SUPPLEMENTARY DATA

**"Alterations of a cellular cholesterol metabolism network is a molecular feature of obesity-related type 2 diabetes and cardiovascular disease"**

Jingzhong Ding, Lindsay M Reynolds, Tanja Zeller, Christian Müller, Kurt Lohman, Barbara J. Nicklas, Stephen B. Kritchevsky, Zhiqing Huang, Alberto de la Fuente, Nicola Soranzo, Robert E. Settlage, Chia-Chi Chuang, Timothy Howard, Ning Xu, Mark O. Goodarzi, Y.-D. Ida Chen, Jerome I. Rotter, David S. Siscovick, John S. Parks, Susan Murphy, David R. Jacobs Jr., Wendy Post, Russell P. Tracy, Philipp S. Wild, Stefan Blankenberg, Ina Hoeschele, David Herrington, Charles E. McCall, Yongmei Liu

**Supporting Document/Data**

***Table of Contents*** ***Page***

SUPPLEMENTARY DATA

## *Methods*

### Purification of monocytes

Monocytes were isolated with anti-CD14 monoclonal antibody coated magnetic beads, respectively, using AutoMACs automated magnetic separation unit (Miltenyi Biotec, Bergisch Gladbach, Germany). Centralized training of technicians, standardized protocols, and extensive quality control (QC) measures were implemented for collection, on-site processing, and shipment of MESA specimens and routine calibration of equipment. Blood was initially collected in sodium heparin-containing Vacutainer CPT$^{TM}$ cell separation tubes (Becton Dickinson, Rutherford, NJ) to separate peripheral blood mononuclear cells from other elements within 2 hours from blood draw. Subsequently, monocytes were isolated with anti-CD14 monoclonal antibody coated magnetic beads, respectively, using AutoMACs automated magnetic separation unit (Miltenyi Biotec, Bergisch Gladbach, Germany). Based on flow cytometry analysis of 18 specimens, monocyte samples were consistently > 90% pure.

### DNA/RNA extraction

DNA and RNA were isolated from samples simultaneously using the AllPrep DNA/RNA Mini Kit (Qiagen, Inc., Hilden, Germany). DNA and RNA QC metrics included optical density (OD) measurements, using a NanoDrop spectrophotometer and evaluation of the integrity of 18s and 28s ribosomal RNA using the Agilent 2100 Bioanalyzer with RNA 6000 Nano chips (Agilent Technology, Inc., Santa Clara, CA) following manufacturer's instructions. RNA with RIN (RNA Integrity) scores > 9.0 was used for global expression microarrays. The median of RIN for our 1,264 samples was 9.9.

### Global expression quantification

The Illumina HumanHT-12 v4 Expression BeadChip and Illumina Bead Array Reader were used to perform the genome-wide expression analysis, following the Illumina expression protocol. The Illumina TotalPrep-96 RNA Amplification Kit (Ambion/Applied Biosystems, Darmstadt, Germany) was used for reverse transcription, and amplification with 500 ng of input total RNA (at 11ul). 700 ng of biotinylated cRNA was hybridized to a BeadChip at $58^0$C for 16 – 17 hours. To avoid potential biases due to batch, chip, and position effects, a stratified random sampling technique was used to assign individual samples (including five common control sample for the first 480 samples) to specific BeadChips (12 samples/chip) and chip position.

### Epigenome-wide methylation quantification

The Illumina HumanMethylation450 BeadChip and HiScan reader were used to perform the epigenome-wide methylation analysis. The EZ-96 DNA Methylation$^{TM}$ Kit (Zymo Research, Orange, CA) was used for bisulfate conversation with 1ug of input DNA (at 45 ul). 4 µl of bisulfite-converted DNA were used for DNA methylation assays, following the Illumina Infinium HD Methylation protocol. This consisted of a whole genome amplification step followed by enzymatic end-point fragmentation, precipitation, and resuspension. The resuspended samples were hybridized on HumanMethylation 450 BeadChips at 48°C for 16 h. The individual samples were assigned to the BeadChips and to chip position using the same sampling scheme as that for the expression BeadChips.

### Quality control and pre-processing of microarray data

Data pre-processing and QC analyses were performed in *R* (http://www.r-project.org/) using *Bioconductor* (http://www.bioconductor.org/) packages. For expression data, data corrected for local background were obtained from Illumina's proprietary software GenomeStudio. QC analyses and bead type summarization (average bead signal for each type after outlier removal) were performed using the *beadarray* package (1). Detection P-values were computed using the negative controls on the array. The *neqc* function of the *limma* (2) package was used to perform a normal-exponential convolution model analysis to estimate non-negative signal, quantile normalization using all probes (gene and control, detected and not detected) and samples, addition of a recommended (small) offset, $\log_2$ transformation,

and elimination of control probe data from the normalized expression matrix. Multidimensional scaling plots showed the five common control samples were highly clustered together and identified three outlier samples, which were excluded subsequently.

The Illumina HumanHT-12 v4 Expression BeadChip included 48K transcripts. Statistical analyses were limited to probes retained after applying the following QC elimination: non-detectable expression in ≥90% of MESA samples using a detection p-value cut-off of 0.0001, overlap with a repetitive element or region, low variance across the samples (<$10^{th}$ percentile), or putative and/or not well-characterized genes, i.e. gene names starting with KIAA, FLJ, HS, Cxorf, MGC, or LOC. We included 14,619 gene transcripts (10,898 unique genes) for analysis.

Bead-level methylation data were summarized in *GenomeStudio*. Because a two-channel system and both Infinium I and II assays were used, normalization was performed in several steps using the *lumi* package (3). We first adjusted for color bias using "smooth quantile normalization". Next, the data were background adjusted by subtracting the median intensity value of the negative control probes. Lastly, data were normalized across all samples by standard quantile normalization applied to the bead-type intensities and combined across Infinium I and II assays and both colors. QC measures included checks for sex and race/ethnicity mismatches, and outlier identification by multidimensional scaling plots. The final methylation value for each methylation probe was computed as the M-value, essentially the log ratio of the methylated to the unmethylated intensity (4). The M-value is well suited for high-level analyses and can be transformed into a beta-value, an estimate of the percent methylation of an individual CpG site that ranges from 0 to 1 (thus M is logit(beta-value)).

The Illumina HumanMethylation450 BeadChip included probes for 485K CpG sites. Of these 485K CpG sites, 448,588 passed the QC elimination criteria including: "detected" methylation levels <90% of MESA samples using a detection p-value cut-off of 0.05, existence of any SNPs within 10 base pairs of the targeted CpG site, or overlap with a repetitive element or region.

Pre-processing with global normalization removed large position and chip effects across all probes; however, probe-specific chip effects were found for some CpG sites and gene transcripts, while probe-specific position effects existed for some CpG sites but were ignorable for all gene transcripts. These probe-specific effects were included as covariates in all subsequent analyses.

**Stability analysis enhanced weighted gene co-expression network identification**

To cluster the subset of 2,807 BMI-associated genes (FDR≤0.05) into network modules of highly correlated transcripts, we applied the weighted correlation network analysis as implemented in the R package WGCNA (5). With this method, we first obtained a weighted network based on the pairwise correlations among all transcripts considered, with an adjustment to produce a scale-free topology. Then, using the topological overlap measure to estimate the network interconnectedness, the transcripts were hierarchically clustered. With respect to the default parameters of WGCNA, we changed only the correlation type from Pearson to biweight midcorrelation (which is more robust to outliers) and set the minimum size for module detection to 10.

Briefly, WGCNA(5) begins with pairwise correlations among all genes which are transformed (continuously) using a power function to obtain a network with weighted edges. The value of the parameter of the power function is chosen such that the network satisfies a scale-free topology criterion. Using these weights, a similarity measure (termed the topological overlap measure) is evaluated for each pair of nodes (transcripts) as a function of the weights for edges between the two nodes and between either of the two nodes and any shared neighbors of the two nodes. Based on the resulting dissimilarity matrix, genes are hierarchically clustered into densely connected sub-networks termed modules which correspond to branches of the dendogram, as determined by a dynamic tree cutting method (5).

Unfortunately, the module structure identified by WGCNA tends to be rather unstable, even when the sample size is relatively large (in the hundreds). Stability of the module structure can be assessed by repeatedly making relatively small, random changes to the data and re-running the analysis,

and then assessing the agreement between the resulting structures. One way of making such changes to the data is by sampling random subsets of the data ("sub-sampling") which contain most but not all of the samples. We performed sub–sampling by obtaining a random sample of 80% of the observations (MESA participants), performing WGCNA on this data subset (with module detection) and repeating this process 200 times. Each of the 200 module assignments was represented by an unsigned network in which all transcripts assigned to the same module were connected by an edge. The Jaccard index (6) was used to evaluate the similarity between any two networks, which is equal to the number of edges shared by both networks divided by the total number of edges present in either network. Hence, the Jaccard index ranges from 0 to 1, with larger values indicating higher similarity between two networks. The values of the Jaccard index for the network constructed from the original data and any network obtained from a sub-sampled data set were low with mean value (across 200 replicates) in the range 0.25 - 0.30. To increase the stability of the module assignment, we calculated a consensus network composed of those edges which were present in at least 70% of the 200 networks constructed from the sub-sampled data sets. We then compared several consensus networks, each based on 200 sub-samples, resulting in Jaccard index values very near 0.90 and indicating much higher stability between consensus networks compared with the (in) stability of networks from individual sub-sampled datasets.

Using this novel approach of stability analysis enhanced WGCNA, we then computed, for each consensus module, the eigengene defined as the first right-singular vector in the singular value decomposition of the standardized $p_m$x$N$ expression matrix (or the first eigenvector of the $N$x$N$ correlation matrix, $N$ = 1,264) based on the $p_m$ genes within module $m$. Association analyses of the eigengenes of individual sub-networks (modules) with T2D and CAC were performed using logistic and linear regression, respectively (**Table S4**). CAC was log transformed (log (Agatston score + 1)). We fit separate models with the first eigenvector for each network as a predictor of T2D or coronary artery calcium. Covariates were age, gender, race/ethnicity, study site, expression/methylation chip, methylation position (for CpG methylation analyses only), residual sample contamination.

**RNA sequencing**

Total RNA samples were enriched for mRNA, by depleting rRNA using the MICROBExpress kit from Ambion and following the manufacturer's instructions. Poly(A) mRNA was enriched, and Illumina compatible, strand-specific libraries were constructed using Illumina's TruSeq Stranded mRNA HT Sample Prep Kit (Illumina, RS-122-2103). 1 ug of total RNA with RIN ≥ 8.0 was converted into a library of stranded template molecules suitable for subsequent cluster generation and sequencing by Illumina HiSeq. The libraries generated were validated using Agilent 2100 Bioanalyzer and quantitated using Quant-iT dsDNA HS Kit (Invitrogen) and qPCR. Six individually indexed cDNA libraries were pooled and sequenced on Illumina HiSeq, resulting in an average of close to 30 million reads per sample. Libraries were clustered onto flow cells using Illumina's TruSeq PE Cluster Kit v3 (PE-401-3001) and sequenced 2X100 cycles using TruSeq SBS Kit -HS (FC-401-3001) on an Illumina HiSeqTM 2500. A total of 64 lanes were run to generate approximately 30 million 2 x 101 Paired End reads per sample. The Illumina HiSeq Control Software (HCS v2.0.12) with Real Time Analysis (RTA v1.3.61) was used to provide the management and execution of the HiSeq 2500.

Illumina sequencing runs were processed to de-multiplex samples and generate FastQ files using the Illumina provided *configureBclToFastq.pl* script to automate running CASAVA 1.8.4 using default parameters for removal of sequencing reads failing the chastity filter (yes) and mismatches in the barcode read (0). Following generation of FastQ files, reads were trimmed to remove poor quality reads (or read tails) using *Btrim* (5 base sliding window average with $Q > 15$) (7) and then trimmed to remove any adaptor sequence present in the reads using custom perl scripts (trim sequences containing 11 base tag of adaptor, final length >40 bases). The *Ensembl* GRCh37 *Homo Sapiens* reference file, annotations and Bowtie2 indexes were downloaded from the *igenomes.com* website (10-Apr-2013) for mapping of the sequencing reads to the genome and read counting. *Bowtie2 (2.1.0)* and *TopHat2 (2.0.8)* were used

to map the sequencing reads to the genome using a mate-inner-distance of 100 bp and '*firststrand*' options (8;9) . Following alignment, *bam* files were merged using the *samtools* (0.1.19) merge function (10), and read counts per gene were obtained using *HTSeq* (0.5.4p3) (http://www-huber.embl.de/users/anders/HTSeq /doc/ overview.html). The '*intersection-strict*' overlap resolution mode and '*stranded reverse*' options were used in *HTSeq*.

Data pre-processing and QC analyses were performed in *R* (http://www.r-project.org/) using *Bioconductor* (http://www.bioconductor.org/) packages. The transcript-based raw count data files for each sample from *TopHat2* were combined into a count matrix with 56,303 features (rows) and 374 MESA samples (columns). The median total count per sample was 28.8 million. Reads denoted by *TopHat2* as "no_feature","ambiguous", "too_low_aQual", "not_aligned", "alignment_not_unique" were removed. Counts were converted to Counts Per Million (CPM) using the *cpm* function of the *edgeR* package (11) , and all features with CPM ≤ 0.25 in ≥90% of the 374 MESA samples were removed. Features assigned to the mitochondrial genome were removed as well. Using the *biomaRt* package and querying the *Ensembl BioMart* database, *Entrez Gene ID*s, Gene Symbols, genome coordinates, gene length and percent GC content were obtained for 12,585 features which had a corresponding *Entrez ID or* Illumina HumanHT-12 v4 probe ID. To be able to continue to use the flexible and computationally efficient linear modeling functions in *R,* we transformed the raw count data to log counts per million ($y$ = logCPM) as recommended by Law et al (2013) (12):

$$y_{gs} = log_2\left(\frac{c_{gs} + 0.5}{T_s + 1} \cdot 10^6\right)$$

where $c_{gs}$ is the raw count of gene transcript $g$ in sample $s$, and $T_s$ is the normalized total count of sample $s$, using the Trimmed Mean of M-values (TMM) normalization method (13) as implemented in the *calcNormFactors* function in the *edgeR Bioconductor* package (11). We either performed only this TMM normalization, or we applied quantile normalization (QN) to the logCPM values. Because the logCPM values' variance tends to decrease with increasing count for smaller counts, we used the voom function of the *limma* package (14) to estimate the mean-variance trend non-parametrically and to predict the residual variance of each individual observation for each gene. Then we incorporated the inverse residual variances into the linear modeling (*lm*) as weights in a standard manner. For the logCPM data, we imposed the same low variance filter that we had used for the microarray data, removing another 192 features with the lowest variance and retaining 12,380 features for analysis. We then performed weighted linear model analyses with the otherwise exact same models as for the microarray data.

**Replication study**

The Gutenberg Health Study (GHS) is designed as a population-based, prospective, observational single-centre cohort study in the Rhine-Main-Region in western Mid-Germany to evaluate and improve cardiovascular risk stratification (15) . The sample was drawn randomly from the governmental local registry offices in the city of Mainz and the district of Mainz-Bingen. The sample was stratified 1:1 for gender and for urban and rural residence with equal strata for decades of age. Individuals between 35 and 74 years of age were enrolled, and written informed consent was obtained from all participants. Exclusion criteria were insufficient knowledge of the German language, and physical or psychological inability to participate in the examinations at the study centre. Blood was drawn under standardized conditions after an overnight fasting period. Routine laboratory parameters were measured immediately after blood withdrawal by standardized methods. DNA and RNA were isolated immediately after blood withdrawal and stored at -80°C until further analyses. Study protocol and sampling design were approved by the local ethics committee, and by local and federal data safety commissioners.

Separation of monocytes was conducted within 60 min after blood collection as described in *Zeller et al, 2010* (16). Briefly, 8 mL blood was collected using the Vacutainer CPT Cell Preparation

Tube System (BD, Heidelberg, Germany) and 400 µL RosetteSep Monocyte Enrichment Cocktail (StemCell Technologies, Vancouver, Canada) was added immediately after blood collection. This cocktail contains antibodies directed against cell surface antigens on human hematopoietic cells (CD2, CD3, CD8, CD19, CD56, CD66b) and glycophorin A on red blood cells. Total RNA was extracted the same day using Trizol extraction and purification by silica-based columns. The integrity of the total RNA was assessed through analysis on an Agilent Bioanalyzer 2100 (Agilent Technologies, Böblingen, Germany). Samples with a RIN less than seven were excluded from further analyses.

Genome-wide expression assessment was performed using the Illumina HT-12 v3 BeadChip as described previously (17) . Pre-processing of data, quantile normalization and log-2 transformation were performed in *R* using *Bioconductor* package *lumi*. Based on expression patterns of probes localised on the X and Y chromosome, we discarded samples which did not match the recorded sex. All 48,803 probes of the BeadChip array were used for further analyses.

Associations between transcripts and BMI were assessed by applying linear regression models using R function *lm* with the probe expression as dependent and BMI as independent variables. The models were adjusted for sex, age and for the technical covariates RIN, storage time and plate layout as described previously (17) . To account for potential cell contamination, we additionally adjusted for 7 surrogate variables reflecting the degree of contamination by different blood cell types (CD4+, CD8+, CD19+, CD56+, CD66b+, erythroblasts and megakaryocytes) (18). Association p-values were corrected for multiple testing using the FDR based method of the *Bioconductor* package *qvalue(19)* . Only probes with a FDR<0.05 were considered significant.

**Bisulfite treatment of genomic DNA and pyrosequencing**

A subset of 176 samples were randomly selected from the 1,264 MESA monocyte samples. Genomic DNA of monocytes were bisulfite modified using the EZ DNA Methylation™ Kit (Zymo Research Co., Irvine, CA) according to the manufacturer's protocol for the Infinium methylation assay. Primers (F: 5'-AAGGTTTGGGGTTATTTTAGTGG-3') for pyrosequencing assays of *ABCG1-cg06500161* were designed using PSQ assay design software version 1.0.6 (Biotage, Uppsala, Sweden): PCR was performed with the PyroMark PCR Master Mix (Qiagen Inc.). Pyrosequencing was conducted using PyroMark Gold Q98 Reagents (Qiagen Inc.). Methylation values for each CpG site were calculated using Pyro Q-CpG software 1.0.9 (Biotage).

SUPPLEMENTARY DATA

**Supplementary Table S1.** Population characteristics for all participants included in analysis and by race in the Multi-Ethnic Study of Atherosclerosis (MESA)

| Variable | All (n = 1,264) | Caucasian (n = 590) | Hispanic (n = 402) | African-American (n = 272) |
|---|---|---|---|---|
| Age (years) | 69 ± 10 | 69 ± 10 | 68 ± 9 | 70 ± 9 |
| Women | 650 (51%) | 285 (48%) | 202 (50%) | 163 (60%) |
| Former smoker | 358 (50%) | 183 (53%) | 104 (49%) | 71 (44%) |
| Current smoker | 69 (10%) | 31 (9%) | 18 (8%) | 20 (12%) |
| BMI ($kg/m^2$) | 29.6 ± 5.5 | 28.7 ± 5.4 | 30.1 ± 5.4 | 30.6 ± 5.7 |
| Hypertension | 596 (61%) | 278 (55%) | 163 (58%) | 155 (78%) |
| Diabetic | 289 (23%) | 81 (14%) | 106 (26%) | 102 (38%) |

Mean ± standard deviation provided for continuous variables; count (percentage) provided for discrete variables

SUPPLEMENTARY DATA

**Supplementary Table S2a.** Top ten BMI-associated gene transcripts (FDR<0.05)

| Official Gene Symbol | Illumina ID | Chr | Beta | SE | P-value | FDR |
|---|---|---|---|---|---|---|
| *GNG10* | 1757074 | 9 | 0.011 | 0.001 | 3.44E-23 | 3.01E-19 |
| *RRAGD* | 1699772 | 6 | 0.012 | 0.001 | 1.87E-17 | 5.77E-14 |
| *C1ORF24* | 1667966 | 1 | 0.014 | 0.002 | 1.97E-17 | 5.77E-14 |
| *GAS6* | 1784749 | 13 | 0.023 | 0.003 | 9.88E-17 | 2.17E-13 |
| *TPST2* | 2329679 | 22 | 0.011 | 0.001 | 4.57E-16 | 8.01E-13 |
| *ABCG1* | 1794782 | 21 | -0.031 | 0.004 | 1.07E-15 | 1.57E-12 |
| *GAS6* | 1779558 | 13 | 0.021 | 0.003 | 2.12E-15 | 2.56E-12 |
| *MYLIP* | 1656111 | 6 | -0.012 | 0.001 | 2.33E-15 | 2.56E-12 |
| *FCGBP* | 2302757 | 19 | -0.043 | 0.005 | 3.55E-15 | 3.46E-12 |
| *CD1D* | 1719433 | 1 | -0.008 | 0.001 | 6.15E-15 | 5.39E-12 |

Effect size (Beta), standard error (SE), and significance (P-value; FDR) from associations of BMI and mRNA expression (Official Gene Symbol; Illumina ID) in 1,263 monocyte samples (FDR<0.05)

SUPPLEMENTARY DATA

**Supplementary Table S2b.** Sex and race subgroup analyses of the associations between the top ten BMI-associated gene transcripts and waist circumference

| | All (N=1,263) | | Men (N=619) | | Women (N=645) | | Caucasians (N=590) | | Hispanics (N=402) | | African Americans (N=272) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gene** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** |
| **GNG10** | 0.004 | 1.8E-22 | 0.004 | 5.2E-09 | 0.004 | 1.2E-15 | 0.004 | 1.5E-09 | 0.004 | 4.7E-08 | 0.005 | 4.1E-08 |
| **RRAGD** | 0.004 | 3.5E-18 | 0.004 | 9.6E-07 | 0.004 | 4.4E-12 | 0.004 | 8.1E-09 | 0.005 | 1.1E-08 | 0.003 | 2.8E-03 |
| **C1ORF24** | 0.005 | 9.8E-18 | 0.006 | 2.0E-09 | 0.005 | 8.4E-10 | 0.004 | 2.4E-07 | 0.006 | 2.8E-07 | 0.006 | 1.9E-06 |
| **GAS6 (1784749)** | 0.007 | 1.4E-14 | 0.009 | 8.4E-09 | 0.006 | 3.4E-07 | 0.006 | 1.4E-05 | 0.007 | 8.5E-05 | 0.011 | 4.2E-07 |
| **TPST2** | 0.004 | 7.9E-17 | 0.005 | 1.8E-09 | 0.003 | 3.8E-09 | 0.004 | 1.4E-08 | 0.005 | 4.0E-08 | 0.003 | 6.5E-03 |
| **ABCG1** | -0.012 | 2.3E-18 | -0.012 | 5.7E-08 | -0.011 | 1.1E-11 | -0.014 | 2.2E-13 | -0.007 | 5.2E-03 | -0.013 | 2.7E-05 |
| **GAS6 (1779558)** | 0.007 | 5.8E-15 | 0.009 | 1.6E-09 | 0.006 | 4.5E-07 | 0.005 | 1.2E-05 | 0.007 | 1.4E-04 | 0.011 | 5.1E-08 |
| **MYLIP** | -0.004 | 1.2E-14 | -0.004 | 5.2E-06 | -0.004 | 3.5E-10 | -0.005 | 2.4E-10 | -0.005 | 6.7E-07 | -0.001 | 2.6E-01 |
| **FCGBP** | -0.015 | 3.8E-14 | -0.012 | 3.5E-05 | -0.016 | 1.1E-09 | -0.013 | 6.5E-09 | -0.016 | 2.7E-05 | -0.019 | 4.7E-04 |
| **CD1D** | -0.002 | 1.4E-11 | -0.003 | 6.0E-07 | -0.002 | 3.3E-06 | -0.002 | 1.5E-03 | -0.004 | 1.0E-07 | -0.003 | 4.7E-04 |

Effect size (Beta) and significance (P-value) from associations of waist circumference and mRNA expression (Official Gene Symbol; Illumina ID) in 1,263 monocyte samples

SUPPLEMENTARY DATA

**Supplementary Table S3.** Enriched Gene Ontology (GO) terms for 1,741 BMI-associated genes

| GO term | Count | Fold Enrichment | P-value |
|---|---|---|---|
| Mitochondrial transport | 21 | 2.2 | 6.16E-04 |
| Regulation of apoptosis | 123 | 1.3 | 8.30E-04 |
| Inflammatory response | 47 | 1.5 | 3.90E-03 |
| Glucose metabolic process | 30 | 1.7 | 4.61E-03 |
| Lipid biosynthetic process | 49 | 1.4 | 9.05E-03 |

Gene set enrichment analyses results (using DAVID; P-value<0.05) included 1,741 genes with expression associated (FDR<0.05) with BMI in monocytes from 1,263 MESA participants, with a background of 14,619 expressed gene transcripts

SUPPLEMENTARY DATA

**Supplementary Table S4a.** Association of each gene co-expression network eigengene with BMI, T2D, and CAC

| | | BMI | | | T2D | | CAC | |
|---|---|---|---|---|---|---|---|---|
| Network module | Gene count | Beta* | P-value | $r^2$ | OR† | P-value | Beta‡ | P-value |
| **cyan** | 11 | 1.30 | 1.60E-18 | 0.06 | 1.71 | **1.82E-09**§ | 1.28 | **2.51E-04** |
| **pink** | 25 | 1.27 | 7.09E-18 | 0.05 | 1.38 | **3.23E-04** | 1.20 | 5.53E-03 |
| **grey60** | 12 | 0.99 | 2.68E-11 | 0.03 | 1.32 | **1.01E-03** | 1.12 | 1.03E-02 |
| **midnightblue** | 14 | 0.98 | 3.54E-11 | 0.03 | 1.49 | **2.47E-06** | 1.11 | 1.12E-01 |
| **greenyellow** | 18 | 0.96 | 1.36E-10 | 0.03 | 1.48 | **4.73E-05** | 1.17 | 1.89E-02 |
| tan | 16 | 0.90 | 1.27E-09 | 0.03 | 1.23 | 1.32E-02 | 1.22 | 3.28E-03 |
| **yellow** | 42 | 0.90 | 1.51E-09 | 0.03 | 1.20 | 3.42E-02 | 1.23 | **2.38E-03** |
| **magenta** | 20 | -0.77 | 2.54E-07 | 0.02 | 0.75 | **9.87E-04** | 0.95 | 5.32E-01 |
| turquoise | 576 | 0.76 | 3.90E-07 | 0.02 | 1.21 | 2.31E-02 | 1.03 | 5.73E-01 |
| brown | 61 | -0.77 | 4.07E-07 | 0.02 | 0.80 | 1.21E-02 | 0.84 | 1.54E-02 |
| **black** | 29 | -0.76 | 4.38E-07 | 0.02 | 0.88 | 1.30E-01 | 0.80 | **1.37E-03** |
| lightcyan | 13 | -0.74 | 7.41E-07 | 0.02 | 0.95 | 4.83E-01 | 0.83 | 4.84E-03 |
| salmon | 16 | 0.73 | 9.29E-07 | 0.02 | 1.06 | 4.55E-01 | 1.02 | 7.54E-01 |
| purple | 19 | 0.72 | 1.72E-06 | 0.02 | 1.19 | 4.00E-02 | 1.03 | 6.77E-01 |
| blue | 63 | 0.67 | 7.39E-06 | 0.01 | 1.26 | 1.92E-02 | 1.14 | 5.55E-02 |

* Changes in eigengene score per unit ($kg/m^2$) increment in body mass index (BMI)
† Odds ratio of type 2 diabetes (T2D) per standard deviation (SD) increment in eigengene score
‡ Fold change in coronary artery calcium (CAC) per SD increment in eigengene score
§ Significant *p* values were bolded for associations with T2D and CAC (using a Bonferroni adjusted significance threshold of $p<0.003$ for testing 15 modules).

SUPPLEMENTARY DATA

**Supplementary Table S4b.** Association of the 11 genes in the cholesterol metabolism network module (cyan) with T2D, CAC, IL-6, and CRP

| | T2D | | CAC | | IL-6 | | CRP | |
|---|---|---|---|---|---|---|---|---|
| **Gene** | **OR*** | **P-value** | **Beta†** | **P-value** | **Beta‡** | **P-value** | **Beta‡** | **P-value** |
| ABCG1 | 0.52 | 3.65E-12 | 0.82 | 3.69E-03 | -0.070 | 2.27E-02 | -0.074 | 2.03E-05 |
| ABCA1 | 0.62 | 3.93E-08 | 0.85 | 1.51E-02 | 0.009 | 7.20E-01 | -0.038 | 6.69E-03 |
| FADS1 | 1.56 | 4.31E-07 | 1.05 | 4.46E-01 | 0.050 | 5.92E-04 | 0.017 | 4.05E-02 |
| MYLIP | 0.65 | 1.21E-06 | 0.86 | 3.18E-02 | -0.042 | 4.83E-04 | -0.040 | 6.59E-09 |
| SQLE | 1.54 | 4.88E-06 | 1.23 | 5.07E-03 | 0.086 | 2.09E-06 | 0.044 | 2.59E-05 |
| LDLR | 1.42 | 9.29E-05 | 1.30 | 2.20E-04 | 0.096 | 4.33E-07 | 0.028 | 1.15E-02 |
| SC4MOL | 1.38 | 1.80E-04 | 1.20 | 7.43E-03 | 0.079 | 1.40E-06 | 0.045 | 1.65E-06 |
| SCD | 1.31 | 1.13E-03 | 1.20 | 8.04E-03 | 0.127 | 1.28E-07 | 0.038 | 5.90E-03 |
| CYP51A1 | 1.28 | 3.96E-03 | 1.10 | 1.79E-01 | 0.047 | 1.43E-03 | 0.030 | 4.49E-04 |
| FDFT1 | 1.25 | 6.47E-03 | 1.26 | 5.91E-04 | 0.019 | 1.06E-02 | 0.009 | 3.20E-02 |
| HMGCS1 | 1.21 | 2.77E-02 | 1.08 | 2.77E-01 | 0.013 | 1.23E-01 | 0.002 | 6.42E-01 |

* Odds ratio of type 2 diabetes (T2D) per standard deviation (SD) increment in each gene transcript
† Fold change in coronary artery calcium (CAC) per SD increment in each gene transcript
‡ Changes in gene transcript per unit increment in logIL-6 or logCRP

SUPPLEMENTARY DATA

**Supplementary Table S5.** Enriched gene sets within the BMI-associated gene network modules

| GO/KEGG Term | Up-regulated Genes | Down-regulated Genes | Fold Enrichment | P-value | FDR |
|---|---|---|---|---|---|
| ***Cyan module*** | | | | | |
| Sterol/cholesterol metabolic process (GOTERM_BP_FAT) | *LDLR, CYP51A1, SQLE, HMGCS1, FDFT1, SC4MOL, FADS1, SCD* | *ABCA1, ABCG1, MYLIP* | 87 | 2.28E-13 | 2.83E-10 |
| ***Magenta module*** | | | | | |
| Antigen processing and presentation (KEGG_PATHWAY) | | *CIITA, RFX5, HLA-DRB4, HLA-DPA1, HLA-DMB, HLA-DOA, HLA-DMA, CD74, HLA-DRA* | 52 | 8.48E-14 | 5.76E-11 |
| ***Brown module*** | | | | | |
| ncRNA metabolic process (GOTERM_BP_FAT) | | *IARS, METTL1, EXOSC5, NPM3, MARS2, POP5, NHP2, TRUB2, RPP40, DTD1* | 9 | 1.11E-06 | 1.55E-03 |
| ***Black module*** | | | | | |
| Ribosome (GOTERM_CC_FAT) | | *MRPS27, RPL32, RPL13A, RPL15, RPL5, RPL7A, RPL12, RPS4X, MRPL45, APEX1, DAP3* | 19 | 2.61E-11 | 2.65E-08 |
| ***Lightcyan module*** | | | | | |
| Translation (GOTERM_BP_FAT) | *RPL4* | *RSL1D1, EIF4B, RPL6, RPL22, RPL3, EEF2* | 16 | 7.53E-07 | 8.17E-04 |

Results from gene set enrichment analysis are presented (from DAVID) for co-expressed gene networks in 1,263 monocyte samples with significantly enriched with GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) terms (FDR < 0.05); no significant enrichment was observed for the other 10 gene networks; background included 14,619 expressed gene transcripts

SUPPLEMENTARY DATA

**Supplementary Table S6. Sex and race** subgroup analyses of the associations between the 11 genes in the cholesterol metabolism network module (cyan) and BMI

| | All (N=1,263) | | Men (N=619) | | Women (N=645) | | Caucasians (N=590) | | Hispanics (N=402) | | African Americans (N=272) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Gene** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** | **Beta** | **P-value** |
| ***ABCG1*** | -0.030 | 1.5E-15 | -0.032 | 1.5E-06 | -0.031 | 2.1E-09 | -0.043 | 4.1E-13 | -0.021 | 3.6E-03 | -0.019 | 5.8E-02 |
| ***MYLIP*** | -0.012 | 4.0E-15 | -0.010 | 3.0E-04 | -0.013 | 7.0E-11 | -0.015 | 9.8E-10 | -0.016 | 4.7E-09 | 0.002 | 5.2E-01 |
| ***SQLE*** | 0.017 | 3.5E-14 | 0.019 | 7.3E-06 | 0.015 | 1.4E-07 | 0.021 | 6.8E-09 | 0.016 | 6.2E-04 | 0.014 | 1.5E-02 |
| ***SC4MOL*** | 0.012 | 2.1E-09 | 0.010 | 6.3E-03 | 0.013 | 1.1E-06 | 0.014 | 2.9E-05 | 0.012 | 1.6E-03 | 0.009 | 6.4E-02 |
| ***LDLR*** | 0.014 | 2.4E-09 | 0.016 | 1.5E-04 | 0.011 | 6.3E-04 | 0.016 | 2.3E-05 | 0.014 | 5.7E-03 | 0.012 | 5.6E-02 |
| ***SCD*** | 0.016 | 2.6E-07 | 0.020 | 3.7E-04 | 0.011 | 4.1E-03 | 0.013 | 6.7E-03 | 0.021 | 7.1E-04 | 0.020 | 1.2E-02 |
| ***CYP51A1*** | 0.009 | 3.1E-07 | 0.009 | 6.1E-03 | 0.008 | 8.0E-04 | 0.012 | 3.9E-05 | 0.010 | 1.1E-02 | 0.007 | 1.0E-01 |
| ***ABCA1*** | -0.015 | 2.4E-06 | -0.020 | 4.7E-04 | -0.013 | 1.3E-03 | -0.027 | 3.9E-08 | -0.004 | 5.2E-01 | -0.005 | 5.1E-01 |
| ***FADS1*** | 0.009 | 3.5E-06 | 0.007 | 2.4E-02 | 0.009 | 4.1E-04 | 0.007 | 1.3E-02 | 0.009 | 1.0E-02 | 0.010 | 3.2E-02 |
| ***FDFT1*** | 0.004 | 1.2E-05 | 0.003 | 5.0E-02 | 0.005 | 1.7E-04 | 0.005 | 3.2E-04 | 0.006 | 1.2E-03 | -0.002 | 6.1E-01 |
| ***HMGCS1*** | 0.003 | 8.6E-04 | 0.004 | 2.0E-02 | 0.002 | 1.9E-01 | 0.003 | 3.8E-02 | 0.001 | 5.9E-01 | 0.003 | 1.9E-01 |

Beta: changes in gene expression levels of monocytes per $kg/m^2$ increment in BMI
Results obtained using linear regression, adjusting for age, sex or race when combined, study site, and residual sample contamination with non-monocytes; the interaction terms between gender (ethnicity) and BMI were not significant (P-value >0.05)

SUPPLEMENTARY DATA

**Supplementary Table S7.** Validation of the cholesterol metabolism network module (cyan) gene expression associations with BMI using RNA sequencing

| Gene | Entrez ID | Beta | SE | P-value |
|---|---|---|---|---|
| *ABCG1* | 9619 | -0.036 | 0.007 | 1.07E-06 |
| *MYLIP* | 29116 | -0.012 | 0.003 | 3.23E-06 |
| *SQLE* | 6713 | 0.022 | 0.005 | 7.58E-06 |
| *FADS1* | 3992 | 0.012 | 0.003 | 4.83E-04 |
| *HMGCS1* | 3157 | 0.006 | 0.002 | 1.36E-03 |
| *ABCA1* | 19 | -0.020 | 0.006 | 1.56E-03 |
| *LDLR* | 3949 | 0.015 | 0.005 | 2.03E-03 |
| *CYP51A1* | 1595 | 0.015 | 0.005 | 4.66E-03 |
| *SCD* | 6319 | 0.017 | 0.006 | 6.99E-03 |
| *FDFT1* | 2222 | 0.004 | 0.002 | 5.63E-02 |

RNA sequencing used to estimate mRNA abundance of cyan module genes in a subset of MESA monocyte samples (N = 374); associations performed using linear regression adjusting for age, gender, race, study site, and residual sample contamination with non-monocytes; cyan module gene (SC4MOL) is not included due to failure to pass the quality control requirement that the reads be uniquely mapped.

SUPPLEMENTARY DATA

**Supplementary Table S8.** Replication of gene expression-BMI associations in the Gutenberg Heart Study (N = 1,285)

| Gene | Illumina ID | Beta | SE | P-value | FDR |
|---|---|---|---|---|---|
| *ABCA1* | 1766054 | -0.025 | 0.003 | 3.89E-15 | **1.85E-12** |
| *ABCG1* | 1794782 | -0.012 | 0.002 | 6.52E-13 | **1.99E-10** |
| *MYLIP* | 1656111 | -0.011 | 0.002 | 3.33E-11 | **5.81E-09** |
| *SC4MOL* | 1720889 | 0.008 | 0.002 | 4.13E-08 | **2.20E-06** |
| *SQLE* | 2041293 | 0.010 | 0.002 | 1.85E-07 | **7.49E-06** |
| *CYP51A1* | 1664718 | 0.006 | 0.002 | 2.43E-03 | **0.02** |
| *FDFT1* | 1741096 | 0.004 | 0.001 | 4.50E-03 | **0.03** |
| *LDLR* | 2053415 | 0.005 | 0.002 | 0.03 | 0.09 |
| *HMGCS1* | 1797728 | 0.002 | 0.002 | 0.17 | 0.27 |
| *SCD* | 1689329 | 0.002 | 0.002 | 0.22 | 0.31 |
| *FADS1* | 1670134 | 0.002 | 0.002 | 0.28 | 0.35 |

Association (beta, standard error – SE) and significance (P-value, genome-wide false discovery rate – FDR) between gene expression and BMI for the cholesterol metabolism network module (cyan) genes in monocytes from 1,285 German participants of the Gutenberg Heart Study (GHS); BMI-expression associations replicating FDR<0.05 are shown in bold

SUPPLEMENTARY DATA

**Supplementary Table S9.** Changes in the 11 genes in the cholesterol metabolism network module (cyan) during weight loss intervention and their correlations with changes in the HOMA measure of insulin resistance in 16 obese adults

| | Expression changes during weight loss | | Correlation coefficients* with HOMA | |
|---|---|---|---|---|
| | Percent changes | p | Correlation coefficients | p |
| SQLE | -0.13 | 0.04 | 0.22 | 0.47 |
| MYLIP | 0.05 | 0.06 | -0.55 | 0.05 |
| SCD | -0.11 | 0.06 | 0.33 | 0.28 |
| FDFT1 | -0.03 | 0.07 | 0.66 | 0.01 |
| LDLR | -0.06 | 0.19 | 0.5 | 0.08 |
| CYP51A1 | -0.05 | 0.26 | 0.5 | 0.08 |
| HMGCS1 | -0.04 | 0.27 | 0.29 | 0.34 |
| ABCA1 | 0.05 | 0.28 | -0.55 | 0.05 |
| ABCG1 | 0.06 | 0.4 | -0.32 | 0.28 |
| FADS1 | -0.01 | 0.74 | 0.02 | 0.95 |
| SC4MOL | -0.01 | 0.83 | 0.12 | 0.69 |

*Pearson product-moment correlation coefficient

SUPPLEMENTARY DATA

**Supplementary Table S10.** Associations of the cyan module eigengene with T2D and CAC by statin use

| | T2D | | | CAC | | |
|---|---|---|---|---|---|---|
| Statin use | OR * | P-value | P for interaction | Beta † | P-value | P for interaction |
| No | 1.63 | 2.36E-04 | 0.42 | 1.08 | 3.74E-01 | 0.23 |
| Yes | 1.32 | 4.75E-02 | | 1.23 | 5.66E-02 | |

* Odds ratio of type 2 diabetes (T2D) per standard deviation (SD) increment in eigengene score
† Fold change in coronary artery calcium (CAC) per SD increment in eigengene score

SUPPLEMENTARY DATA

**Supplementary Table S11.** *Cis*-gene expression-associated CpG methylation sites (eMS) linked with expression of the cholesterol metabolism network (cyan) module genes

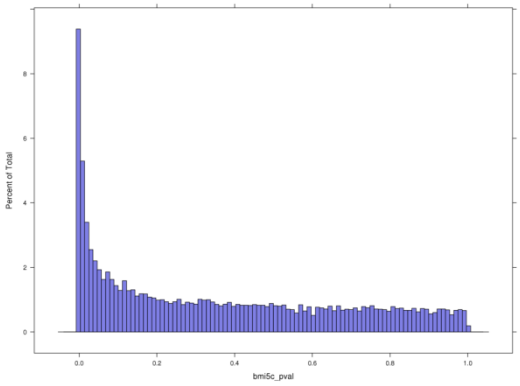
| eMS | | Methylation-gene expression | | |
|---|---|---|---|---|
| **CpG ID** | **CpG location** | **Gene (Illumina ID)** | **prho** | **FDR** |
| cg13571501 | Body (ABCA1) | ABCA1 (ILMN_1766054) | -0.19 | 7.31E-09 |
| cg06500161 | Body (ABCG1) | ABCG1 (ILMN_1794782) | -0.29 | 7.98E-23 |
| cg01881899 | Body (ABCG1) | ABCG1 (ILMN_1794782) | -0.23 | 5.30E-14 |
| cg10192877 | Body (ABCG1) | ABCG1 (ILMN_1794782) | -0.19 | 1.05E-08 |
| cg27243685 | Body (ABCG1) | ABCG1 (ILMN_1794782) | -0.18 | 5.27E-08 |
| cg02370100 | Body (ABCG1) | ABCG1 (ILMN_1794782) | -0.17 | 1.80E-06 |
| cg07397296 | Body (ABCG1) | ABCG1 (ILMN_1794782) | -0.13 | 7.90E-04 |
| cg22117172 | TSS1500 (CYP51A1) | CYP51A1 (ILMN_1664718) | -0.15 | 3.49E-05 |
| cg01689657 | TSS1500 (CYP51A1) | CYP51A1 (ILMN_1664718) | -0.13 | 8.35E-04 |
| cg01400685 | Body (FADS1) | FADS1 (ILMN_1670134) | -0.22 | 1.21E-11 |
| cg00603274 | Body (FADS1) | FADS1 (ILMN_1670134) | -0.15 | 5.49E-05 |
| cg13008174 | Body (FDFT1) | FDFT1 (ILMN_2144088) | -0.14 | 3.73E-04 |
| cg18564928 | | FDFT1 (ILMN_2144088) | -0.13 | 7.66E-04 |
| cg19751789 | TSS200 (LDLR) | LDLR (ILMN_2053415) | -0.23 | 6.09E-14 |
| cg07960944 | TSS1500 (LDLR) | LDLR (ILMN_2053415) | -0.22 | 8.04E-12 |
| cg12013591 | TSS200 (LDLR) | LDLR (ILMN_2053415) | -0.19 | 7.61E-09 |
| cg03567652 | 3'UTR (LDLR) | LDLR (ILMN_2053415) | 0.17 | 7.07E-07 |
| cg05323251 | Body (LDLR) | LDLR (ILMN_2053415) | 0.14 | 1.04E-04 |
| cg18596381 | TSS200 (LDLR) | LDLR (ILMN_2053415) | -0.14 | 3.97E-04 |
| cg05119988 | 5'UTR (SC4MOL) | SC4MOL (ILMN_1720889) | -0.18 | 7.27E-08 |
| cg03440556 | Body (SCD) | SCD (ILMN_1689329) | -0.30 | 9.50E-24 |
| cg24503796 | Body (SCD) | SCD (ILMN_1689329) | -0.24 | 5.22E-15 |
| cg06400428 | Body (SCD) | SCD (ILMN_1689329) | -0.19 | 3.88E-09 |
| cg02237755 | Body (SCD) | SCD (ILMN_1689329) | -0.18 | 2.24E-07 |
| cg09984392 | 1stExon (SQLE) | SQLE (ILMN_2041293) | -0.14 | 1.50E-04 |
| cg14660676 | Body (SQLE) | SQLE (ILMN_2041293) | -0.14 | 3.64E-04 |

The partial correlation (prho) and genome-wide significance (FDR) between CpG methylation and mRNA expression of cholesterol network genes in monocytes (N=1,264), estimated using linear regression adjusting for age, sex, race, study site, and % residual cellular contamination with non-monocytes; some eMS were located within gene bodies (gene containing CpG in parentheses), untranslated regions (UTR) of genes, within 1,500 base pairs of a transcription start site (TSS1500); genes are ordered by gene name, and then the association of methylation with gene expression (FDR<0.001).

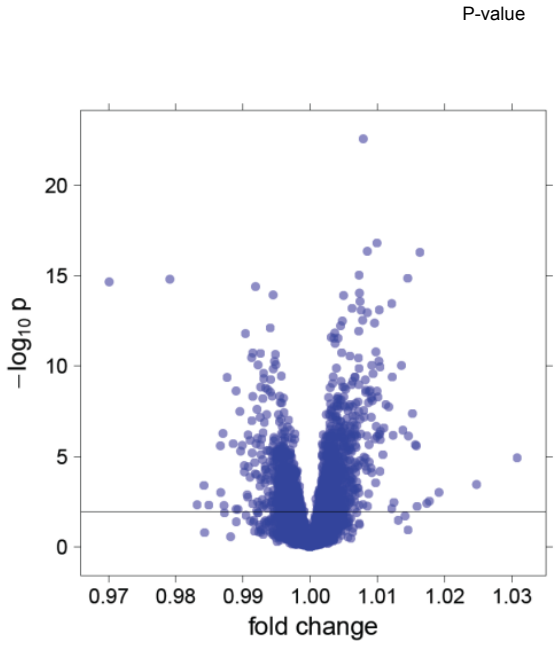
SUPPLEMENTARY DATA

**Supplementary Figure S1. BMI associations with the monocyte transcriptome. (A)** Histogram showing the percent of total associations plotted on the y-axis corresponding to the significance of mRNA expression associations with BMI, plotted on the x-axis. **(B)** Volcano plot showing the significance of mRNA expression associations with BMI ($-\log_{10}$ P-value) plotted on the y-axis and fold change of expression per 1 unit increase in BMI ($kg/m^2$) plotted on the x-axis; the black line represents gene expression and BMI association significance FDR<0.05 ($-\log_{10}$ P-value = 1.95). The fold change is derived from linear modeling of gene expression changes ($\log_2$ expression = a + beta x BMI ($kg/m^2$); fold change = $2^{beta}$ of P-values from 14,619 expressed gene transcripts (from 10,898 genes) tested for association with body mass index (BMI) in 1,263 monocyte samples using linear regression, adjusting for age, gender, race, study site, and residual sample contamination with non-monocytes.

**A**

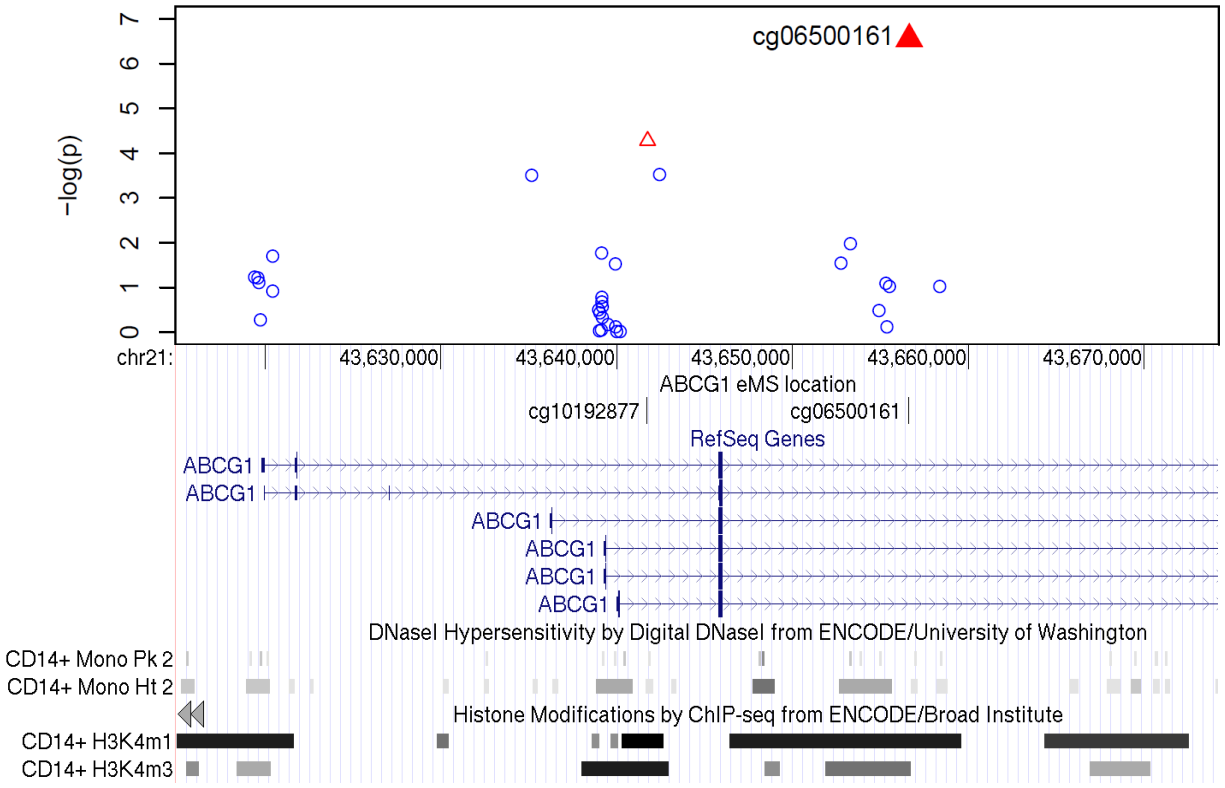



**B**



SUPPLEMENTARY DATA

**Supplementary Figure S2. Regional association plot between BMI and *ABCG1* methylation.** The significance of the associations detected between BMI and CpG methylation of *ABCG1* (ATP-binding cassette, sub-family G (WHITE), member 1) in 1,263 monocyte samples is shown on the y-axis ($-\log_{10}$ P-values) with the CpG genomic position plotted on the x-axis. Two CpG sites with methylation associated with BMI (FDR<0.05) were also *ABCG1* expression-associated methylation sites (*ABCG1*-eMS, shown as red triangles). Methylation of these two *ABCG1*-eMS are independently associated with decreasing expression of *ABCG1* expression (from multiple regression analysis, P-value<0.05). Methylation of 29 CpG sites in this region was not associated with BMI (FDR>0.05, shown as blue circles). Below the regional association plot is the surrounding genomic landscape (from UCSC genome browser, hg19), revealing the *ABCG1*-eMS overlap ENCODE-annotated features predicted to be regulatory, such as DNase-I sensitive regions (hotspots = Ht) and hypersensitive peaks (peaks = Pk) detected in monocytes, and histone methylation marks detected in monocytes indicative of enhancer and promoter regions (H3K4m1 and H3K4m3).



SUPPLEMENTARY DATA

## References


1. Dunning,MJ, Smith,ML, Ritchie,ME, Tavare,S: beadarray: R classes and methods for Illumina bead-based data. Bioinformatics 23:2183-2184, 2007
2. Smyth,GK, Michaud,J, Scott,HS: Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21:2067-2075, 2005
3. Du,P, Kibbe,WA, Lin,SM: lumi: a pipeline for processing Illumina microarray. Bioinformatics 24:1547-1548, 2008
4. Du,P, Zhang,X, Huang,CC, Jafari,N, Kibbe,WA, Hou,L, Lin,SM: Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11:587, 2010
5. Langfelder,P, Horvath,S: WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559, 2008
6. Diao,H, Li,X, Hu,S: The identification of dysfunctional crosstalk of pathways in Parkinson disease. Gene 515:159-162, 2013
7. Kong,Y: Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics 98:152-153, 2011
8. Langmead,B, Trapnell,C, Pop,M, Salzberg,SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25, 2009
9. Trapnell,C, Pachter,L, Salzberg,SL: TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111, 2009
10. Li,H, Handsaker,B, Wysoker,A, Fennell,T, Ruan,J, Homer,N, Marth,G, Abecasis,G, Durbin,R: The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079, 2009
11. Robinson,MD, McCarthy,DJ, Smyth,GK: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140, 2010
12. Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. Voom! Precision weights unlock linear model analysis tools for RNA-seq read counts. Preprint available at http://www.statsci.org/smyth/pubs/ . 5-1-2013.

Ref Type: Generic

13. Robinson,MD, Oshlack,A: A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25, 2010
14. Smyth GK: Limma: linear models for microarray data. In Bioinformatics and Computational Biology Solutions using R and Bioconductor. Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, Eds. New York, Springer, 2005, p. 397-420
15. Wild,PS, Zeller,T, Beutel,M, Blettner,M, Dugi,KA, Lackner,KJ, Pfeiffer,N, Munzel,T, Blankenberg,S: [The Gutenberg Health Study]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 55:824-829, 2012
16. Zeller,T, Wild,P, Szymczak,S, Rotival,M, Schillert,A, Castagne,R, Maouche,S, Germain,M, Lackner,K, Rossmann,H, Eleftheriadis,M, Sinning,CR, Schnabel,RB, Lubos,E, Mennerich,D, Rust,W, Perret,C, Proust,C, Nicaud,V, Loscalzo,J, Hubner,N, Tregouet,D, Munzel,T, Ziegler,A, Tiret,L, Blankenberg,S, Cambien,F: Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. PLoS One 5:e10693, 2010
17. Schurmann,C, Heim,K, Schillert,A, Blankenberg,S, Carstensen,M, Dorr,M, Endlich,K, Felix,SB, Gieger,C, Grallert,H, Herder,C, Hoffmann,W, Homuth,G, Illig,T, Kruppa,J, Meitinger,T, Muller,C, Nauck,M, Peters,A, Rettig,R, Roden,M, Strauch,K, Volker,U, Volzke,H, Wahl,S, Wallaschofski,H, Wild,PS, Zeller,T, Teumer,A, Prokisch,H, Ziegler,A: Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PLoS One 7:e50938, 2012

18. Rotival,M, Zeller,T, Wild,PS, Maouche,S, Szymczak,S, Schillert,A, Castagne,R, Deiseroth,A, Proust,C, Brocheton,J, Godefroy,T, Perret,C, Germain,M, Eleftheriadis,M, Sinning,CR, Schnabel,RB, Lubos,E, Lackner,KJ, Rossmann,H, Munzel,T, Rendon,A, Erdmann,J, Deloukas,P, Hengstenberg,C, Diemert,P, Montalescot,G, Ouwehand,WH, Samani,NJ, Schunkert,H, Tregouet,DA, Ziegler,A, Goodall,AH, Cambien,F, Tiret,L, Blankenberg,S: Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet 7:e1002367, 2011
19. Storey,JD, Tibshirani,R: Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100:9440-9445, 2003