87E-14	-0.00247	1.84E-02	2.65E-02
cg09349128	CRELD2	22	50327986	---	1.15E-13	-0.00590	1.40E-08	2.66E-07
cg16611584	AKAP10	17	19809078	+++	7.37E-12	0.00810	2.08E-04	5.27E-04
cg03940776	SYNJ2	6	158490013	---	8.82E-12	-0.00306	1.40E-04	4.09E-04
cg04610187	SOCS3	17	76360794	---	8.95E-12	-0.00510	1.89E-03	3.34E-03
cg10919522	ELMSAN1	14	74227441	---	1.18E-11	-0.00548	4.00E-06	3.04E-05
cg08548559	PIK3IP1	22	31686097	---	1.49E-11	-0.00743	1.40E-07	1.77E-06
cg12593793	LMNA	1	156074135	---	1.79E-11	-0.00385	1.59E-04	4.32E-04
cg06178669	HRASLS2	11	63334608	---	3.50E-11	-0.00507	3.12E-03	5.15E-03
cg21585138	CISH	3	50645106	---	4.52E-11	-0.00572	1.10E-06	9.29E-06
cg07136133	PRR5L	11	36422377	---	6.77E-11	-0.00484	6.70E-09	2.10E-07
cg01671681	PLCH1	3	155421735	---	2.02E-10	-0.00440	1.48E-04	4.17E-04
cg03957124	COX6A1P2	6	37016869	---	2.91E-10	-0.00268	2.77E-03	4.67E-03
cg16936953	VMP1	17	57915665	---	3.17E-10	-0.00777	2.40E-05	1.22E-04
cg10601624	PLEKHG6	12	6404377	---	3.34E-10	-0.00454	6.90E-05	2.91E-04
cg07037944	DAPK2	15	64290807	---	5.48E-10	-0.00349	2.47E-04	6.06E-04
cg17178175	NFE2L2	2	178109973	---	5.56E-10	-0.00397	1.38E-03	2.69E-03
cg06568880	SMG6	17	2166583	+++	6.00E-10	0.00258	1.63E-03	3.10E-03
cg10508317	SOCS3	17	76355146	---	7.45E-10	-0.00197	1.31E-03	2.62E-03
cg09152259	MAP3K2	2	128156114	---	8.03E-10	-0.00429	7.10E-04	1.54E-03
cg21163717	DOK2	8	21769903	---	8.41E-10	-0.00339	1.27E-04	3.86E-04
cg07573872	SBNO2	19	1126342	---	1.12E-09	-0.00689	2.80E-08	4.26E-07
cg23005227	CISH	3	50645426	---	1.26E-09	-0.00416	1.73E-03	3.20E-03
cg22866430	SNORA59A	1	12566378	+++	2.68E-09	0.00179	1.30E-02	1.90E-02
cg12054453	VMP1	17	57915717	---	2.71E-09	-0.00745	8.10E-05	3.08E-04
cg07094298	TNIP2	4	2748026	---	3.08E-09	-0.00534	8.10E-05	3.08E-04
cg24174557	VMP1	17	57903544	---	3.35E-09	-0.00556	1.91E-04	5.01E-04
cg19750657	UFM1	13	38935967	+++	5.50E-09	0.00805	5.50E-06	3.80E-05
cg23076949	FBRSL1	12	133016265	+++	5.58E-09	0.00316	1.01E-03	2.13E-03
cg06647068	CHST11	12	104853274	---	6.48E-09	-0.00367	1.20E-02	1.78E-02
cg17061862	WEE1	11	9590431	---	6.70E-09	-0.00458	1.26E-03	2.59E-03
cg03183540	LOC101927822	8	134931756	---	6.97E-09	-0.00282	2.50E-03	4.31E-03
cg05845030	DCN	12	91573247	---	9.27E-09	-0.00260	2.51E-02	3.54E-02
cg24531955	LOXL2	8	23154691	---	1.20E-08	-0.00416	3.80E-05	1.70E-04
cg03257930	RNU6-996P	9	110399293	---	1.50E-08	-0.00251	6.11E-03	9.67E-03
cg00760203	PLXDC1	17	37254921	+++	1.64E-08	0.00458	9.20E-05	3.18E-04
cg20849150	LINC01358	1	59522993	---	2.34E-08	-0.00401	3.40E-05	1.62E-04
cg18942579	VMP1	17	57915773	---	2.52E-08	-0.00658	8.90E-05	3.18E-04
cg02734358	GPRIN3	4	90227074	---	4.22E-08	-0.00525	3.29E-04	7.81E-04
cg22143698	ANKRD33B	5	10608058	+++	4.46E-08	0.00628	1.50E-05	8.14E-05
cg08857797	VPS25	17	40927699	+++	4.98E-08	0.00634	3.40E-09	2.10E-07
cg07960624	SAMD12	8	119208486	---	6.13E-08	-0.00578	6.98E-04	1.54E-03
cg03929796	ALAS1	3	52231840	---	6.94E-08	-0.00317	8.36E-03	1.30E-02
cg08239103	SOX6	11	16623818	---	7.32E-08	-0.00342	4.74E-03	7.66E-03
cg06946797	RIM2	16	11422409	---	7.38E-08	-0.00702	8.30E-09	2.10E-07
cg04163119	PRICKLE4	6	41754006	---	7.77E-08	-0.00153	9.49E-03	1.44E-02
cg13134297	CRHR2	7	30737556	---	8.47E-08	-0.00374	1.10E-04	3.48E-04
cg01409343	VMP1	17	57915740	---	8.53E-08	-0.00546	1.03E-04	3.40E-04
cg14531564	SDF4	1	1154853	---	8.54E-08	-0.00410	1.10E-05	6.97E-05
cg06164260	BCL6	3	187454439	---	9.42E-08	-0.00410	5.30E-07	5.04E-06
cg23018755	MAFG	17	79881529	+++	9.85E-08	0.00339	1.86E-03	3.34E-03

Newly discovered associations are highlighted in gray

**+ indicates that DNA methylation levels increase with obesity status or BMI. The order is Epigo, LACHY, and BP stress cohort

**DNA methylation changes with 1 unit increase in BMI

Furthermore, of these 17 genes, the expression levels of 5 genes were significantly associated with obesity ($p < 0.05$) (Additional file 2: Table S5). Figure 3 shows the corresponding relationships between DNA methylation

and gene expression for these 5 genes. With the exception of the correlation between cg06178669 and *HRASLS2* (*HRAS*-like suppressor 2), increased methylation was associated with decreased gene expressions. For

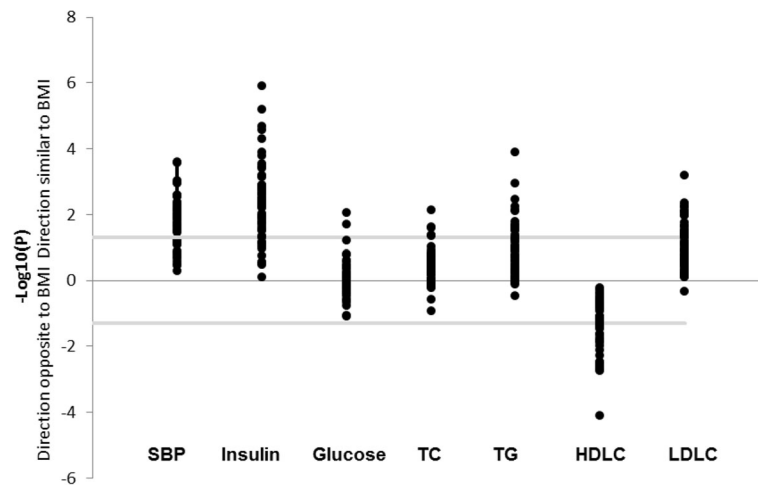


Fig. 2 Manhattan plot depicting the $-\log_{10}(p)$ values and effect direction (respectively to obesity) of the associations between the 55 replicated CpG sites and each cardiometabolic phenotype. The lines indicate the threshold of 0.05 for significance

SOCS3, *CISH*, *PIM3* (Pim-3 proto-oncogene, serine/threonine kinase), and *KLF4* (Kruppel-like factor 4), obesity was associated with decreased methylation and increased gene expression, while for *HRASLS2*, obesity was associated with decreased methylation and decreased gene expression (Additional file 2: Table S5).

Pathway analyses

The pathway analyses yielded significant ($FDR < 0.05$) enrichment of 33 KEGG pathways for obesity-related DNA methylation changes in peripheral leukocytes (Additional file 2: Table S6). These pathways represent processes involved in inflammatory pathways (e.g., cell adhesion molecules, JAK-STAT signaling pathway, intestinal immune network for IGA production), autoimmune response (e.g., type 1 diabetes, allograft rejection, graft versus host disease), and obesity-related comorbidities (e.g. type 2 diabetes, vascular smooth muscle contraction, cancer, and asthma).

Replication of previously reported CpG sites for obesity in middle-aged and elderly adults

We attempted to replicate the previously reported CpG sites for obesity. At the time this study was conducted, 4 large EWASs (with replication cohorts) for obesity were available: the ARIC study by Demerath E et al. [2] (reported 37 CpG sites), the Cardiogenics Consortium study by Dick K et al. [4] (reported 3 CpG sites), the GOLDN study by Aslibekyan S et al. [5] (reported 8 CpG sites), and the study reported by Wahl S et al. [3] (reported 254 CpG sites). These 4 studies resulted in 277 unique CpG sites associated with obesity. Out of these 277 CpG sites, 198 (71.5%) could be replicated with consistent direction of effect and a $FDR < 0.05$ in our study (Additional file 2: Table S7), indicating that

majority of the obesity-related CpG sites identified in middle-aged and elderly adults already show changes in youth and young adulthood.

Discussion

In this epigenome-wide association study of African American youth and young adults, we identified 29 novel CpG sites associated with obesity and the replicated majority of the CpG sites previously identified in middle-aged and older adults. We further demonstrated that majority of the obesity-related CpG sites identified from leukocytes are not driven by cell compositions or obesity-related cardio-metabolic risk factors. We also provided direct links between DNA methylation, gene expression, and obesity status for 7 CpG sites in 5 genes.

Unlike genetic markers that are the same in all cells and unlikely to change over a lifetime, epigenetic regulation is tissue specific and plastic (i.e., findings may reflect the consequence rather than the cause of the disease) [18]. In addition to availability, the choice of leukocytes in this study as well as in the previous epigenome-wide association studies of obesity in middle-aged and elder people is based on the fact that obesity is characterized by a state of chronic, low-grade inflammation [19]. With the established role of epigenetic regulation in shaping the immune and inflammatory response [20], identification of the epigenetic markers specifically involved in obesity-induced inflammation has the potential to provide novel insight into the pathogenesis of obesity-related disorders. The choice of leukocytes also assumes that the observed epigenetic changes will reflect consequences rather than causes of obesity. Although the cross-sectional design of the current and previous studies [2–5] does not provide information on the direction of causality, the Mendelian randomization model

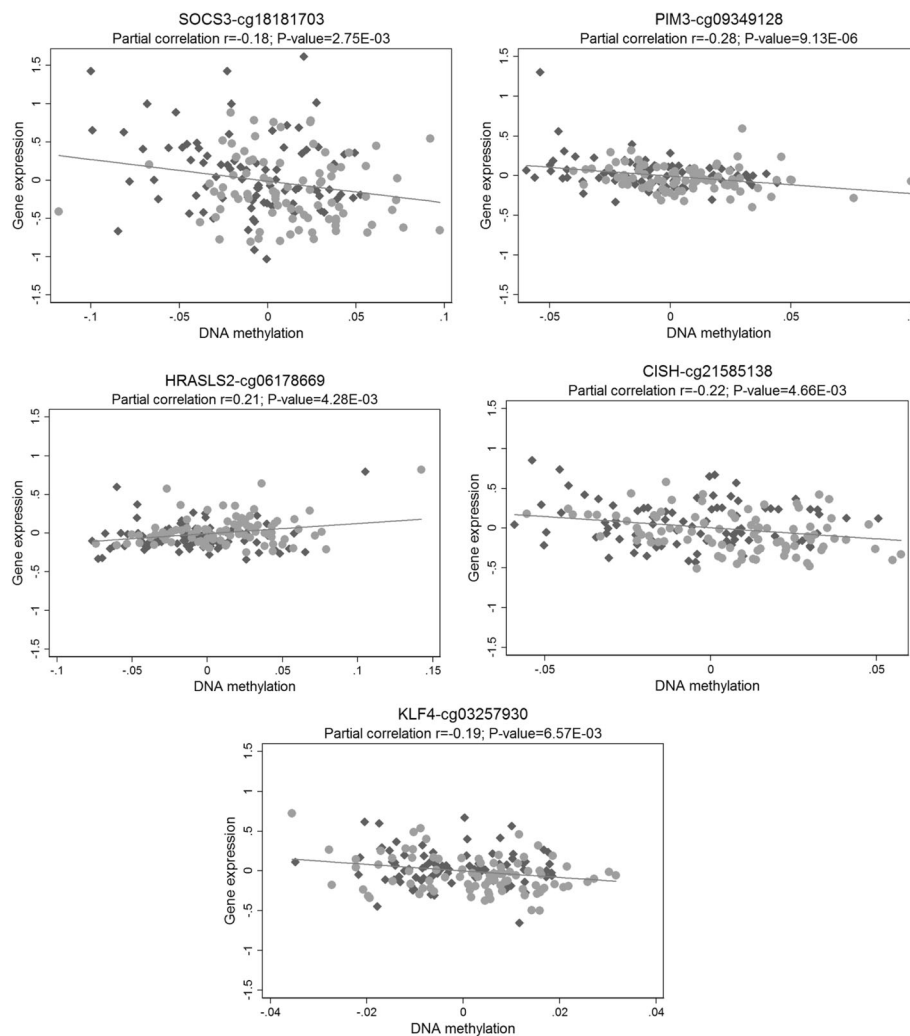


Fig. 3 Partial correlations between DNA methylation and gene expression. Covariates included age, sex, and obesity status. Gray dots represent the lean group and black diamonds represent the obese group. For the *SOCS3* gene, similar correlation was observed for cg04610187 ($r = -0.20$, $p = 1.77E-03$); for the *CISH* gene, similar correlation was observed for cg23005227 ($r = -0.21$, $p = 7.14E-03$)

used in one [3] of the previous adult studies strongly suggested that obesity-related DNA methylation changes identified in peripheral leukocytes are predominately the consequence rather than the cause of obesity. One concern in analyzing epigenetic changes in peripheral leukocytes in obese individuals is that the effect of obesity itself on the immune system can be hidden or biased by the coexistence of obesity-related diseases and the use of medication, which is very common in middle-aged and elderly individuals. To avoid this problem, in this study, we focused on youth and young adults. This is a population having shown obesity-related metabolic risk that has not developed into clinical disease yet. We did observe that many obesity-related CpG sites were associated with cardiometabolic traits. However, after the adjustment of obesity, this number decreased dramatically, indicating that a large majority of the associations

of these CpG sites with cardiometabolic phenotypes is driven by obesity.

In comparing with the previous studies in middle-aged adults and elderly [2–5], we identified 29 novel CpG sites (annotated to 25 genes). We were also able to present triangular relationships between DNA methylation, gene expression, and obesity at 5 genes including *SOCS3*, *CISH*, *PIM3*, *KLF4*, and *HRASLS2*. With the exception of *HARSL2*, obesity was associated with decreased DNA methylation level and increased gene expressions. Interestingly, both *SOCS3* and *CISH* are involved in the JAK-STAT signaling pathway and are the key negative regulators of the activation of this pathway [21]. The JAK-STAT signaling pathway is activated during inflammation and used by a variety of cytokines. The JAK-STAT signaling pathway activates its own suppressors: suppressors of cytokine signaling (SOCS)

molecules including SOCS1-7 and CISH. This negative feedback control is essential for the effective dissipation of cytokine signaling to prevent excessive inflammation and detrimental effect on other signaling pathways. However, the increased levels of SOCS proteins can induce insulin resistance in peripheral organs [22] and leptin resistance in the central nervous system [23]. There are very limited studies on *PIM3*, *KLF4*, and *HARSLS2*. However, *PIM1*, a protein from the same family and structurally and functionally similar to *PIM3*, is also primarily involved in the JAK/STAT signaling pathway. *PIM-1* transcription is initiated by STAT factors and can also bind to regulators of the JAK/STAT pathway, resulting in a negative feedback loop [24]. Taken together, the activation of the JAK/STAT signaling pathway in obesity transcriptionally upregulates its own suppressors through DNA methylation mechanisms. On the one hand, the increased production of negative regulators can prevent excessive inflammation, while on the other hand, these increased negative feedback products may increase the risk of obesity-related disorders, a compensatory mechanism that has been observed in the development of many diseases.

Our study in healthy youth and young adults also replicated 71.5% of obesity-related CpG sites previously reported in middle-aged and elderly adults, indicating that the majority of the obesity-related CpG sites already show changes in adolescence. It will be interesting to find out the age of onset of these changes. In a recent study exploring this question on the first EWAS-identified gene (*HIF3A*), Pan et al. [25] observed that DNA methylation levels at the three previously described *HIF3A* CpG sites were already associated with greater weight and adiposity at birth. On the other hand, the strong consistency of obesity-related DNA methylation signals across ethnic groups (African Americans in this study comparing with Caucasians, Asians, and African Americans in previous studies) and age emphasizes the importance of conducting large-scale epigenome-wide meta-analysis similar to what has been done in the GWAS. This action will discover more obesity-related epigenome-wide changes and enable the recognition of the overall picture of the role of DNA methylation in obesity and its related disorders.

One possible concern in these EWAS studies on obesity using peripheral blood is the heterogeneity of leukocytes since different cell populations have distinct epigenetic signatures [26]. This concern was dramatically reduced with the development of algorithms estimating the cell subset compositions based on genome-wide DNA methylation data [12]. Our study further confirms this through successfully replication of 68.5% of the obesity-related CpG sites from leukocytes in one purified cell type: neutrophils. The reason that neutrophils were selected is that previous studies including ours [9, 27]

have observed that obesity is consistently associated with neutrophilia, manifesting not only by elevated neutrophil count but also increased percentage. However, we believe that replication of the findings from leukocytes in a variety of other single cell types remains important to guide further functional in vitro studies.

Several limitations to the present study need to be recognized. First, although the study by Wahl S et al. [3] in adults have shown that obesity-related DNA methylation changes identified in peripheral leukocytes are predominantly the consequence rather than the cause of obesity, we cannot test the direction of the causality of these newly identified obesity-related CpG sites in youth and young adults due to the cross-sectional design and the lack of genotype data. Second, although 71% of the signals from the discovery cohort were validated in the replication cohort, the lack of replication of other signals might be due to the age difference between the discovery cohort and the replication cohort. Third, due to the lack of genotype data, we cannot adjust the potential bias resulted from ancestry compositions.

Conclusion

In summary, this EWAS study in healthy youth and young adult confirmed the majority of previously identified obesity-related DNA methylation loci in middle-aged and elderly adults and identified 29 novel CpG sites associated with obesity. The findings from the current study have the potential to enable development of new strategies for early prevention and treatment.

Additional files

Additional file 1: Detailed description of the participants for each study. (DOCX 22 kb)

Additional file 2: Table S1. CpG sites with $p < 1E-7$ in this youth and young adult study. **Table S2.** Performance of the 55 CpG sites in neutrophils. **Table S3.** Association of the 55 CpG sites with cardiometabolic traits without adjustment of obesity/BMI. **Table S4.** The significant associations between CpG sites and metabolic traits after the adjustment of obesity/BMI. **Table S5.** DNA methylation, gene expression, and obesity. **Table S6.** Significantly enriched pathways (FDR < 0.05). **Table S7.** Replication of previous reported CpG sites identified in middle-aged and elderly population. (XLSX 64 kb)

Additional file 3: Figure S1. Positions and correlations of the multiple CpG sites in *SBNO2*, *SOCS3*, *VMP1*, and *CISH* genes. (PDF 425 kb)

Abbreviations

ARIC: Atherosclerosis Risk in Communities; BMI: Body mass index; BP: Blood pressure; CISH: Cytokine-Inducible Src-Homology 2-containing protein; CPACOR: Incorporating Control Probe Adjustment and reduction of global CORrelation; CpG: 5'-C-phosphate-G-3'; DNA: Deoxyribonucleic acid; EpiGO: Epigenetic Basis of Obesity-Induced Cardiovascular Disease and Type 2 Diabetes; EWAS: Epigenome-wide association study; FDR: False discovery rate; GOLDN: Genetics of Lipid Lowering Drugs and Diet Network; GSEA: Gene set enrichment analysis; GWAS: Genome-wide association study; HDLC: Fasting HDL-cholesterol; HIF3A: Hypoxia-inducible factor 3 alpha subunit; HRASLS2: HRAS-like suppressor 2; IL-10: Interleukin-10; JAK-STAT: Janus kinase/signal transducers and activators of transcription; KEGG: Kyoto Encyclopedia of Genes and Genomes; KLF4: Kruppel-like factor

4; LACHY: Lifestyle, Adiposity, and Cardiovascular Health in Youth; LDLC: Fasting LDL-cholesterol; PCA: Principal component analysis; PIM1: Pim-1 proto-oncogene, serine/threonine kinase; PIM3: Pim-3 proto-oncogene, serine/threonine kinase; QC: Quality control; RIN: RNA integrity; SBNO2: Strawberry notch homolog 2; SBP: Systolic blood pressure; SNP: Single-nucleotide polymorphism; SOCS: Suppressors of cytokine signaling; SOCS3: Suppressor of cytokine signaling 3; TC: Fasting total cholesterol; TG: Fasting triglycerides; VMP1: Vacuole membrane protein

Acknowledgements

Not applicable.

Funding

This study was supported by grant HL105689 and HL125577 from the National Institute of Health (NIH) and GRNT20480211 from AHA.

Availability of data and materials

The genome-wide data that support the findings of this study are available from the authors upon reasonable request and with permission of the Institutional Review Board of Augusta University.

Authors' contributions

XW and SS designed the study and drafted the manuscript; YP, GH, YH, and WG conducted the data analysis; HZ, XW, VB, HS, and YD participated in the data collection; HS made critical changes to the manuscript; JP, KN, MC, and ED conducted the in silico replication of the study. All authors read and approved the final manuscript.

Ethics approval and consent to participate

These studies were approved by the Institutional Review Board of Augusta University and performed following the guidelines of the Declaration of Helsinki. Written informed consent was provided by all participants or by their parents if they were less than 18 years.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Pediatrics, Georgia Prevention Institute, Medical College of Georgia, Augusta University, HS-1640, Augusta, GA 30912, USA. ²Cancer Center, Augusta University, Augusta, GA, USA. ³Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. ⁴Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55454, USA. ⁵Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA. ⁶Human Genetics Center, School of Public Health, University of Texas Health Sciences Center at Houston, Houston, TX 77030, USA.

Received: 15 March 2017 Accepted: 15 December 2017

Published online: 05 January 2018

References

- van Dijk SJ, Tellam RL, Morrison JL, Muhlhauser BS, Molloy PL. Recent developments on the role of epigenetics in obesity and metabolic disease. *Clin Epigenetics*. 2015;7:66.
- Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, Hedman AK, Sandling JK, Li LA, Irvin MR, 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*. 2015; 24(15):4464–79.
- Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*. 2017;541(7635):81–6.
- Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, Meduri E, Morange PE, Gagnon F, Grallert H, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383(9933):1990–8.
- Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, Sha J, Pankow JS, Liu C, Irvin MR, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity (Silver Spring)*. 2015;23(7):1493–501.
- Xu X, Su S, Barnes VA, De Miguel C, Pollock J, Ownby D, Shi H, Zhu H, Snieder H, Wang X. A genome-wide methylation study on obesity: differential variability and differential methylation. *Epigenetics*. 2013;8(5):522–33.
- Gutin B, Johnson MH, Humphries MC, Hatfield-Laube JL, Kapuku GK, Allison JD, Gower BA, Daniels SR, Barbeau P. Relationship of visceral adiposity to cardiovascular disease risk factors in black and white teens. *Obesity (Silver Spring)*. 2007;15(4):1029–35.
- Wang X, Poole JC, Treiber FA, Harshfield GA, Hanevold CD, Snieder H. Ethnic and gender differences in ambulatory blood pressure trajectories: results from a 15-year longitudinal study in youth and young adults. *Circulation*. 2006;114(25):2780–7.
- Xu X, Su S, Wang X, Barnes V, De Miguel C, Ownby D, Pollock J, Snieder H, Chen W. Obesity is associated with more activated neutrophils in African American male youth. *Int J Obes*. 2015;39(1):26–32.
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9.
- Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*. 2015;16:37.
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
- Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24(13):1547–8.
- Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*. 2005;21(9):2067–75.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*. 2010;26(17):2190–1.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
- El Kasmī KC, Smith AM, Williams L, Neale G, Panopoulos AD, Watowich SS, Hacker H, Foxwell BM, Murray PJ. Cutting edge: a transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway. *J Immunol*. 2007;179(11):7215–9.
- Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004;429(6990):457–63.
- Karalis KP, Giannogonas P, Kodala E, Koutmani Y, Zoumakis M, Teli T. Mechanisms of obesity and related pathology: linking immune responses to metabolic stress. *FEBS J*. 2009;276(20):5747–54.
- Wilson AG. Epigenetic regulation of gene expression in the inflammatory response and relevance to common diseases. *J Periodontol*. 2008;79(8 Suppl):1514–9.
- Wunderlich CM, Hovelmeyer N, Wunderlich FT. Mechanisms of chronic JAK-STAT3-SOCS3 signaling in obesity. *JAKSTAT*. 2013;2(2):e23878.
- Galic S, Sachithanandan N, Kay TW, Steinberg GR. Suppressor of cytokine signalling (SOCS) proteins as guardians of inflammatory responses critical for regulating insulin sensitivity. *Biochem J*. 2014;461(2):177–88.
- Lubis AR, Widia F, Soegondo S, Setiawati A. The role of SOCS-3 protein in leptin resistance and obesity. *Acta Med Indones*. 2008;40(2):89–95.
- Bachmann M, Moroy T. The serine/threonine kinase Pim-1. *Int J Biochem Cell Biol*. 2005;37(4):726–30.
- Pan H, Lin X, Wu Y, Chen L, Teh AL, Soh SE, Lee YS, Tint MT, Maclsaac JL, Morin AM, et al. HIF3A association with adiposity: the story begins before birth. *Epigenomics*. 2015;7(6):937–50.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, Soderhall C, Scheynius A, Kere J. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One*. 2012;7(7):e41361.
- Pecht T, Gutman-Tirosh A, Bashan N, Rudich A. Peripheral blood leucocyte subclasses as potential biomarkers of adipose tissue inflammation and obesity subphenotypes in humans. *Obes Rev*. 2014;15(4):322–37.