UCLA

UCLA Previously Published Works

Title

Gene expression associations with body mass index in the Multi-Ethnic Study of Atherosclerosis

Permalink

https://escholarship.org/uc/item/5cq3c096

Journal

International Journal of Obesity, 47(2)

ISSN

0307-0565

Authors

Vargas, Luciana B
Lange, Leslie A
Ferrier, Kendra
et al.

Publication Date

2023-02-01

DOI

10.1038/s41366-022-01240-x

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

HHS Public Access

Author manuscript

Int J Obes (Lond). Author manuscript; available in PMC 2023 March 07.

Published in final edited form as:

Int J Obes (Lond). 2023 February ; 47(2): 109–116. doi:10.1038/s41366-022-01240-x.

Gene expression associations with body mass index in the Multi-Ethnic Study of Atherosclerosis

Luciana B. Vargas[1], **Leslie A. Lange**[1,2], **Kendra Ferrier**[1], **François Aguet**[3,4], **Kristin Ardlie**[3], **Stacey Gabriel**[5], **Namrata Gupta**[6], **Joshua D. Smith**[7], **Thomas W. Blackwell**[8], **Jingzhong Ding**[9], **Peter Durda**[10], **Russell P. Tracy**[10], **Yongmei Liu**[11], **Kent D. Taylor**[12], **W. Craig Johnson**[13], **Stephen S. Rich**[14], **Jerome I. Rotter**[12], **Ethan M. Lange**[1,15], **Iain R. Konigsberg**[1,✉]

[1]Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[2]Department of Epidemiology, University of Colorado School of Public Health, Aurora, CO, USA

[3]The Broad Institute of MIT and Harvard, Cambridge, MA, USA

[4]Illumina Artificial Intelligence Laboratory, Illumina, Inc, San Diego, CA, USA

[5]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA

[6]Genomics Platform, Broad Institute, Cambridge, MA, USA

[7]Northwest Genomics Center, University of Washington, Seattle, WA, USA

[8]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA

[9]Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[10]Department of Pathology and Laboratory Medicine, The Robert Larner, M.D. College of Medicine at University of Vermont, Burlington, VT, USA

[11]Divisions of Cardiology & Neurology, Department of Medicine, Duke University Medical Center, Durham, NC, USA

[12]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA

Reprints and permission information is available at http://www.nature.com/reprints

✉**Correspondence** and requests for materials should be addressed to Iain R. Konigsberg. iain.konigsberg@cuanschutz.edu.

AUTHOR CONTRIBUTIONS

LBV and IRK performed all statistical analysis, data visualization, and drafted the manuscript. LAL, KF, and EML contributed to conceptualizing analyses and critical editing of the manuscript. JDS, SG, and NG performed RNA-Sequencing. FA and KA processed RNA-Seq data to create the final MESA transcriptomics dataset. TWB provided advice and support as a member of the TOPMed Informatics Research Center. JD provided advice and support as a member of the MESA Multi-omics Adiposity Working Group. PD, RPT, YL, WCJ, SSR, JIR, and KDT designed the RNA-Seq study in MESA. All authors critically reviewed and approved the manuscript.

COMPETING INTERESTS

FA is an employee and shareholder of Illumina, Inc. The remaining authors have no competing interests to disclose.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41366-022-01240-x.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[13]Department of Biostatistics, University of Washington, Seattle, WA, USA

[14]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

[15]Department of Biostatistics and Informatics, University of Colorado School of Public Health, Aurora, CO, USA

Abstract

BACKGROUND/OBJECTIVES: Obesity, defined as excessive fat accumulation that represents a health risk, is increasing in adults and children, reaching global epidemic proportions. Body mass index (BMI) correlates with body fat and future health risk, yet differs in prediction by fat distribution, across populations and by age. Nonetheless, few genetic studies of BMI have been conducted in ancestrally diverse populations. Gene expression association with BMI was assessed in the Multi-Ethnic Study of Atherosclerosis (MESA) in four self-identified race and ethnicity (SIRE) groups to identify genes associated with obesity.

SUBJECTS/METHODS: RNA-sequencing was performed on 1096 MESA participants (37.8% white, 24.3% Hispanic, 28.4% African American, and 9.5% Chinese American) and linear models were used to assess the association of expression from each gene for its effect on BMI, adjusting for age, sex, sequencing center, study site, five expression and four genetic principal components in each self-identified race group. Sample-size-weighted meta-analysis was performed to identify genes with BMI-associated expression across ancestry groups.

RESULTS: Within individual SIRE groups, there were zero to three genes whose expression is significantly ($p < 1.97 \times 10^{-6}$) associated with BMI. Across all groups, 45 genes were identified by meta-analysis whose expression was significantly associated with BMI, explaining 29.7% of BMI variation. The 45 genes are expressed in a variety of tissues and cell types and are enriched for obesity-related processes including erythrocyte function, oxygen binding and transport, and JAK-STAT signaling.

CONCLUSIONS: We have identified genes whose expression is significantly associated with obesity in a multi-ethnic cohort. We have identified novel genes associated with BMI as well as confirmed previously identified genes from earlier genetic analyses. These novel genes and their biological pathways represent new targets for understanding the biology of obesity as well as new therapeutic intervention to reduce obesity and improve global public health.

INTRODUCTION

Obesity is a global pandemic with increasing prevalence and is associated with excess mortality and morbidity [1]. Obesity is a risk factor for many diseases, including cardiovascular disease, the leading cause of death globally [2]. Body mass index (BMI) is a commonly collected non-invasive anthropometric measure used as a proxy for fat mass to assess obesity. The World Health Organization defines obesity as a BMI greater than or equal to 30 kg/m^2.

Recent large-scale genome-wide association studies (GWAS) have uncovered hundreds of genetic associations with BMI [3–12], with a meta-analysis of approximately 700,000

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

individuals identifying 941 BMI-associated loci [11]. Despite these high-confidence genetic associations, in aggregate they only explain 6.0% of variance in BMI [11]; further, polygenic risk scores for BMI have low prediction accuracy. Additional studies are required to further disentangle the biological basis of human obesity.

As ’omics technologies have developed and matured, it is possible now to interrogate multiple ‘omics in thousands of individuals. Transcriptome-wide association studies (TWAS), originally using gene microarrays, but currently, RNA-sequencing (RNA-seq) approaches, have led to the discovery of associations between human traits and expression of discrete genes. While TWAS uses the GWAS methodologic framework, it can capture more phenotype-relevant associations. GWAS only captures causal genetic variant (typically single-nucleotide polymorphism, SNP) associations, and the functional impact of these variants is often unclear. Multiple studies of gene expression with BMI (as a continuous trait) and obesity (as a binary trait) have been published; however, these studies are often limited by small number of subjects, variable phenotype definitions, and limited genetic diversity (primarily from individuals of European descent).

Obesity rates differ by genetic ancestry, with 47% of African American and Hispanic individuals being classified as obese relative to 38% of individuals of European descent [13]. In order to identify potential SIRE-specific and shared gene expression effects on BMI, we leveraged a multi-ethnic cohort. The Multi-Ethnic Study of Atherosclerosis (MESA). MESA is a longitudinal cohort study initiated in 2000–2002 to investigate emergent risk factors for subclinical atherosclerosis and risk of cardiovascular disease [13]. In order to increase participant diversity, MESA recruited 6814 participants at six field centers across the U.S. that self-identify as white (38%), African American (28%), Hispanic (22%), and Chinese American (12%). We examined the association of gene expression from RNA-seq data with BMI in each self-identified race and ethnicity (SIRE) group. We further aggregated ancestry-specific results through meta-analysis to identify genes and pathways leading to the risk of obesity.

METHODS

Samples

Participants from MESA were included in this study. MESA is a longitudinal, community-based sample of 6814 men and women aged 45–84 years without evidence of clinical cardiovascular disease assembled to investigate the prevalence, risk factors, and progression of subclinical atherosclerosis. Participants were enrolled at six sites (Columbia University, New York, NY; Johns Hopkins University, Baltimore, MD; Northwestern University, Chicago, IL; University of California – Los Angeles, Los Angeles, CA; University of Minnesota, Twin Cities, Minneapolis and Saint Paul, MN; Wake Forest University, Winston Salem, NC) who self-identify as white (38%), African American (28%), Hispanic (22%), and Chinese American (12%) [14] between 2000 and 2002 [13]. Detailed clinical and biological data and samples were collected at baseline and across follow-up examinations. Blood samples were collected for biochemical risk factors and DNA extraction, while peripheral blood mononuclear cells (PBMCs) were viably cryopreserved for future studies (including RNA-seq). Anthropometric measures, including BMI, were collected at baseline

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and over the course of the five follow-up examinations. All participants provided informed consent to be included in MESA.

RNA-sequencing

MESA is a study contributing to the NHLBI Trans-Omics for Precision Medicine (TOPMed) program. In addition to whole genome sequence data currently available for ~4600 participants (in the same demographic proportions as the entire sample of 6814 subjects), the TOPMed MESA Multi-Omics project was designed to test whether archived samples could be used for generation of other ʻomics data (transcriptomics, methylation, metabolomics, proteomics). In MESA, PBMCs were collected at baseline (Exam 1, 2000–2002) and 10 years later (Exam 5, 2010–2012), cryopreserved, and stored at the MESA Biochemistry Laboratory at the University of Vermont.

RNA was extracted at the University of Vermont, using samples from cryopreserved PBMCs collected during Exam 1 (2000–2002). RNA-seq was conducted by two NHLBI-sequencing laboratories (the Northwest Genomics Center at the University of Washington and the Broad Institute of MIT and Harvard) on 1096 unique MESA participants using harmonized protocols. RNA libraries were prepared from at least 250 ng RNA using the Illumina TruSeq™ Stranded mRNA Kit and sequenced using the Illumina HiSeq 4000 (Illumina, San Diego, CA) platform, for a target depth of 40 M 2 × 101 bp paired-end reads. Alignment was performed using the TOPMed RNA-seq pipeline. Briefly, reads were aligned to GRCh38 with STAR [15] and collapsed to the gene-level using RNASEQC v2 [16] and the GENCODE 34 reference [17]. Comprehensive pipeline information is provided at https://github.com/broadinstitute/gtex-pipeline/blob/master/TOPMed_RNAseq_pipeline.md.

Normalization by library size was performed using the *estimateSizeFactors()* function in the *DESeq2* R package (v1.36.0) [18]. Principal component analysis was performed to identify sample outliers and major sources of variation. Samples at least 3 standard deviations from the mean of the first two principal components were excluded from analysis ($n = 5$). Genes with low expression were filtered using the *apply_filters()* function in the *Olivia* R package (v0.1.0) using default parameters. The final dataset for analysis included the expression of 25,416 genes [19].

Statistical analysis

In each SIRE group, linear regression models were fit to the expression (from RNA-seq) of each gene, testing for association with BMI (measured at Exam 1), adjusted for age, sex, sequencing center, study site, and principal components (4 genetic and 5 expression) using the *Olivia* R package [19]. P values were adjusted for bias and inflation using the *BACON* R package (v1.24.0) [20]. *BACON* is a Bayesian method developed for transcriptomic and epigenomic association studies, and it has been shown to perform better in these scenarios than methods devised for genomic association studies such as genomic control. Statistical significance was determined using a Bonferroni-adjusted threshold of $P = 1.97 \times 10^{-6}$ (25,416 tests/0.05). Sample-size-weighted meta-analysis within self-identified race/ancestry groups was performed using METAL [21]. The percent variation between SIRE groups for each gene due to heterogeneity, rather than chance, in the meta-analysis was assessed by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cochran's *Q* test [22] and the I^2 statistic [23]. The percentage of BMI variance explained by gene expression identified through meta-analysis was calculated using a partial R^2 approach implemented in the *rsq* R package (v2.5) [24], adjusted for age, sex, sequencing center, study site, and nine principal components (4 genetic and 5 expression).

Annotation

Genes whose expression was identified as significantly associated with BMI from our meta-analysis were further queried for previous association with BMI in the NHGRI-EBI GWAS Catalog [25], the GTEx portal (https://gtexportal.org/home/), and the TWAS Hub (http://twas-hub.org/). In the GWAS Catalog, 134 publications were identified; summary statistics were downloaded and GWAS SNPs annotated to the meta-analysis. Summary statistics from the 2018 GIANT consortium meta-analysis [11] were not available in the GWAS Catalog, but were included from the study web portal. SNPs annotated to genes from the BMI meta-analysis were interrogated for potential regulatory effects in the GTEx portal. Expression quantitative trait locus (eQTL) and splicing QTL (sQTL) targets of identified SNPs, as well as tissue- and cell-specific expression from GTEx bulk tissue and single-nuclear RNA-seq, were identified and included in our annotation. Results from PrediXcan models (https://github.com/hakyimlab/MetaXcan) predicting gene expression associations with BMI were downloaded from the TWAS Hub. The TWAS Hub models were generated from genetic data obtained across multiple studies [12, 26] and made publicly available (http://www.nealelab.is/uk-biobank).

Pathways and targets

Overrepresentation enrichment analysis was performed on genes that attained statistical significance from meta-analysis using clusterProfiler [27, 28]. Genes were tested for enrichment in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [29–31] and the Gene Ontology Resource (GO) Biological Process (GO:BP), Cellular Component (GO:CC) and Molecular Function (GO:MF) databases [32, 33]. Potential drug targets and disease associations were identified for the 45 significant genes using Open Targets (https://genetics.opentargets.org/).

RESULTS

Demographics

A total of 1091 MESA participants were included in the analyses. Participants self-identified as white (37.8%), African American (28.4%), Hispanic (24.3%), or Chinese American (9.53%). There were no statistically significant differences with respect to sex within SIRE group among participants; however, there were significant differences by age at Exam 1 (baseline) and BMI (Table 1). Those of Chinese American ancestry showed significantly lower BMI with reduced variation compared to participants in other SIRE groups, as observed previously.

SIRE-stratified analyses

Across groups, between 0 and 3 genes exhibited a statistically significant association of BMI with gene expression (Fig. 1, Supplementary Fig. 1, and Supplementary Table 1;

$p < 1.97 \times 10^{-6}$), independent of age, sex, study site, sequencing center, and principal components accounting for structure in genetics and gene expression. In African American subjects, *SOCS3* (suppressor of cytokine signaling 3) and *FAM20A* (FAM20A golgi-associated secretory pathway pseudokinase) were significantly associated with BMI. In Hispanic subjects, *CD209, JAK3* (Janus kinase 3), and *PIM1* (Pim-1 proto-oncogene, serine/threonine kinase) were significantly associated with BMI. In white subjects, *ARL6IP5* (ADP ribosylation factor like GTPase 6 interacting protein 5) and *JAK3* were significantly associated with BMI. No genes attained statistical significance with BMI in the Chinese American population, consistent with the smaller sample size relative to other populations.

Meta-analysis

Multi-ethnic sample-size-weighted meta-analysis was performed to identify significant associations of BMI with gene expression across SIRE groups. A total of 45 genes were identified with differential expression significantly associated with BMI after adjustment for bias and inflation [20] (Fig. 2, Supplementary Fig. 3, and Supplementary Table 2; BACON $p < 1.97 \times 10^{-6}$). Expression of the 45 meta-analysis significant genes explains 29.7% of BMI variation among individuals. The highest percentage of BMI variance explained was in Chinese American (39.9%), followed by African American (30.8%), Hispanic (30.6%), and white (26.4%) participants. Four of the five most highly associated genes were significant in at least one individual ancestry group: *JAK3* (Janus kinase 3) in white and Hispanic groups, *FAM20A* and *SOCS3* in the African American group, and *PIM1* (Pim-1 Proto-oncogene, serine/threonine kinase) in the Hispanic group. Although no genes showed significant expression-BMI associations in the Chinese American group, these 5 genes were among the 12 most associated in unadjusted analyses.

A total of 44 of 45 genes whose expression was significantly associated with BMI had a consistent direction of effect in each SIRE group (Supplementary Fig. 4). *NRG1* (Neuregulin 1) showed increased expression with increased BMI in the African American, Hispanic, and white populations, but decreased expression with increased BMI in the Chinese American population. Accordingly, *NRG1* and two other genes (*PIM1* and *SOCS2* (Suppressor Of Cytokine Signaling 2)) exhibited significant evidence of heterogeneity between groups. For *NRG1*, there was a strong support of gene expression associated with BMI in African American and Hispanic participants (4th most significant in each group), with modest support in white (gene rank #1840) and little support in Chinese American individuals (gene rank #23,121). *SOCS2* was strongly supported in the white, Chinese American (3rd most significant gene in each group), and Hispanic groups (4th most significant gene) but exhibited limited support in African American individuals (gene rank #1900). Lastly, *PIM1* showed the strongest support in Hispanic (3rd most significant gene) and Chinese American (6th most significant gene) groups but limited support in the white (gene rank #186) and African American (gene rank #204) groups. Of the five most significant associations of gene expression with BMI in the meta-analysis, only *GRAMD1B* (GRAM Domain Containing 1B) did not have an individual group with a significant association. However, *GRAMD1B* expression with BMI was ranked highly in white (#50), African American (#6), Hispanic (#10), and Chinese American analyses (#2), suggesting a global, rather than SIRE-specific effect of its expression with BMI.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In the GWAS Catalog [25], 134 publications report genetic (SNP) associations with BMI and obesity-related traits. Eight publications identify SNPs in *FLT3* (Fms Related Receptor Tyrosine Kinase 3), *NRG1, PACSIN1* (Protein Kinase C and Casein Kinase Substrate in Neurons 1), and *PTPRS* (Protein Tyrosine Phosphatase Receptor Type S) as associated with BMI [3–10]. SNPs in *PACSIN1* and *FLT3* were also present in the recent GIANT meta-analysis [11] (Supplementary Table 4). While these GWAS results suggest genetic variants may be associated with BMI, it is unclear if they act as eQTLs or sQTLs [34]. As shown in Supplementary Table 5, the *FLT3*-associated SNP is an eQTL for *FLT3* in many tissues (but not an sQTL), including subcutaneous adipose and multiple brain regions. The *PACSIN1*-associated SNP is an eQTL and sQTL in multiple tissues, although the target effector gene may not be *PACSIN1*. The SNP in *NRG1* associated with BMI is an eQTL in lung, liver, and whole blood, but it is not an sQTL. The *PTPRS* BMI-associated SNP is not an eQTL or an sQTL. The Open Targets database provides further evidence that these genes are involved in obesity biology, predicting 18 genes identified from meta-analysis are potential therapeutic targets for diabetes mellitus and 16 genes may be targets for obesity.

Pathways and enrichment analyses

Enrichment analyses using the KEGG (https://www.genome.jp/kegg/) and GO (http://geneontology.org/docs/go-enrichment-analysis/) resources (Fig. 3 and Supplementary Table 3) were conducted for the 45 genes exhibiting significant association of BMI with gene expression. Upregulated genes are enriched for KEGG pathways “type 2 diabetes mellitus”, “JAK-STAT signaling”, and the GO categories “hemoglobin complex”, “endocytic vesicle lumen”, “phosphatidylinositol 3-kinase complex”, “response to growth hormone”, “erythrocyte differentiation”, “erythrocyte homeostasis”, “myeloid cell homeostasis”, “oxygen carrier activity”, and “phosphatidylinositol 3-kinase regulator activity”. Downregulated genes are enriched for KEGG pathways “hematopoietic cell lineage” and GO categories “cytokine-cytokine receptor interaction”, “leukocyte proliferation”, “negative regulation of cytokine production”, “immune receptor activity”, “growth factor binding”, and “monosaccharide binding”.

Cells and tissue types related to the expression of genes identified from the BMI meta-analysis were interrogated using the Genotype-Tissue Expression (GTEx) Portal [34] (Supplementary Figs. 6–8). Some genes, including *HBA2* (Hemoglobin Subunit Alpha 2), *HBB* (Hemoglobin Subunit Beta), *SLC4A1* (Solute Carrier Family 4 Member 1 – Diego blood group), *AHSP* (Alpha hemoglobin stabilizing protein), and *ALAS2* (5′-Aminolevulinate Synthase 2), are highly expressed in whole blood. Of the 45 significantly associated genes, many have expression restricted to specific tissues (e.g., *SERPINF2* in liver). Consistent with the hypothesis of brain-body (obesity) crosstalk, regulation of appetite, and the overrepresentation of brain tissue in GTEx, many of the 45 significantly associated genes had the highest expression in brain tissues, including *PACSIN1* and *KCNH3* (Potassium Voltage-Gated Channel Subfamily H Member 3). As expected, a number of the 45 genes have expression in obesity-relevant cell types, including *GPX3* (Glutathione Peroxidase 3) and *PTPRS*, which are expressed primarily in adipocytes and fibroblasts. Other cell populations represented included immune cells (*SYTL3*, Synaptotagmin Like 3), dendritic cells and macrophages (*VSIG4*, V-Set and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Immunoglobuilin Domain Containing 4), mast cells (*IL18R1*, Interleukin 18 Receptor 1), neuromuscular-junction-rich myocytes (*FAM20A*), and neuroendocrine and neuronal cells (*NRG1*). These results suggest potential cellular and temporal heterogeneity of gene expression related to obesity and it associated traits.

Sensitivity analyses

Sensitivity analyses were conducted to compare the stratified meta-analysis to a unified modeling including all individuals and adjusting for SIRE as a covariate. In contrast to the 45 genes exhibiting a significant association of BMI with gene expression in the meta-analysis, the unified model identified a total of 56 genes. Of the 56 genes, 41 (73%) were shared with the meta-analysis, while 15 genes (*DAB2IP, DNASE1L3, HBD, IRF8, ITGA7, KCNK10, LILRA4, NRP1, P2RY2, P2RY6, RP11–76E17.3, RRAS, S100A8, SERPINF1, TEX2*) were different. Four of the genes in the meta-analysis were not seen in the unified model (*AHSP*, rank #22; *HDAC4* (Histone Deacetylase 4), rank #43; *IMPA2* (Inositol Monophosphatase 2, rank #45); and *PTGER2* (Prostaglandin E Receptor 2, rank #40)) and were mostly of lower ranking by significance (Supplementary Fig. 5). Genes unique to the meta-analysis or unified model did not show significant evidence of heterogeneity by SIRE group and displayed a consistent direction of effect across groups, with the exception of *TEX2* (Testis Expressed 2). In the unified analysis, *TEX2* did exhibit evidence of heterogeneity and showed a negative association with BMI only in the Chinese American group.

DISCUSSION

In this report, we identified genes whose expression in peripheral blood is associated with BMI in a study with participants of diverse genetic ancestries. Sample-size weighted meta-analysis of RNA-seq from peripheral blood identified 45 genes whose expression is significantly associated with BMI. These genes are enriched for a variety of processes relevant to obesity, including pathways relating to brain, heme, and erythrocyte function and cytokine signaling, expressed across a variety of tissues and cell types.

A number of genes identified in this study through SIRE-specific analyses have been implicated previously in obesity and obesity-related traits. *SOCS3* encodes a negative regulator of cytokine signaling induced by cytokines including IL6, IL10, and IFN-γ. SOCS3 protein binds multiple active tyrosine kinases, including IGF1 receptor, insulin receptor, and JAK2. *CD209* encodes a C-type lectin involved in cell adhesion and pathogen recognition and is highly expressed in antigen-presenting cells such as dendritic cells and macrophages. It has been shown that dendritic cells obtained from post-menopausal women with obesity and type 2 diabetes have elevated levels of *CD209* mRNA compared to normal-weight individuals with and without diabetes, as well as individuals with obesity but without diabetes [35], leading to the hypothesis that inflammation drives altered adhesion characteristics of dendritic cells. Decreased *CD209* levels were seen in peripheral blood, indicative of monocyte-derived inflammatory cells having migrated to tissue. Expression of *PIM1* has been shown to be upregulated in placenta of women with obesity and *PIM1* small interfering RNA knockdown in primary trophoblast cells results in a reduction of

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

pro-inflammatory cytokines. In mice, the loss of Jak3 results in increased body weight, chronic low-grade inflammation, and hyperinsulemia [36]. Mutations in *ARL6*, which physically interacts with *ARL6IP5*, result in the autosomal recessive disorder Bardet–Biedl syndrome-3, which is characterized by a history of obesity [37, 38].

Genes identified through meta-analysis also demonstrate biological relevance to obesity and related traits. *FLT3* driver mutations are significantly associated with obesity [39]. *CPT1A* has been implicated as a potential biomarker (from rat PBMC gene expression) for metabolically obese normal-weight syndrome; its expression was correlated with obesity phenotypes regardless of weight gain when rats were fed a high-fat diet [40]. Mice overexpressing *CPT1A* have increased hepatic fatty acid oxidation and reduced obesity-induced weight gain and inflammation [41]. *GPX3* regulates insulin receptor expression in adipose tissue [42] and its expression in adipose tissue is downregulated in subjects with type 2 diabetes and obesity [43]. Mice overexpressing *HDAC4* in adipocytes have reduced adiposity in response to either normal or high-fat diets, with *HDAC4* driving expansion of beige adipocytes [44]. HDAC4 protein levels are reduced in individuals with obesity and upregulated by 3 months of physical exercise [45]. *IL18R1* knockout mice have increased weight and reduced glucose and insulin tolerance [46]. NRG1 administered to mice results in reduced food intake and weight gain and increased leptin levels [47]. *SOCS2*-deficient mice exhibit increased adipose tissue mass with reduction in multiple cytokines, increased M2 macrophages (that resolve inflammation), and regulatory T cells [48]. *VSIG4* inhibits macrophage activation and resulting inflammation, and *Vsig4* knockout mice display obesity following administration of a high-fat diet [49].

In our meta-analysis, multiple genes whose expression is associated with BMI are involved in heme and erythrocyte processes, as well as JAK/STAT signaling. JAK/STAT signaling mediates cellular responses to cytokines, growth factors, and hormones. Leptin signaling requires JAK/STAT signal transduction and leptin binding to its receptor drives JAK/STAT-mediated transcription of proopiomelanocortin that drives cortisol production [50]. Mutations in *JAK2* have been associated with increased adiposity and waist circumference as well as protection from metabolic syndrome [51, 52]. Whole-body and specific-tissue knockouts of various JAK/STAT components display changes in body weight and measures of adiposity, metabolism, and insulin resistance [53]. Hemoglobin concentrations, higher in individuals with obesity, are consistent with the significance of hemoglobin-related genes (*ALAS2*, *HBA2*, and *HBB*) upregulated in skeletal muscle of rhesus macaques fed a high-fat western-style diet [54]. Hemin treatment increases adipocyte differentiation in mouse fibroblast cell lines [55], and obesity alters red blood cell function; thus, individuals with obesity have higher rates of hemolysis and altered adhesiveness and aggregation [56, 57].

The 45 genes identified through meta-analysis explain 29.7% of BMI variance in this study. This estimate is much higher than the 6.0% of variance explained by genetic associations and is indicative of the different architectures of genetic and gene expression data. While genetic associations with phenotypes are causal except in rare situations, gene expression studies capture the effects of phenotype on gene expression. While this results in transcriptomic studies often failing to identify causal targets in disease, this additional information uncovered by these studies can provide valuable insights into biology and

Author Manuscript

can be leveraged for biomarkers and patient stratification. Recent studies incorporating transcriptional risk scores into predictive disease models have shown better prediction than genetic risk scores alone [58]. This also explains the limited overlap of genetic and transcriptomic associations for many phenotypes including BMI.

Despite identifying many biologically relevant findings, this study does have limitations. In large observational studies, BMI is used as a surrogate measure for adipose tissue fat mass given ease of collection and (incomplete) correlation with fat and lean mass. Validation of results from BMI should be made through targeted studies of tissue-specific adiposity in human and model systems to better understand how altered gene expression relates to physiologically defined measures of obesity. In addition, Gene expression was assessed in peripheral blood, and not the likely targeted tissues (e.g., adipose). Tissue-specificity of gene expression could explain the limited overlap with previous GWAS findings (including both common and rare variants) as well as a smaller sample size, particularly in SIRE-specific analyses. GWAS of BMI and obesity-related traits have identified associations with variants in the leptin-melanocortin pathway and genes expressed in the brain, central nervous system, and adipose tissue [59]. Despite being conducted in blood tissue, our analyses did identify genes involved in other tissue-specific processes, such as neuronal regulation and signaling (e.g., *NRG1, PACSIN1*, and *KCNH3*), and JAK/STAT signaling which drives leptin gene expression in white adipose tissue [60]. Ideally, studies should incorporate tissue- and cell-specific molecular signatures of obesity, however access to relevant tissues remains a limitation. Furthermore, as TWAS captures non-causal associations, many of the results reported in this manuscript are likely driven by increased BMI as opposed to causal associations detected by GWAS. In addition, these analyses were performed using only cross-sectional data. Future work in this (and other) populations would benefit from longitudinal analysis exploring the relationship between change in gene expression with change in obesity (BMI) in order to identify temporally stable relationships in diverse populations.

In summary, we provide evidence of 45 genes whose expression is associated with BMI in the MESA, including novel and replicated targets relating to obesity biology. These genes and pathways may provide insights into early biomarkers of obesity for targeted therapeutics and behavioral modification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). Genome sequencing for “NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)” (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1 and HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1), and TOPMed MESA Multi-Omics (HHSN2682015000031/ HHSN26800004). The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the Multi-Ethnic Study of Atherosclerosis (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutes can be found at http://www.mesa-nhlbi.org.

DATA AVAILABILITY

MESA data are available through the application to dbGaP. Phenotypes are available in MESA study accession phs000209.v13.p3, and transcriptomic data has been deposited and will become available through the TOPMed MESA study accession phs001416.v2.p1.

REFERENCES

1. GBD 2015 Obesity Collaborators. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med. 2017;377:13–27. [PubMed: 28604169]
2. World Health Organization. Global health estimates 2020: deaths by cause, age, sex, by country and by region, 2000–2019. 2020.
3. Hoffmann TJ, Choquet H, Yin J, Banda Y, Kvale MN, Glymour M, et al. A large multiethnic genome-wide association study of adult body mass index identifies novel loci. Genetics. 2018;210:499–515. [PubMed: 30108127]
4. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518:197–206. [PubMed: 25673413]
5. Turcot V, Lu Y, Highland HM, Schurmann C, Justice AE, Fine RS, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. Nat Genet. 2018;50:26–41. [PubMed: 29273807]
6. Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L, et al. Genome-wide physical activity interactions in adiposity – a meta-analysis of 200,452 adults. PLoS Genet. 2017;13:e1006528. [PubMed: 28448500]
7. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet. 2019;104:65–75. [PubMed: 30595370]
8. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet. 2019;28:166–74. [PubMed: 30239722]
9. Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. J Allergy Clin Immunol. 2020;145:537–49. [PubMed: 31669095]
10. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, et al. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. PLoS Genet. 2015;11:e1005378. [PubMed: 26426971]
11. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet. 2018;27:3641–9. [PubMed: 30124842]
12. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics. 2010;42:937–48. [PubMed: 20935630]
13. Young KL, Graff M, Fernandez-Rhodes L, North KE. Genetics of obesity in diverse populations. Curr Diab Rep. 2018;18:145. [PubMed: 30456705]
14. Bild DE. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol. 2002;156:871–81. [PubMed: 12397006]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. [PubMed: 23104886]
16. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics. 2012;28:1530–2. [PubMed: 22539670]
17. Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland Jane E, Mudge JM, et al. Gencode 2021. Nucleic Acids Res. 2021;49:D916–23. [PubMed: 33270111]
18. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. [PubMed: 25516281]
19. Sofer T, Kurniansyah N, Aguet F, Ardlie K, Durda P, Nickerson DA, et al. Benchmarking association analyses of continuous exposures with RNA-seq in observational studies. Brief Bioinform. 2021;22:bbab194. [PubMed: 34015820]
20. van Iterson M, van Zwet EW, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18:19. [PubMed: 28129774]
21. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1. [PubMed: 20616382]
22. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954;10:101–29.
23. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539–58. [PubMed: 12111919]
24. Zhang D rsq: R-squared and related measures. R Package. 2022.
25. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47:D1005–12. [PubMed: 30445434]
26. Loh P-R, Genovese G, Handsaker RE, Finucane HK, Reshef YA, Palamara PF, et al. Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. Nature. 2018;559:350–5. [PubMed: 29995854]
27. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2:100141. [PubMed: 34557778]
28. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7. [PubMed: 22455463]
29. Kanehisa M KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30. [PubMed: 10592173]
30. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49:D545–51. [PubMed: 33125081]
31. Ye Y, Doak TG. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol. 2009;5:e1000465. [PubMed: 19680427]
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9. [PubMed: 10802651]
33. Gene Ontology Consortium The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2021;49:D325–34. [PubMed: 33290552]
34. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5. [PubMed: 23715323]
35. Paccosi S, Pala L, Cresci B, Silvano A, Cecchi M, Caporale R, et al. Insulin resistance and obesity affect monocyte-derived dendritic cell phenotype and function. Diabetes Res Clin Pract. 2020;170:108528. [PubMed: 33157116]
36. Mishra J, Verma RK, Alpini G, Meng F, Kumar N. Role of Janus Kinase 3 in predisposition to obesity-associated metabolic syndrome. J Biol Chem. 2015;290:29301–12. [PubMed: 26451047]
37. Ghadami M, Tomita HA, Najafi MT, Damavandi E, Farahvash MS, Yamada K, et al. Bardet-Biedl syndrome type 3 in an Iranian family: clinical study and confirmation of disease localization. Am J Med Genet. 2000;94:433–7. [PubMed: 11050632]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

38. Young TL, Woods MO, Parfrey PS, Green JS, O'Leary E, Hefferton D,et al. Canadian Bardet-Biedl syndrome family reduces the critical region of BBS3 (3p) and presents with a variable phenotype. Am J Med Genet. 1998;78:461–7. [PubMed: 9714014]
39. Mazzarella L, Botteri E, Matthews A, Gatti E, Di Salvatore D, Bagnardi V, et al. Obesity is a risk factor for acute promyelocytic leukemia: evidence from population and cross-sectional studies and correlation with FLT3 mutations and polyunsaturated fatty acid metabolism. Haematologica. 2020;105:1559–66. [PubMed: 31515354]
40. Díaz-Rúa R, Palou A, Oliver P. Cpt1agene expression in peripheral blood mononuclear cells as an early biomarker of diet-related metabolic alterations. Food Nutr Res. 2016;60:33554. [PubMed: 27885970]
41. Orellana-Gavaldà JM, Herrero L, Malandrino MI, Pañeda A, Sol Rodríguez-Peña M, Petry H, et al. Molecular therapy for obesity and diabetes based on a long-term increase in hepatic fatty-acid oxidation. Hepatology. 2011;53:821–32. [PubMed: 21319201]
42. Hauffe R, Stein V, Chudoba C, Flore T, Rath M, Ritter K, et al. GPx3 dysregulation impacts adipose tissue insulin receptor expression and sensitivity. JCI Insight. 2020;5:e136283. [PubMed: 32369454]
43. Langhardt J, Flehmig G, Klöting N, Lehmann S, Ebert T, Kern M, et al. Effects of weight loss on glutathione Peroxidase 3 serum concentrations and adipose tissue expression in human obesity. Obesity Facts. 2018;11:475–90. [PubMed: 30537708]
44. Paulo E, Wu D, Hecker P, Zhang Y, Wang B. Adipocyte HDAC4 activation leads to beige adipocyte expansion and reduced adiposity. J Endocrinol. 2018;239:153–65. [PubMed: 30121575]
45. Bereswill S, Abu-Farha M, Tiss A, Abubaker J, Khadir A, Al-Ghimlas F, et al. Proteomics analysis of human obesity reveals the epigenetic factor HDAC4 as a potential target for obesity. PLoS One. 2013;8:e75342. [PubMed: 24086512]
46. Netea MG, Joosten LAB, Lewis E, Jensen DR, Voshol PJ, Kullberg BJ, et al. Deficiency of interleukin-18 in mice leads to hyperphagia, obesity and insulin resistance. Nat Med. 2006;12:650–6. [PubMed: 16732281]
47. Ennequin G, Boisseau N, Caillaud K, Chavanelle V, Etienne M, Li X, et al. Neuregulin 1 affects leptin levels, food intake and weight gain in normal-weight, but not obese, db/db mice. Diabetes Metab. 2015;41:168–72. [PubMed: 25573691]
48. Val CH, de Oliveira MC, Lacerda DR, Barroso A, Batista NV, Menezes-Garcia Z, et al. SOCS2 modulates adipose tissue inflammation and expansion in mice. J Nutr Biochem. 2020;76:108304. [PubMed: 31816561]
49. Li J, Diao B, Guo S, Huang X, Yang C, Feng Z, et al. VSIG4 inhibits proinflammatory macrophage activation by reprogramming mitochondrial pyruvate metabolism. Nat Commun. 2017;8:1322. [PubMed: 29109438]
50. Gurzov EN, Stanley WJ, Pappas EG, Thomas HE, Gough DJ. The JAK/STAT pathway in obesity and diabetes. FEBS J. 2016;283:3002–15. [PubMed: 26972840]
51. Ge D, Gooljar SB, Kyriakou T, Collins LJ, Swaminathan R, Snieder H, et al. Association of common JAK2 variants with body fat, insulin sensitivity and lipid profile. Obesity (Silver Spring). 2008;16:492–6. [PubMed: 18239666]
52. Penas-Steinhardt A, Tellechea ML, Gomez-Rosso L, Brites F, Frechtel GD, Poskus E. Association of common variants in JAK2 gene with reduced risk of metabolic syndrome and related disorders. BMC Med Genet. 2011;12:166. [PubMed: 22185674]
53. Dodington DW, Desai HR, Woo M. JAK/STAT – emerging players in metabolism. Trends Endocrinol Metab. 2018;29:55–65. [PubMed: 29191719]
54. Messaoudi I, Handu M, Rais M, Sureshchandra S, Park BS, Fei SS, et al. Long-lasting effect of obesity on skeletal muscle transcriptome. BMC Genomics. 2017;18:411. [PubMed: 28545403]
55. Chen J-J, London IM. Hemin enhances the differentiation of mouse 3T3 cells to adipocytes. Cell. 1981;26:117–22. [PubMed: 6799206]
56. Hazegh K, Fang F, Bravo MD, Tran JQ, Muench MO, Jackman RP, et al. Blood donor obesity is associated with changes in red blood cell metabolism and susceptibility to hemolysis in cold storage and in response to osmotic and oxidative stress. Transfusion. 2020;61:435–48. [PubMed: 33146433]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

57. Samocha-Bonet D, Lichtenberg D, Tomer A, Deutsch V, Mardi T, Goldin Y, et al. Enhanced erythrocyte adhesiveness/aggregation in obesity corresponds to low-grade inflammation. Obes Res. 2003;11:403–7. [PubMed: 12634437]

58. Moll M, Boueiz A, Ghosh AJ, Saferali A, Lee S, Xu Z, et al. Development of a blood-based transcriptional risk score for chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2022;205:161–70. [PubMed: 34739356]

59. Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. Nat Rev Genet. 2022;23:120–33. [PubMed: 34556834]

60. Banks AS, Davis SM, Bates SH, Myers MG. Activation ofdownstream signals by the long form of the leptin receptor. J Biol Chem. 2000;275:14563–72. [PubMed: 10799542]

Author Manuscript

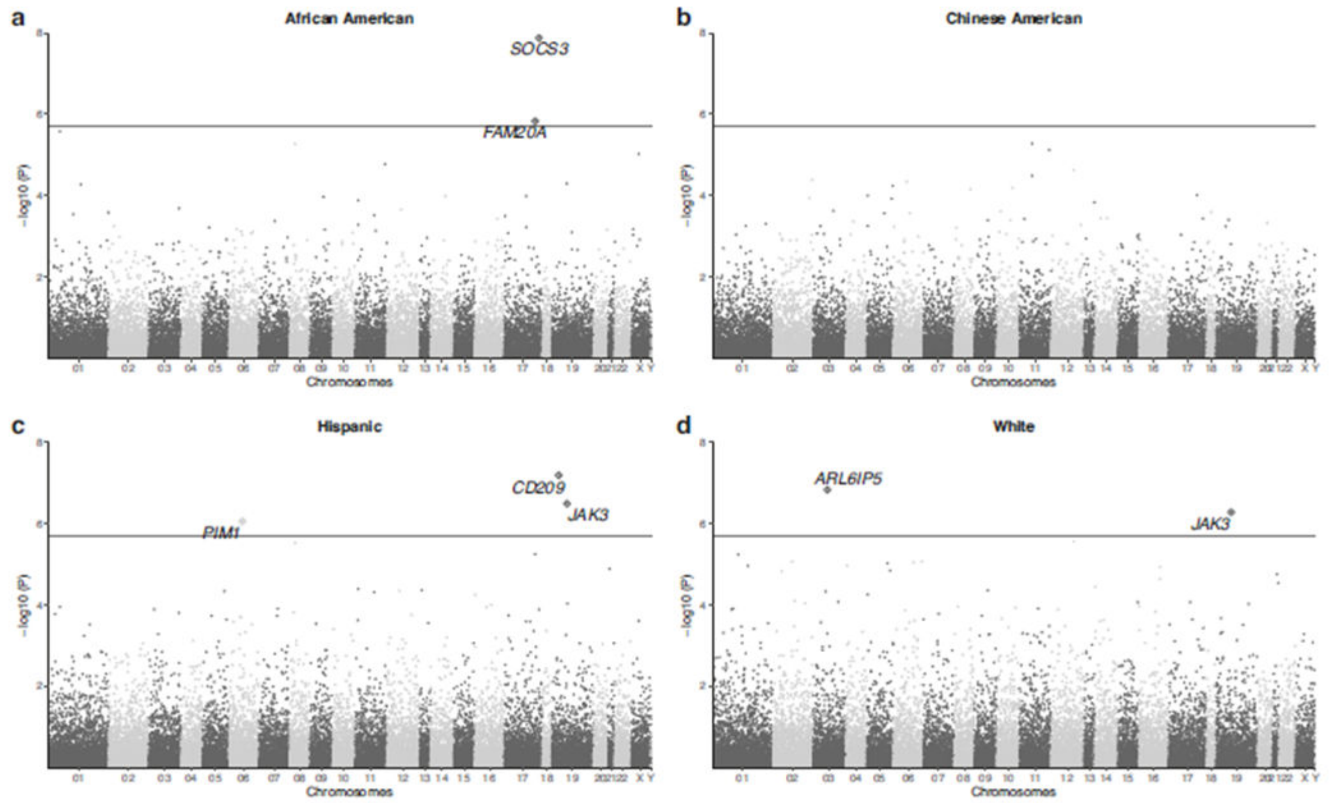

Fig. 1. Transcriptome-wide associations with BMI in each SIRE group.

Panels represent associations in **a** African American, **b** Chinese American, **c** Hispanic, and **d** white individuals. Horizontal lines represent the Bonferroni significance threshold for 25,416 tests ($p < 1.97 \times 10^{-6}$). Significant associations are labeled by gene name.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

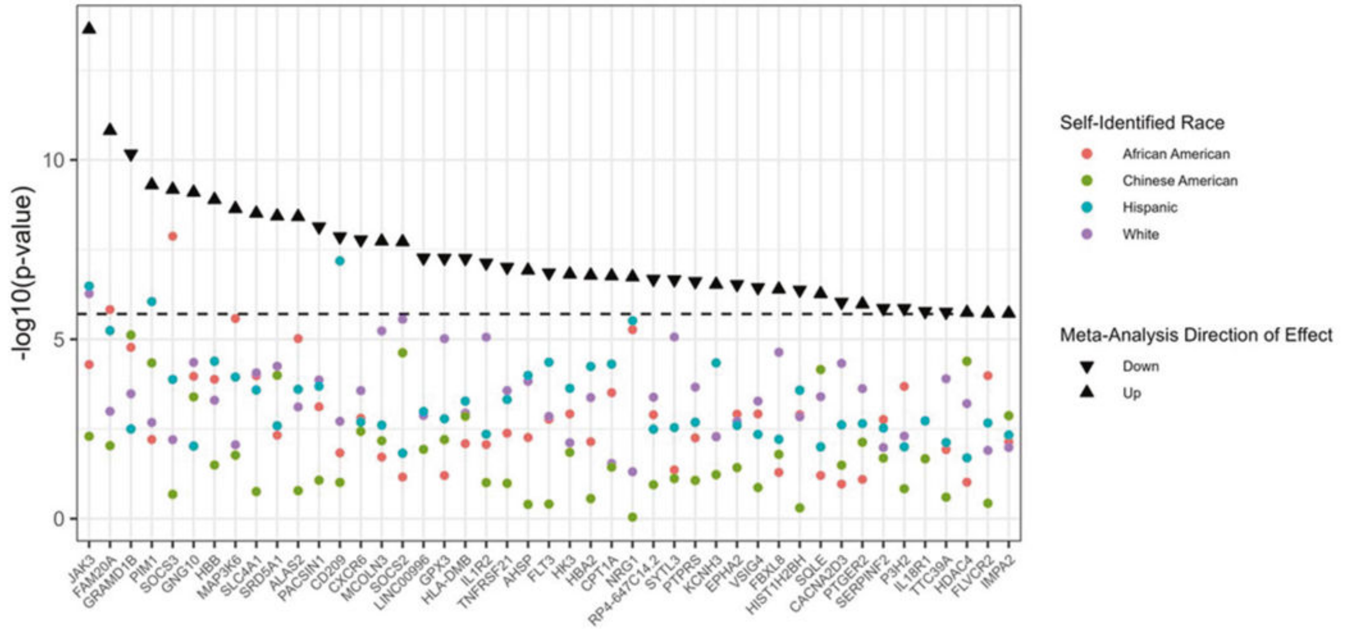

Fig. 2. Foothill plot of multi-ethnic meta-analysis of BMI associations with gene expression.

Colored circles represent p values from individual analyses and triangles represent p values from meta-analysis. Horizontal dashed line indicates Bonferroni significance threshold for 25,416 tests ($p < 1.97 \times 10^{-6}$).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

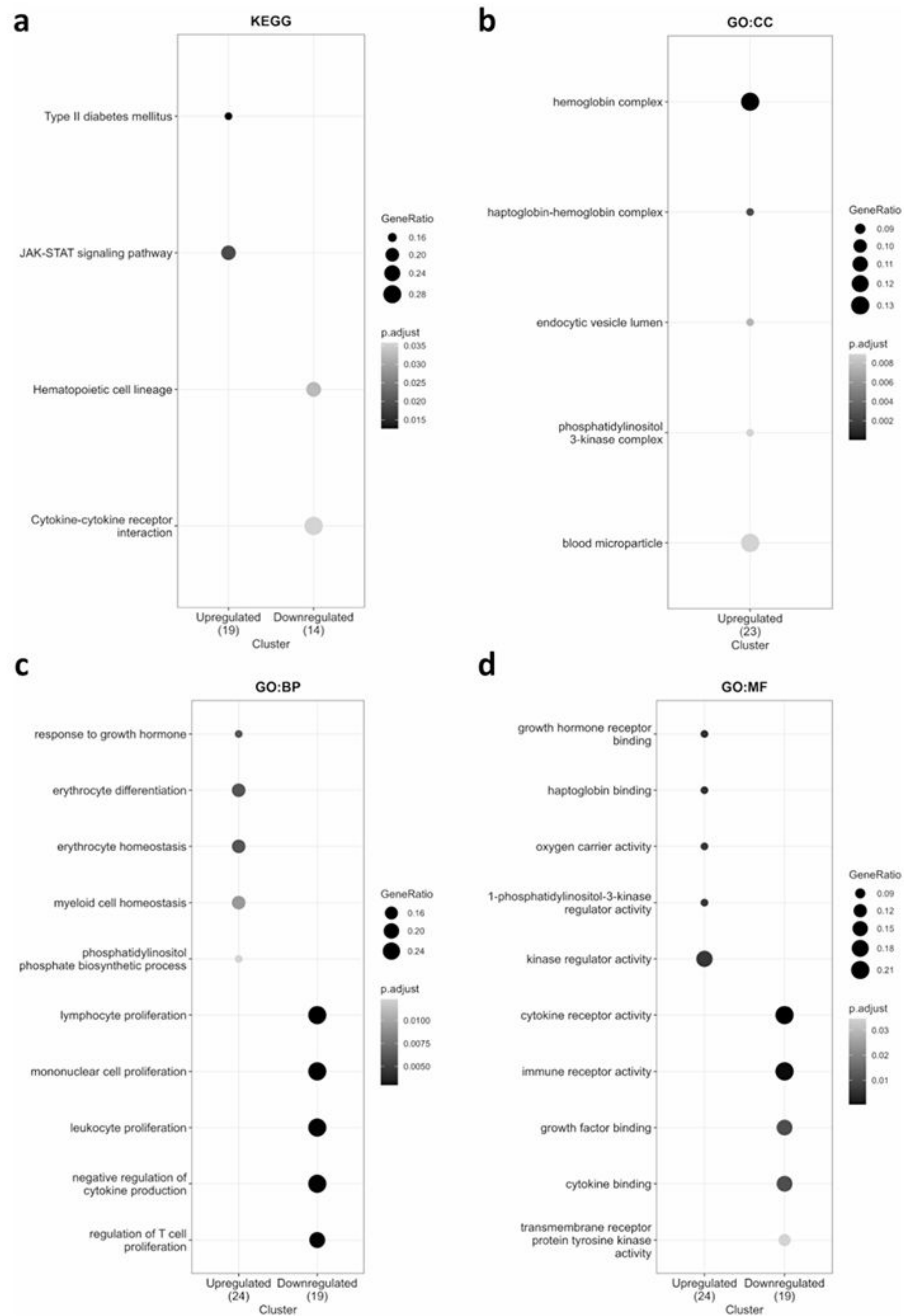

Fig. 3. Pathway overrepresentation enrichment analysis.

Panels represent most significant enriched pathways in **a** KEGG, **b** GO: Cellular Component (GO:CC), **c** GO: Biological Process (GO:BP), and **d** GO: Molecular https://doi.org/10.1038/s41366-022-01240-x Function (GO:MF) databases.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table 1.

Demographic information of MESA participants by SIRE.

	African American (*n* = 310)	Chinese American (*n* = 104)	Hispanic (*n* = 265)	White (*n* = 412)	*p* value
Age (years, SD)	60.5 ± 9.6	61.7 ± 9.9	58.1 ± 9.3	61.2 ± 9.6	2.3×10^{-4}*
Sex (M, %)	133 (42.9)	52 (50)	123 (46.4)	200 (48.5)	0.42#
BMI (kg/m^2, SD)	31.0 ± 5.7	24.6 ± 3.1	29.9 ± 5.5	27.9 ± 4.7	$<2.2 \times 10^{-16}$*

Statistical significance was assessed using *one-way ANOVA and #χ^2 test.