Supplementary Materials:

Integrative analysis of glucometabolic traits, adipose tissue DNA methylation and gene expression identifies epigenetic regulatory mechanisms of insulin resistance and obesity in African Americans

Short title: DNA methylation and insulin resistance in African Americans

Neeraj K. Sharma[1], Mary E. Comeau[2], Dennis Montoya[3], Matteo Pellegrini[3], Timothy D Howard[4], Carl D. Langefeld[2], and Swapan K. Das [1,*]

[*] Corresponding author and person to whom reprint requests should be addressed:

Swapan K. Das, Ph.D.

Section on Endocrinology and Metabolism

Department of Internal Medicine, Wake Forest School of Medicine

Medical Center Boulevard, Winston-Salem, North Carolina 27157

Email: sdas@wakehealth.edu, Telephone: 336-713-6057; Fax: 336-713-7200

Supplementary Technical Information:

DNA isolation from adipose tissue: Subcutaneous adipose tissue near the umbilicus was collected by Bergstrom needle from participants after an overnight fast. Biopsies were immediately rinsed in sterile saline, quick-frozen in liquid nitrogen, and stored at -80^{0}C. Genomic DNA from ~100 mg of frozen subcutaneous adipose tissue biopsies was isolated with the Qiagen DNeasy tissue kit (Cat#69504). Briefly, tissue samples were incubated at 56 ^{0}C for 3 hours in 450 µl of ATL buffer and 30 µl of proteinase K (Qiagen, Cat # 19131) to achieve complete lysis. RNase A solution (4 µl, 100 mg/ml; Qiagen RNase A Solution Cat # 19101) was then added to the lysed tissue samples and were further incubated at room temperature for four minutes. DNA was eluted through DNeasy mini spin column by adding 60 µl Elution Buffer (AE Buffer). Concentration and purity of isolated DNA samples were checked spectrophometrically by Nanodrop. Extraction for samples with low yield (<40ng/µl) and quality (A260/280 <1.8) was repeated to obtain DNA suitable for DNA methylation profiling.

DNA methylation profiling: Epigenome-wide profiling of DNA methylation levels was performed by Reduced Representation Bisulfite Sequencing (RRBS). The DNAm profiling and bioinformatics analysis of RRBS data was conducted by Diagenode RRBS service (Diagenode, Belgium). Genomic DNA samples were checked to ensure availability of high quality DNA for library preparation with the capillary electrophoretic system Fragment Analyzer™ and the DNF-487 Standard Sensitivity Genomic DNA Analysis Kit (Advanced Analytical). DNA concentration was measured fluorometrically using the Qubit® dsDNA BR Assay Kit (Thermo Fisher Scientific). RRBS libraries were prepared using the Premium Reduced Representation Bisulfite Sequencing Kit (Diagenode Cat# C02030033). Briefly, 100ng of genomic DNA were used to start library preparation for each sample. Following library preparation, samples were analyzed by qPCR to assess the efficiency of this step. According to the Ct values obtained by the qPCR and making sure that each pool incorporates samples with different indexes, 7 or 8 samples were pooled together. Pooled samples were bisulfite-converted and a second qPCR was carried out to estimate the cycle number for library PCR amplification. Following final amplification, clean-up was performed using a 1.45x beads:sample ratio of Agencourt® AMPure® XP (Beckman Coulter). After completing the library preparation, the pools were quality checked before sequencing. DNA concentration of the pools was measured using the Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific) and the profiles were analyzed using the High Sensitivity DNA chip for 2100 Bioanalyzer (Agilent). RRBS library pools were sequenced on a HiSeq3000 (Illumina) using 50 bp single-read sequencing (SR50) to obtain at least 30 million reads/sample.

Bioinformatic Analysis of RRBS data: Adapter removal and read trimming of raw sequence data was performed using Trim Galore! version 0.5.0 (1) for RRBS libraries (--rrbs parameter), followed by a quality control with the FastQC tool (--fastqc parameter) (2). Reads smaller than 15 nucleotides after adapter removal and trimming were discarded (--length 15 parameter). The cleaned reads were then aligned to the *Homo sapiens* reference genome (Genome Reference Consortium 37, hg19) using Bismark v0.20.0 (3) with default parameters. Bismark is a specialized tool for mapping bisulfite-treated reads such as the ones generated in RRBS-Seq experiments. Bismark requires that the reference genome first undergo an *in-silico* bisulfite conversion while transforming the genome to forward (C -> T) and reverse strand (G -> A). The reads producing a unique best hit to one of the bisulfite genomes were then compared to the unconverted genome to identify cytosine contexts (CpG, CHG or CHH - where H is A, C or T). The cytosine2coverage and bismark_methylation_extractor modules of Bismark were used to infer the methylation state of all cytosines (for every single mappable read) and their context, and to compute the percentage methylation. The reported cytosines were filtered to get only the CpGs covered. The spike-in control sequences were used at this step to check the bisulfite conversion rates and to validate the efficiency of the bisulfite treatment. On average, 40.8 ± 5.8 million sequencing reads were obtained per sample. Alignment of this data to the human genome (hg19) resulted in an average of 21.4 ± 3.2 million (median 20.9 million) uniquely aligned reads per sample, corresponding to 52.4 ± 2.1% average mappability, and 15.1 ± 2.3X average coverage. The aligned RRBS data covered 4,708,455 CpGs (≥10X coverage in at least one sample, average 3,770,679 ± 189,047 CpG detected per sample). MethylKit (4), a R/Bioconductor package, was used to filter the CpG dataset for low coverage and for extremely high coverage to discard reads with possible PCR bias (CpGs with coverage higher than the 99.9th percentile were discarded). The data was then normalized for read coverage distribution between samples. The filtering and normalization was done using the functions filterByCoverage(myobj,lo.count=10,lo.perc=NULL, hi.count=NULL,hi.perc=99.9) and normalizeCoverage(filtered.myobj), respectively. The normalization was done with the methylKit default settings and therefore performed median normalization. Coverage tables and normalized methylation values for the CpG sites covered more than 10X in at least 75% of the samples were generated with methylKit by specifying the minimal number of samples to meet the required percentage (min.per.group option of methylKit *unite* function set to 172L for 75% of the samples; unite(normalized.myobj,min.per.group= x)). Normalized methylation values summarizing methylation information over windows of 1000bp across the genome, and covered more than 10X in at least 75% of the samples were also generated. Specific functionality in the methylKit package [e.g., tiles <-

tileMethylCounts(meth, win.size=1000, step.size=1000)] was used for this purpose and was considered as a CpG_methylation region.

Annotation of CpG sites and regions: Selected CpG sites and regions were annotated with the R/Bioconductor package 'annotatr' (http://bioconductor.org/packages/release/bioc/html/annotatr.html) (5). The annotation consisted of two categories: (i) distance to a CpG island and (ii) annotation to gene region. The distance related annotations identify whether CpG sites/regions overlapp a known CpG island, 2000bp of the flanking regions of the CpG islands (shores), 2000bp of the flanking regions of the shores (shelves), or outside these regions (cpg_inter or open sea). The gene region analysis classifies methylation site/regions in groups in the context of genes, namely, exons, UTRs, introns, promoters, and intergenic regions. Annotations were derived using hg19 data from the built-in annotations for 'hg19_cpgs', 'hg19_basicgenes', 'hg19_genes_intergenic', and 'hg19_genes_intronexonboundaries'. Additional annotation of CpG sites with respect to nearby SNPs was determined using Fast Locus Annotator (FALCAN), an in-house program and the UCSC table browser. To identify the regulatory potential of CpG sites, each site was categorized based on its predicted chromatin state. Chromatin state annotation data (ChromeHMM, Core 15-state model) was downloaded from the Roadmap Epigenomics Project (http://www.roadmapepigenomics.org/) for the three cell types/adult tissue source most likely to be related to adipose tissue [Mesenchymal Stem Cell Derived Adipocyte Cultured Cells (E023), Adipose Derived Mesenchymal Stem Cell Cultured Cells (E025), and Adipose nuclei (E063)], and merged with the CpG site chromosomal position.

Gene ontology and biological pathway enrichment analysis: Several insulin sensitivity and BMI-associated CpG sites are located in genes or gene promoters. The g:GOSt Gene group functional profiling analysis tool (in g:Profiler; https://biit.cs.ut.ee/gprofiler/) and DAVID (http://david.abcc.ncifcrf.gov/home.jsp) were used to determine enrichment of these CpG-annotated genes in biological pathways, and cellular and molecular functions.

***In silico* deconvolution of adipose tissue cell types:** Adipocytes are the most common type of cell in adipose tissue. However, adipose tissue is heterogeneous, composed of other cell types including endothelial cells and several immune cell subtypes, and its cellular composition changes in response to obesity and other pathophysiological conditions. Direct assessment of cell type composition in adipose tissue is challenging (6). In this study, we utilized *in silico* deconvolution to estimate the relative proportions of cell types in adipose tissue. We estimated proportions of cell types using both transcriptome and methylome data and two different reference-based methods.

In the first method, we used "Adipose Signature Matrix" recently developed by Glastonbury et al (6) using the CIBERSORT analytical tool and RNA-seq datasets from reference *in vitro* cells, including both primary cells and iPSC-derived cells that are known to be present in subcutaneous adipose tissue. The final CIBERSORT adipose signature matrix is comprised of 658 genes. Transcript level data for 506 of these 658 signature genes were available for our samples. We used this signature matrix in the CIBERSORTx analytical tool (https://cibersortx.stanford.edu/) (7) to deconvolute adipose tissue transcriptome data and estimate the relative proportions of cell types (adipocytes, CD4+ T cells, microvascular endothelial cells, M1- and M2-macrophage) of 253 individuals in AAGMEx cohort.

In the second method, a reference-based deconvolution algorithm was utilized as detailed in a previous study (8). In this method, references were collected from the methylation profiles of purified cell types from the Blueprint Epigenome Project and the International Human Epigenome Consortium (IHEC) data portal. In total, 30 methylation profiles were analyzed from macrophages, neutrophils, B cells, T cells, NK cells, endothelial cells, and adipocytes. Given the paucity of adipocyte references with adequate coverage, average methylation levels across all adipose samples in the AAGMEx cohort were also used as additional adipocyte reference (since the main cell type of the samples is presumably adipocytes), similarly used in Orozco et al (8). Since the sequencing coverage varies across RRBS samples, the methylation data from our cohort was aggregated to only include CpG loci that are covered in at least 80% of the samples which resulted in 868,592 CpG loci with sufficient common coverage. The remaining CpG loci were then imputed using nearest neighbor averaging. References were subsequently filtered to form common sample loci for subsequent analysis. First, to determine cell-specific methylation across all references, we used a sliding window to aggregate the methylation profiles into regions of at least two CpG loci with similar methylation (within 40% methylation difference across neighboring CpG within 500 bp). Selected regions were those uniquely methylated for each cell type to provide 259 cell-specific hypomethylated regions (composed of 743 reference and 609 sample CpG loci). Second, a non-negative least squares regression was performed using cell-specific regions to estimate the proportion of each cell type within each sample.

Genotyping: As described previously (9), Infinium HumanOmni5Exome-4 v1.1 DNA Analysis BeadChips (Illumina) were used to genotype DNA samples from the AAGMEx cohort based on the manufacturer's instructions. Genotype data were examined to verify sample and SNP quality. Genotype assays of 4,210,443 SNPs passed technical quality filters. Genotypes from 2,296,925 autosomal SNP assays (representing 2,210,735 unique high-quality genotyped SNPs with minor allele frequency (MAF) >0.01 and Hardy-Weinberg equilibrium-p-value $>1x10^{-6}$) were used in our published

adipose tissue expression quantitative trait (eQTL) analyses (9). We imputed these genotyped SNPs to the 1000 Genomes dataset (1KGP, phase 3 cosmopolitan reference panel) using the genotype imputation program Minimac3, implemented on the Michigan Imputation Server (https://imputationserver.sph.umich.edu/). The expanded genotype data set includes 14,502,313 autosomal genotyped and imputed SNPs (10). In this study, we have used genotype data to determine association between methylation sites and the genotype of local SNPs (*cis*-methylation quantitative trait loci; *cis*-mQTL) only for a selected gene region.

Supplementary References:

1. Krueger, Felix. Trim galore,A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. 2015.
Ref Type: Online Source

2. Andrews, Simon. FastQC: A quality control tool for high throughput sequence data; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010.
Ref Type: Online Source

3. Krueger,F, Andrews,SR: Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571-1572, 2011

4. Akalin,A, Kormaksson,M, Li,S, Garrett-Bakelman,FE, Figueroa,ME, Melnick,A, Mason,CE: methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 13:R87-13, 2012

5. Cavalcante,RG, Sartor,MA: annotatr: genomic regions in context. *Bioinformatics* 33:2381-2383, 2017

6. Glastonbury,CA, Couto,AA, El-Sayed Moustafa,JS, Small,KS: Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs. *Am J Hum Genet* 104:1013-1024, 2019

7. Newman,AM, Steen,CB, Liu,CL, Gentles,AJ, Chaudhuri,AA, Scherer,F, Khodadoust,MS, Esfahani,MS, Luca,BA, Steiner,D, Diehn,M, Alizadeh,AA: Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 37:773-782, 2019

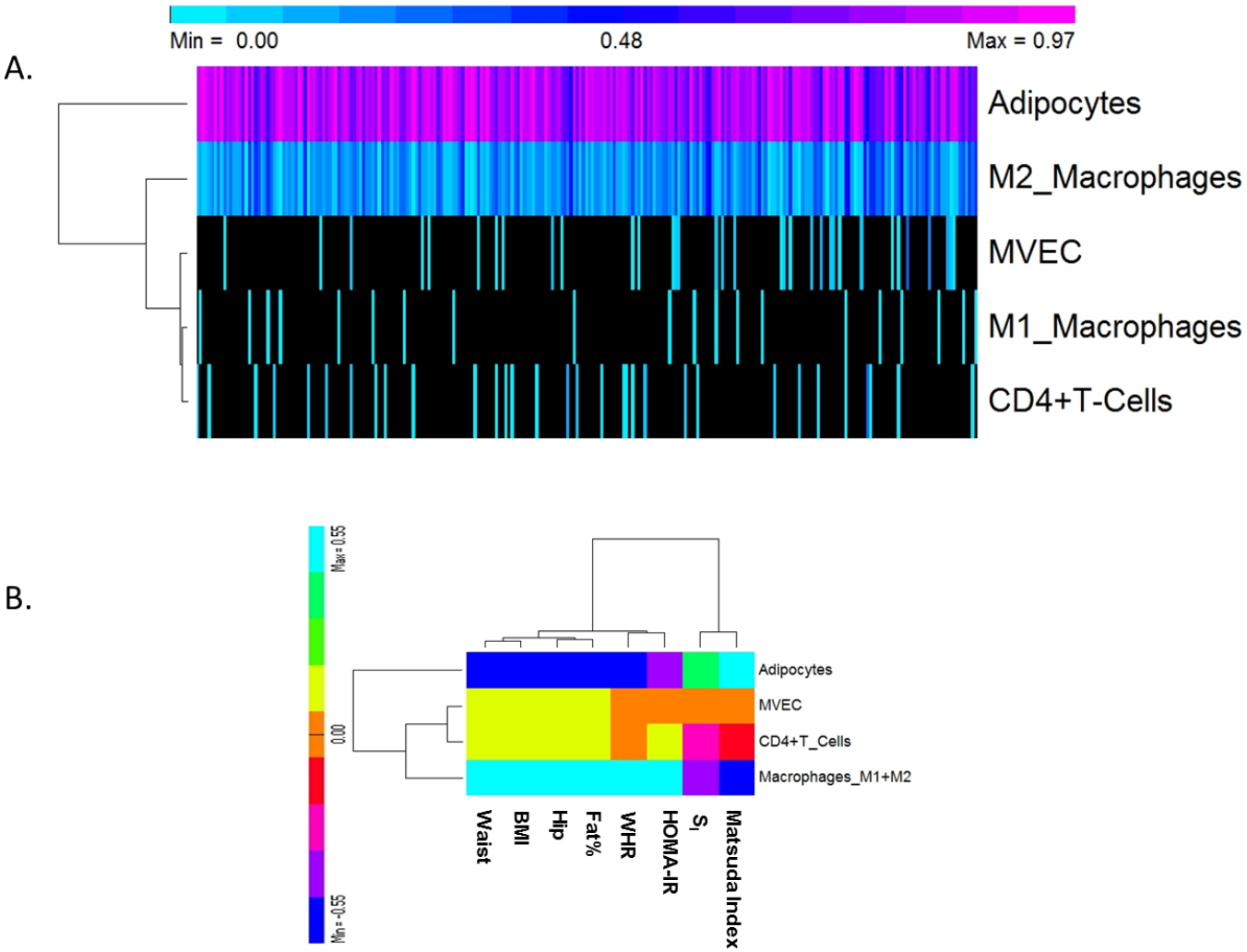
8. Orozco,LD, Farrell,C, Hale,C, Rubbi,L, Rinaldi,A, Civelek,M, Pan,C, Lam,L, Montoya,D, Edillor,C, Seldin,M, Boehnke,M, Mohlke,KL, Jacobsen,S, Kuusisto,J, Laakso,M, Lusis,AJ, Pellegrini,M: Epigenome-wide association in adipose tissue from the METSIM cohort. *Hum Mol Genet* 27:1830-1846, 2018

9. Sajuthi,SP, Sharma,NK, Chou,JW, Palmer,ND, McWilliams,DR, Beal,J, Comeau,ME, Ma,L, Calles-Escandon,J, Demons,J, Rogers,S, Cherry,K, Menon,L, Kouba,E, Davis,D, Burris,M, Byerly,SJ, Ng,MC, Maruthur,NM, Patel,SR, Bielak,LF, Lange,LA, Guo,X, Sale,MM, Chan,KH, Monda,KL, Chen,GK, Taylor,K, Palmer,C, Edwards,TL, North,KE, Haiman,CA, Bowden,DW, Freedman,BI, Langefeld,CD, Das,SK: Mapping adipose and muscle tissue expression quantitative trait loci in African Americans to identify genes for type 2 diabetes and obesity. *Hum Genet* 135:869-880, 2016

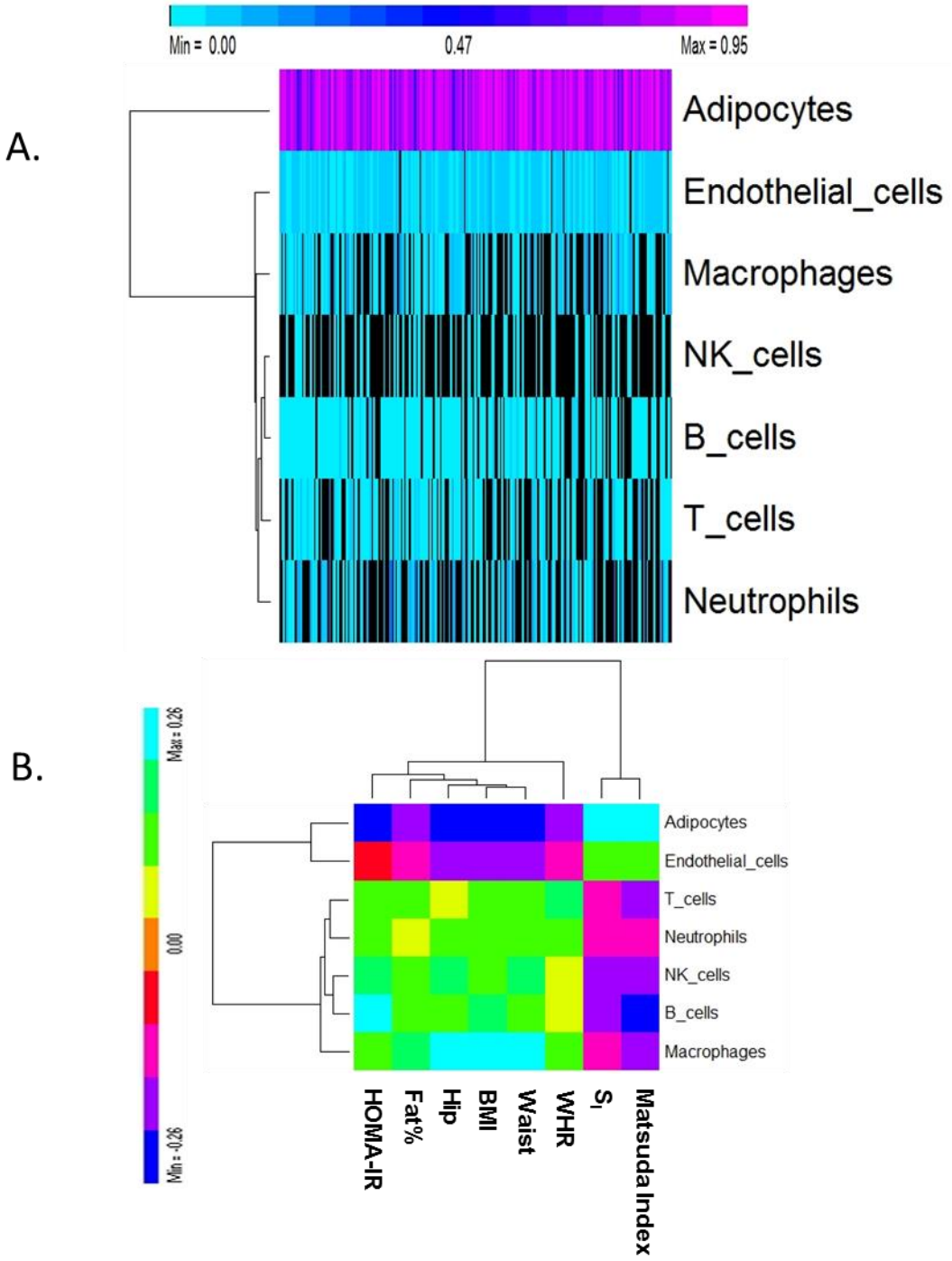
10. Sharma,NK, Chuang Key,CC, Civelek,M, Wabitsch,M, Comeau,ME, Langefeld,CD, Parks,JS, Das,SK: Genetic Regulation of Enoyl-CoA Hydratase Domain-Containing 3 in Adipose Tissue Determines Insulin Sensitivity in African Americans and Europeans. *Diabetes* 68:1508-1522, 2019

Supplementary Figures:

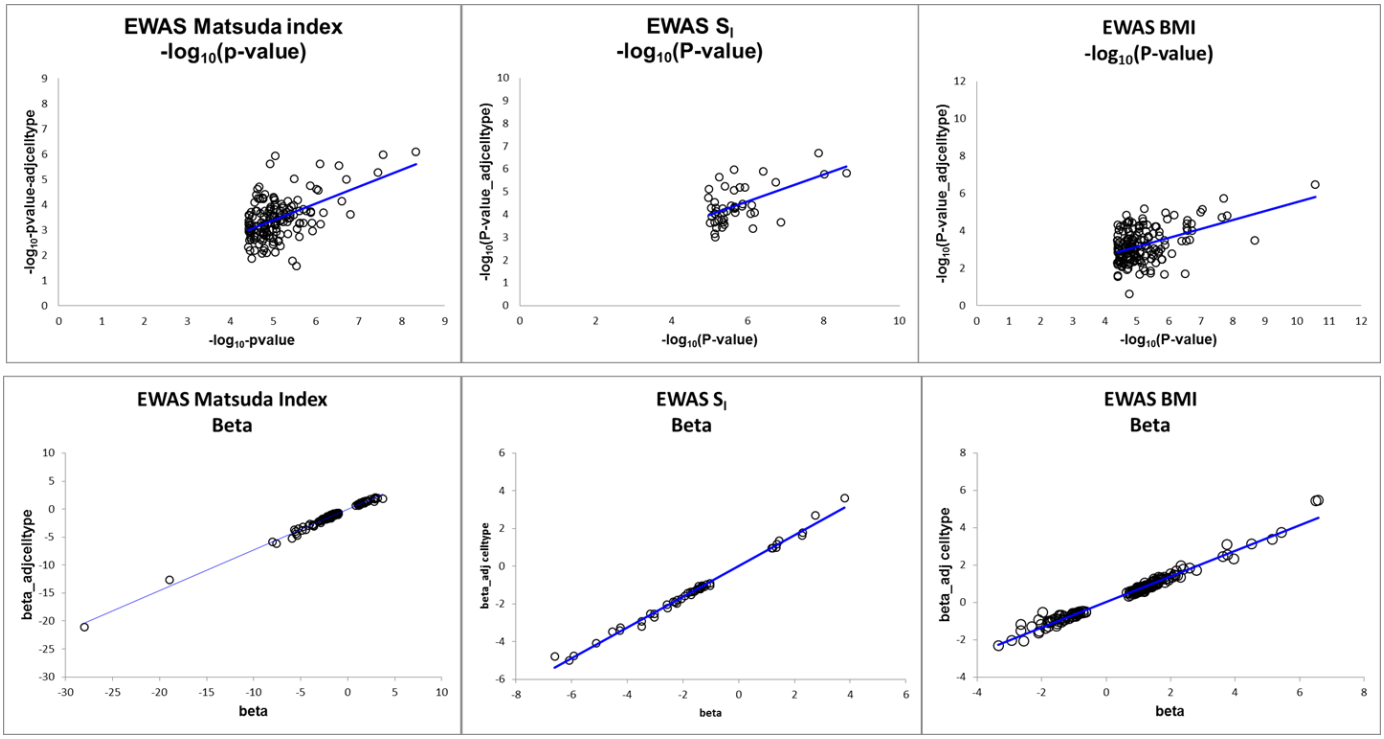
Supplementary Figure 1. ***In silico*** **deconvolution of adipose tissue cell types based on transcriptome data. A**) Heat map shows relative fraction of different cell-types (rows) across adipose tissue samples from 253 African American individuals in AAGMEx cohort (columns) estimated by CIBERSORTx analytical tool. The color in the heat map represents the relative fraction that each cell contributes to the total in each sample. **B**) Correlation between cell-type composition and glucometabolic phenotypes. The color in the heat map represents the partial correlation coefficient adjusted for age and gender.



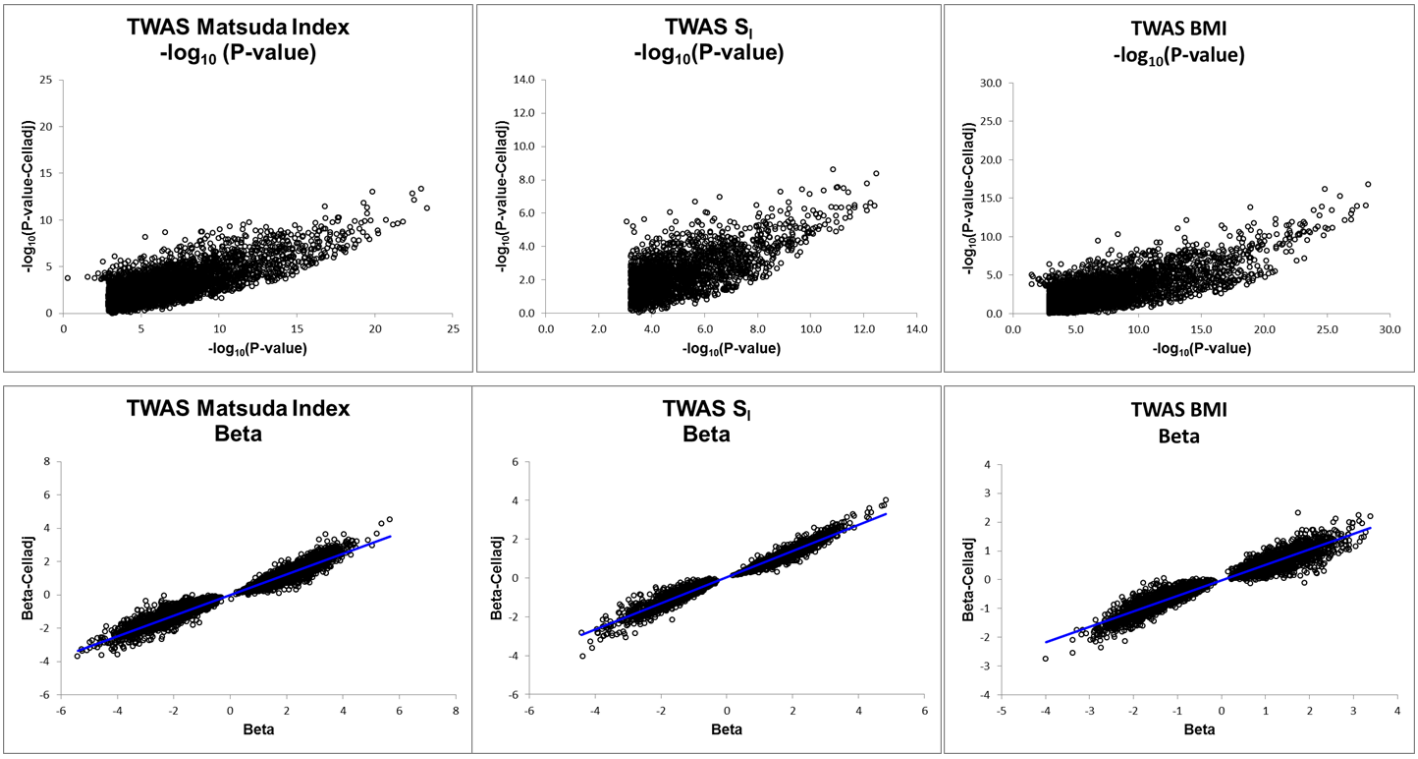
Supplementary Figure 2. ***In silico*** **deconvolution of adipose tissue cell types based on methylome data. A**) Heat map shows relative fraction of different cell-types (rows) across adipose tissue samples from 230 African American individuals in AAGMEx cohort (columns) estimated by reference-based algorithm based on CpG DNA methylation profile data. The color in the heat map represents the relative fraction that each cell contributes to the total in each sample. B) Correlation between cell-type composition and glucometabolic phenotypes. The color in the heat map represents the partial correlation coefficient adjusted for age and gender.



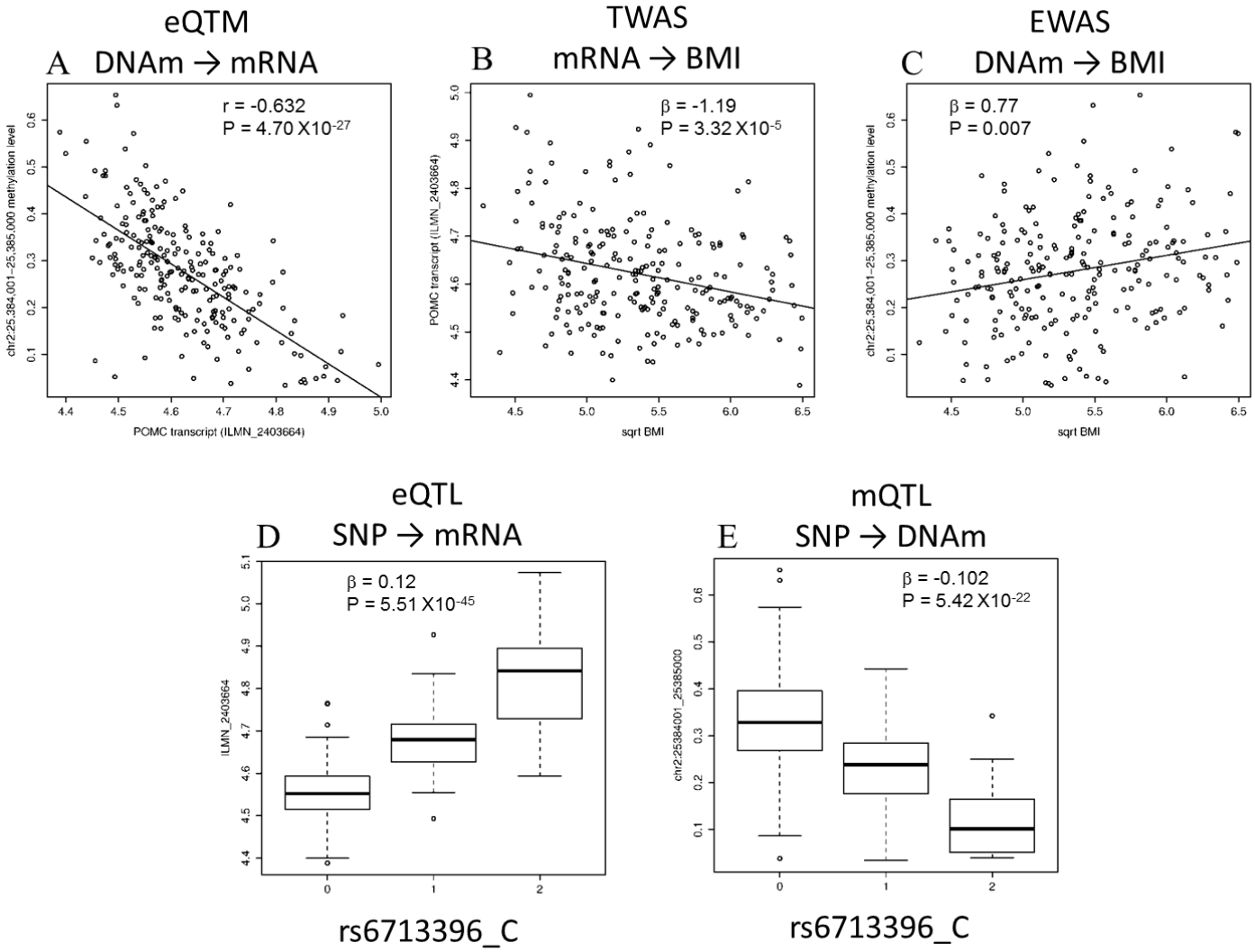
Supplementary Figure 3. Comparison of EWAS results for Matsuda index, S_I and BMI in cell-type adjusted and unadjusted analyses. The X-axis of the scatter plot indicates EWAS results from linear regression analysis adjusted for age, gender, and admixture, while the Y-axis indicates EWAS results from linear regression analysis adjusted for age, gender, admixture and cell-type proportions. The P-value, and Beta for CpG regions that were significantly associated (FDR-p≤0.05) with a trait in either cell-type adjusted or unadjusted analysis are shown.



Supplementary Figure 4. Comparison of TWAS results for Matsuda index, S_I and BMI in cell-type adjusted and unadjusted analyses. The X-axis of the scatter plot indicates TWAS results from linear regression analysis adjusted for age, gender, and admixture, while the Y-axis indicates TWAS results from in linear regression analysis adjusted for age, gender, admixture and cell-type proportions. The P-value, and Beta for transcript expression levels that were significantly associated (FDR-p≤0.05) with a trait in either cell-type adjusted or unadjusted analysis are shown.



Supplementary Figure 5. DNA Methylation level at chr2:25,384,001-25,385,000 CpG region determine BMI by regulating expression of POMC, and are genetically regulated in adipose tissue of AAGMEx participants. Scatter plots show **(A)** correlation of DNA methylation levels at chr2:25,384,001-25,385,000 CpG region with POMC transcript (ILMN_2403664) levels, **(B)** correlation of POMC transcript (ILMN_2403664) levels with BMI, and **(C)** correlation of DNA methylation levels at chr2:25,384,001-25,385,000 CpG region with BMI. Box plots showing the expression of *POMC* transcript (ILMN_2403664) and DNA methylation level in the top eQTM locus (CpG region chr2:25384001-25385000) in genotype groups for SNP rs6713396 (chr2: 25384705) and suggest **(D)** expression quantitative loci (eQTL), and **(E)** methylation quantitative trait loci (mQTL) for *POMC* in adipose tissue.



Supplementary Tables:

(in Microsoft –Excel file: Supplementary tables-eQTM-DB20-0117-R1-Diabetes.xlsx; can be viewed in Google drive link; https://drive.google.com/file/d/1sHZV526fttTjnkigyvebjO0p1Yq4rUIY/view?usp=sharing)

Supplementary Table-1: Summary of anthropometric and glucometabolic phenotypes in AAGMEx cohort

Supplementary Table-2A: *cis*-eQTM sites for glucometabolic trait-associated transcripts in adipose tissue.

Analysis identifies association between methylation level of adipose tissue CpG sites (independent variable) and adipose tissue transcript level (dependent variable). Models included age, sex and admixture as covariates.

Of the 7,464 probes for glucometabolic trait-associated transcripts, 7,461 were included in *cis*-eQTM analysis based on CpG sites within and ±100kb of the transcript. Final results focused on 6,774 high quality transcripts.

cis-eQTMs significant in linear regression analysis (FDR-P<0.05) and partial correlation r (Spearman's) ≥ +0.2 or ≤ -0.2 are shown.

Results are sorted by absolute value of partial correlation coefficient.

Genomic annotations of CpG sites were shown primarily from ANNOTATR analysis and adipose tissue/ adipocyte chromatin state annotation data (ChromeHMM, Core 15-state model). Additional annotation of CpG sites to SNPs are based on FALCAN (Fast Locus Annotator) and UCSC browser SNPbuild151. When CpG sites are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

RRBS sequence coverages are from Bismark and methylKit analysis.

Supplementary Table-2B: *cis*-eQTM regions for glucometabolic trait-associated transcripts in adipose tissue.

cis-eQTM analysis is based on CpG regions (tilling window of 1000bp region) within and ±100kb of the transcript.

Analysis identifies association between methylation level of adipose tissue CpG regions_1000bp (independent variable) and adipose tissue transcript level (dependent variable). Models included age, sex and admixture as covariates.

cis-eQTM regions significant in linear regression analysis (FDR-P<0.05) and partial correlation r (Spearman's) ≥ +0.2 or ≤ -0.2 are shown.

Results are sorted by absolute value of partial correlation coefficient.

Genomic annotations of CpG regions were shown from ANNOTATR analysis. When CpG regions are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

Supplementary Table-3: CpG sites associated with Matsuda insulin sensitivity index (from OGTT)

A total of 26 CpG sites are significantly associated at FDR-P<0.05. Results for CpG sites associated with Matsuda index at FDR-P<0.1 in epigenome wide association study (EWAS) are shown.

Results are sorted by EWAS p-value.

Analysis identifies association between methylation level of adipose tissue CpG sites (independent variable) and Matsuda index (dependent variable) in linear regression analysis. Models included age, sex and admixture as covariates.

Genomic annotations of CpG sites were shown primarily from ANNOTATR analysis and adipose tissue/ adipocyte chromatin state annotation data (ChromeHMM, Core 15-state model). Additional annotation of CpG sites to SNPs are based on FALCAN (Fast Locus Annotator) and UCSC browser SNPbuild151. When CpG sites are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

RRBS sequence coverages are from Bismark and methylKit analysis.

Supplementary Table-4: CpG sites associated with S_I (Insulin sensitivity index from FSIGT)

No CpG site is significantly associated at FDR-P<0.05. Results for CpG sites associated with S_I at FDR-P<0.1 in epigenome wide association study (EWAS) are shown.

Results are sorted by EWAS p-value.

Analysis identifies association between methylation level of adipose tissue CpG sites (independent variable) and S_I (dependent variable) in linear regression analysis. Models included age, sex and admixture as covariates.

Genomic annotations of CpG sites were shown primarily from ANNOTATR analysis and adipose tissue/ adipocyte chromatin state annotation data (ChromeHMM, Core 15-state model). Additional annotation of CpG sites to SNPs are based on FALCAN (Fast Locus Annotator) and UCSC browser SNPbuild151. When CpG sites are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

RRBS sequence coverages are from Bismark and methylKit analysis.

Supplementary Table-5: CpG sites associated with BMI (Body Mass Index)

A total of 15 CpG sites are significantly associated at FDR-P<0.05. Results for CpG sites associated with BMI at FDR-P<0.1 in epigenome wide association study (EWAS) are shown.

Results are sorted by EWAS p-value.

Analysis identifies association between methylation level of adipose tissue CpG sites (independent variable) and BMI (dependent variable) in linear regression analysis. Models included age, sex and admixture as covariates.

Genomic annotations of CpG sites were shown primarily from ANNOTATR analysis and adipose tissue/ adipocyte chromatin state annotation data (ChromeHMM, Core 15-state model). Additional annotation of CpG sites to SNPs are based on FALCAN (Fast Locus Annotator) and UCSC browser SNPbuild151. When CpG sites are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

RRBS sequence coverages are from Bismark and methylKit analysis.

Supplementary Table-6: CpG regions associated with Matsuda insulin sensitivity index (from OGTT)

Results for CpG regions (tilling window of 1000bp region) associated with Matsuda index at FDR-P<0.05 in epigenome wide association study (EWAS) are shown.

Results are sorted by EWAS p-value.

Analysis identifies association between methylation level of adipose tissue CpG regions (independent variable) and Matsuda index (dependent variable) in linear regression analysis. Models included age, sex and admixture as covariates.

Genomic annotations of CpG regions were shown from ANNOTATR analysis. When CpG regions are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

Supplementary Table-7: CpG regions associated with S_I (Insulin sensitivity index from FSIGT)

Results for CpG regions (tilling window of 1000bp region) associated with S_I at FDR-P<0.05 in epigenome wide association study (EWAS) are shown.

Results are sorted by EWAS p-value.

Analysis identifies association between methylation level of adipose tissue CpG regions (independent variable) and S_I (dependent variable) in linear regression analysis. Models included age, sex and admixture as covariates.

Genomic annotations of CpG regions were shown from ANNOTATR analysis. When CpG regions are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

Supplementary Table-8: CpG regions associated with BMI (Body Mass Index)

Results for CpG regions (tilling window of 1000bp region) associated with BMI at FDR-P<0.05 in epigenome wide association study (EWAS) are shown.

Results are sorted by EWAS p-value.

Analysis identifies association between methylation level of adipose tissue CpG regions (independent variable) and BMI (dependent variable) in linear regression analysis. Models included age, sex and admixture as covariates.

Genomic annotations of CpG regions were shown from ANNOTATR analysis. When CpG regions are included in multiple annotations, only one annotation was prioritized in the order of hg19_cpg_islands > cpg_shores cpg_shelves > cpg_inter (for distance to CpG island annotation) and hg19_genes_promoters > 5UTRs > exons > 3UTRs > intronexonboundaries > introns > genes_1to5kb > intergenic (for annotation to gene).

Supplementary Table-9: CpG sites associated with Matsuda index are also *cis*-eQTM site for Matsuda index-associated transcripts.

Integration of Matsuda Index CpG site EWAS, *cis*-eQTM site analysis, and transcript to Matsuda index association (TWAS) results.

CpG sites that are eQTM (linear regression analysis and partial correlation FDR-P<0.05 and partial correlation r+/-0.2) for a Matsuda index-associated transcript (TWAS, FDR-P<0.01) and CpG site is also associated with Matsuda index (EWAS, P<0.01) are shown.

Results are sorted by P_CIT (omnibus P-value from Causal Inference test).

Supplementary Table-10: CpG sites associated with S_I are also *cis*-eQTM sites for S_I- associated transcripts.

Integration of S_I CpG site EWAS, *cis*-eQTM site analysis, and transcript to S_I association (TWAS) results.

CpG sites that are eQTM (linear regression analysis and partial correlation FDR-P<0.05 and partial correlation r+/-0.2) for a S_I-associated transcript (TWAS, FDR-P<0.01) and are also associated with S_I (EWAS, P<0.01) are shown.

Results are sorted by P_CIT (omnibus P-value from Causal Inference test).

Supplementary Table-11: CpG sites associated with BMI are also *cis*-eQTM sites for BMI- associated transcripts.

Integration of BMI CpG site EWAS, *cis*-eQTM site analysis, and transcript to BMI association (TWAS) results.

CpG sites that are eQTM (linear regression analysis and partial correlation FDR-P<0.05 and partial correlation r+/-0.2) for a BMI-associated transcript (TWAS, FDR-P<0.01) and are also associated with BMI (EWAS, P<0.01) are shown.

Results are sorted by P_CIT (omnibus P-value from Causal Inference test).

Supplementary Table-12: CpG regions associated with Matsuda index are also *cis*-eQTM regions for Matsuda index-associated transcripts.

Integration of Matsuda Index CpG region (tilling window of 1000bp region) EWAS, *cis*-eQTM region analysis, and transcript to Matsuda index association (TWAS) results.

CpG regions that are are eQTM (linear regression analysis and partial correlation FDR-P<0.05 and partial correlation r+/-0.2) for a Matsuda index-associated transcript (TWAS, FDR-P<0.01) and are also associated with Matsuda insulin sensitivity index (EWAS, P<0.01) are shown.

Results are sorted by P_CIT (omnibus P-value from Causal Inference test).

Supplementary Table-13: CpG regions associated with S_I are also *cis*-eQTM regions for S_I-associated transcripts.

Integration of S_I CpG region (tilling window of 1000bp region) EWAS, *cis*-eQTM region analysis, and transcript to S_I association (TWAS) results.

CpG regions that are eQTM (linear regression analysis and partial correlation FDR-P<0.05 and partial correlation r+/-0.2) for a S_I-associated transcript (TWAS, FDR-P<0.01) and are also associated with S_I (EWAS, P<0.01) are shown.

Results are sorted by P_CIT (omnibus P-value from Causal Inference test).

Supplementary Table-14: CpG regions associated with BMI are also cis-eQTM regions for BMI-associated transcripts.

Integration of BMI CpG region (tilling window of 1000bp region) EWAS, *cis* -eQTM region analysis, and transcript to BMI association (TWAS) results.

CpG regions that are eQTM (linear regression analysis and partial correlation FDR-P<0.05 and partial correlation r+/-0.2) for a BMI-associated transcript (TWAS, FDR-P<0.01) and are also associated with BMI (EWAS, P<0.01) are shown.

Results are sorted by P_CIT (omnibus P-value from Causal Inference test).