bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Epigenome-wide association in adipose tissue from the METSIM cohort

Luz D. Orozco[1]*,*** Colin Farrell[1], Christopher Hale[1], Liudmilla Rbi[1], Arturo Rinaldi[1], Mete Civelek[2]**, Calvin Pan[2], Larry Lam[1], Dennis Montoya[1], Chantle Edillor[2], Marcus Seldin[2], Karen L Mohlke[3], Steve Jacobsen[1,4,5], Johanna Kuusisto[6], Markku Laakso[6], Aldons J Lusis[2], Matteo Pellegrini1,[4,5]

[1]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, 90095, USA.
[2]Departments of Human Genetics, Medicine, and Microbiology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, 90095, USA.
[3]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA.
[4]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA
[5]Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, University of California Los Angeles, Los Angeles, California 90095, USA.
[6]Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, California 90095, USA.
[7]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital
*Present address: Department of Bioinformatics, Genentech, Inc., South San Francisco, CA, 94158, USA.
**Present address: Center for Public Health Genomics, Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, 22904, USA.
***Corresponding Author: Phone: 323-491-4999, Email: luz.d.orozco@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**ABSTRACT**

Most epigenome-wide association studies to date have been conducted in blood. However, metabolic syndrome is mediated by a dysregulation of adiposity and therefore it is critical to study adipose tissue in order to understand the effects of this syndrome on epigenomes. Therefore, to determine if natural variation in DNA methylation was associated with metabolic syndrome traits, we profiled global methylation levels in subcutaneous abdominal adipose tissue. We measured association with 32 clinical traits related to diabetes and obesity in 201 people from the Metabolic Syndrome In Men cohort. We performed epigenome-wide association studies between DNA methylation levels and traits, and identified significant associations for 13 clinical traits in 21 loci. We prioritized candidate genes using eQTL, and identified 18 high confidence candidate genes, including known and novel genes associated with diabetes and obesity traits. We also carried out an analysis to identify which cell types may be mediating the associations, and concluded that most of the loci we identified were specific to adipocytes. We determined whether the abundance of cell types varies with metabolic traits, and found that macrophages increased in abundance with the severity of metabolic syndrome traits. Finally, we developed a DNA methylation based biomarker to assess type II diabetes risk in adipose tissue. In conclusion, our results demonstrate that profiling DNA methylation in adipose tissue is a powerful tool for understanding the molecular effects of metabolic syndrome on adipose tissue, and can be used in conjunction with traditional genetic analyses to further characterize this disorder.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## INTRODUCTION

Metabolic syndrome traits such as obesity, dyslipidemia, insulin resistance, and hypertension underlie the common forms of atherosclerosis, type 2 diabetes (T2D) and heart failure, which together account for the majority of deaths in Western populations. Metabolic syndrome affects 44% of adults over the age of 50 in the US, and people affected with metabolic syndrome have higher risk of heart attacks, diabetes and stroke (1). Numerous studies have investigated the genetic basis of metabolic syndrome traits such as diabetes (2), and accumulating evidence suggests that epigenetics is associated with these phenotypes (3, 4).

Methylation of DNA cytosine bases is evolutionarily conserved and plays important roles in development, cell differentiation, imprinting, X-chromosome inactivation, and regulation of gene expression. Aberrant DNA methylation in mammals is associated with both rare and complex traits including cancer, aging (5), and imprinting disorders such as Prader-Willi syndrome. Recent studies have demonstrated that much like genome sequence variation, DNA methylation is variable among individuals in human (6), plant (7) and mouse (8) populations. Moreover, differences in DNA methylation of cytosines are in part heritable and controlled by genetics both in *cis* and in *trans*. However, sex and environmental factors such as smoking and diet can also influence DNA methylation differences, leading to changes in methylation levels over an individual's lifetime (9).

DNA methylation states have been shown to be associated with biological processes underlying metabolic syndrome, including obesity, hypertension and diabetes (10). Environment-induced changes in DNA methylation have also been associated with fetal origins of adult disease (11), and alterations in maternal diet during pregnancy can affect the methylation levels of the placenta, inducing transcriptional changes in key metabolic regulatory genes (12, 13). Recent studies have also shown that diet-induced obesity in adults affects methylation of obesogenic genes such as leptin (14), *Scd1* (15), and *LPK* (16).

Similar to genome-wide association studies (GWAS), *epi*genome-wide association studies (EWAS) aim to identify candidate genes for traits by using epigenetic factors instead of SNP genotypes in the association model. EWAS have recently identified associations for gene expression and protein levels in humans (6), and complex traits such as bone mineral density, obesity, and insulin resistance in mice (17). However, to date, most EWAS studies have been carried out in blood, which is the tissue that is most readily collected for large scale studies.

By contrast, in this study we examined the association of DNA methylation with metabolic traits in humans using adipose tissue samples from the Metabolic Syndrome in Men (METSIM) cohort. Metabolic syndrome is characterized by a clustering of three or more of the following conditions: elevated blood pressure, elevated serum triglycerides, elevated blood sugar, low HDL levels, and abdominal obesity. As adipose is known to be a central organ in metabolic syndrome manifestation, adipose tissue should be one of the most relevant for defining and studying metabolic syndrome traits (18). The METSIM cohort has been thoroughly characterized for longitudinal clinical data of metabolic traits including a 3-point oral glucose tolerance test, cardiovascular disorders, diabetes complications, drug and diet questionnaire, as well as high density genotyping, and genome-wide expression in adipose (19, 20) We performed EWAS on clinical traits using reduced representation bisulfite sequencing data and identified 51 significant associations for metabolic syndrome traits, corresponding to 21 loci. These associations include previously known genes, *FASN* (21–23) and *RXRA* (24–26), as well as loci harboring 22 new candidate genes for diabetes and obesity in humans. We identify the types of cells that are likely to be mediating these associations, and conclude that adipocytes are involved. We also examine the abundance of cell types and show that macrophages increase with the severity of metabolic syndrome

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

traits. Finally, we develop a biomarker for to assess type II diabetes status in adipose tissue. Our results demonstrate that DNA methylation profiling is both useful and complementary to GWAS for characterizing the molecular and cellular basis of metabolic syndrome.

## RESULTS

### METSIM COHORT

The METSIM cohort consists of 10,197 men from Kuopio Finland between 45-73 years of age. Laakso and colleagues (20) have characterized this cohort for numerous clinical traits involved in diabetes and obesity, and genome-wide expression levels in adipose tissue biopsies (19). In this study, we examined 32 clinical traits related to metabolic syndrome (Table S1), adipose tissue expression levels using microarrays, and DNA methylation profiles from adipose tissue biopsies in 201 individuals from the METSIM cohort.

### DNA METHYLATION OF ADIPOSE BIOPSIES

To examine methylation patterns in the METSIM cohort we constructed reduced representation bisulfite sequencing libraries (RRBS) from adipose tissue biopsies, corresponding to 228 individuals. The sequences obtained from RRBS libraries are enriched in genes and CpG islands, and cover 4.6 million CpGs out of the ~30 million CpGs in the human genome (~15%). We sequenced the libraries using the Illumina HiSeq platform and obtained on average 34.3 +/- 6.7 million reads per sample. We aligned the data to the human genome using BSMAP (27) and obtained on average 21.9 +/- 4.6 million uniquely aligned reads per sample (Figure S1A), corresponding to 64% average mappability (Figure S1B), and 19x average coverage (Figure S1B). We focused our analyses on CpGs, since CHG and CHH (H = A, C or T) methylation in mammals is on average only 1-2%, which makes it difficult to detect significant variation in our samples (17, 28). We and others have previously validated RRBS data relative to traditional bisulfite sequencing by cloning DNA fragments into bacterial colonies followed by Sanger sequencing and found a high degree of concordance between RRBS and traditional bisulfite sequencing results in mice (8, 12) and humans (29). RRBS shows limited overlap with the Illumina 450k arrays, a small study (n=11) found an overlap between 24,000 - 120,000 CpG sites (30).

We filtered our dataset for CpGs with at least 10x coverage, and present in at least 75% of the samples, corresponding to 2,320,297 CpGs. However, the methylation state of individual CpGs may be subject to stochastic variation or measurement error, and we observed a single or a few outlier samples with methylation levels that are very different from the rest of the population (Figure S1C). This variability is likely to lead to spurious associations between methylation and traits, and we observed that was indeed the case when we performed EWAS using individual CpGs. In contrast to individual CpGs, the methylation level of a CpG methylation region (a unit comprised of several CpGs) is a much more robust measure of DNA methylation levels (for example see Figure S1D). Methylation regions are likely a more biologically relevant genomic unit than individual CpGs, and methylation levels of proximal CpGs tend to be correlated over distances of a few hundred bases to 1kb, roughly the typical size of CpG islands (31). Therefore, we defined 149,191 methylation regions, where each region is defined as the average methylation of multiple CpGs that are near each other and highly correlated. We require that a region have a minimum of 2 CpGs whose methylation is correlated (Pearson's r >= 0.9), and the region has a maximum size of 3kb. The distribution of methylation levels for these regions is shown in Figure S1E, with average methylation levels of 58.4% +/- 4.7. The range of methylation across all individuals

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

can vary between 0 and 100%. The average range was 33% for individual CpGs, and 25% for methylation regions (Figure S1F). These regions are located throughout the genome with no major gaps in coverage, with the exception of centromeric and acrocentric regions, and in the y-chromosome where there was minimal coverage (Figure S2).

**EWAS**

We performed epigenome-wide association (EWAS) between CpG methylation regions and 32 clinical traits related to obesity and diabetes, including body weight, body mass index (BMI), body fat percentage, oral glucose tolerance test (OGTT) glucose and insulin measurements (Table S1). All clinical traits were transformed using inverse normal transformation (see Methods), as is common practice for genome-wide association studies of quantitative traits (32). We used the linear mixed model package pyLMM to determine associations between DNA methylation patterns and phenotypes. Others and we have previously demonstrated that this approach corrects for spurious associations due to population structure (17, 33) and tissue heterogeneity (34). Associations were considered significant if the *p*-value for the association was below $1x10^{-7}$, based on the Bonferroni correction for the number of CpG regions tested.

In total, we found 51 significant associations, corresponding to 21 distinct methylation loci and 15 unique phenotypes (Figure 1 and Table1) where the *p*-value was below $1x10^{-7}$. Of the 21 distinct loci, 15 methylation loci were intragenic, and 6 loci were intergenic. The distance between intergenic loci and nearby flanking genes ranged between 23kb and 440kb. Candidate genes listed for each association in Table1 correspond to the gene itself for intragenic associations, and the two nearest flanking genes by distance for intergenic associations, with the distance between the locus and each flanking gene listed for intergenic associations (Table1). Figure 1 summarizes the genomic distribution of all EWAS hits, where each dot represents an association between a phenotype and a methylation region.

Some may argue that a significance threshold of $1x10^{-7}$ is insufficiently low, since we tested for 32 traits. Of the 51 associations described above, 26 associations would remain significant using a Bonferroni correction ($p<1x10^{-8}$) which accounts for both CpG regions and the 32 traits. However, we believe the additional Bonferroni correction for 32 traits would be too stringent given that several of the traits are not independent, for example BMI and fat mass, or plasma insulin and glucose levels. Alternatively, 40 associations would remain significant at the commonly used GWAS significance threshold ($p<5x10^{-8}$).

We found no evidence of inflation in our EWAS results, where the inflation factor lambda was on average 0.99, and maximum of 1.01 (Table 1). Sample EWAS *p*-value distributions and *qq*plots are shown in Figure S3A-C.

**CANDIDATE GENES**

We initially identified a total of 24 candidate genes and non-coding RNAs by proximity to an EWAS signal (Table 1). To prioritize candidate genes, we examined adipose expression associations from 770 individuals of the METSIM cohort previously published by our laboratories (19, 35). We asked if there were significant expression quantitative trait loci that overlapped with the methylation loci identified in the EWAS. We narrowed down the candidate gene list from 24 to 18 (75%) high confidence candidate genes that had significant cis-eQTL in adipose tissue samples from the METSIM cohort (Table 1). The reported cis-eQTL were significant for the candidate gene reported.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We identified 3 loci where multiple clinical traits mapped to the same methylation region. These loci include chromosome 17 at the *FASN* gene (Figure 2A-B), in chromosome 2 near *SLC1A4*, and in chromosome 5 near *CPEB4* (Figure 1, Table 1). The *FASN* gene has a *cis*-eQTL ($p$=9.3x10$^{-10}$, Figure 2C), suggesting that genetic variation in the population affects expression levels of this gene. One of the traits associated with this locus is BMI, and we observed a positive correlation between methylation levels in the *FASN* locus and BMI (Figure 2D). Remarkably, the observed correlation of 0.4 suggests that the methylation of FASN alone is able to capture 16% of the variation in BMI observed in our cohort. Moreover, we also observed an inverse correlation between methylation and *FASN* expression in adipose tissue biopsies (Figure 2E), and an inverse correlation between *FASN* expression and BMI (Figure 2F).

A second locus is located upstream of *SLC1A4* and was associated with waist circumference, lean mass, fat mass, plasma insulin levels, BMI, and indices of insulin resistance and insulin sensitivity MATSUDA, and HOMAIR (Figure 3A-B). *SLC1A4* has a *cis*-eQTL ($p$=1.6x10$^{-10}$, Figure 3C), and we observed an inverse correlation between methylation at this locus and the insulin resistance index HOMAIR (Figure 3D), an inverse correlation between methylation and expression of *SLC1A4* (Figure 3E), and a positive correlation between *SLC1A4* expression and HOMAIR (Figure 3F). A third locus is located upstream of *CPEB4*, and was associated with basal plasma insulin levels, OGTT plasma insulin, and the indices of insulin resistance and insulin sensitivity MATSUDA, and HOMAIR (Table 1). A *cis*-eQTL for *CPEB4* expression in adipose tissue biopsies ($p$=2.4x10$^{-174}$) makes this gene strong candidate gene for this locus.

## CELL TYPE DECOMPOSITION OF ADIPOSE TISSUE

Whenever we examine molecular phenotypes such as DNA methylation and gene expression in tissues, the question arises, what cell-types within the tissue are responsible for the signal we observe? We know that subcutaneous adipose tissue is composed primarily of adipocytes, but also contains endothelial cells and immune cells such as resident and infiltrating macrophages. Moreover, we know that obese individuals show increased macrophage content in their adipose tissue, and hence that heterogeneity in people's phenotypes can influence cell-type composition in the adipose tissue biopsies (36). To examine macrophage content in the adipose tissue biopsies, we examined expression levels of genes expressed in adipocytes, namely *PPARG, CFD, ADIPOQ, FABP4, CIDEC* and *LEP, and TNMD*, and genes highly expressed macrophages including *TLR1, TLR2, TLR3, TLR4, ABCG, IL10*, and *TNF*. We found high expression levels of adipocyte-specific genes and low expression of macrophage-specific genes (Figure S3D). These results suggest that there is minimal macrophage content in the adipose biopsies. However, the genes selected may not fully reflect the transcriptome of adipocytes and macrophages, or additional cell-types that may be present in adipose tissue.

To further explore the contribution of different cell-types to the METSIM adipose tissue biopsies, we performed cell-type deconvolution using BS-seq methylation data from our samples and from multiple reference cell-types including adipocytes, endothelial cells, macrophages, neutrophils, NK-, T-, and B-cells. Using this approach, we can determine the relative content of different cell types by comparing DNA methylation at cell-specific methylation markers in our test samples, to DNA methylation signatures derived from purified cell types (see Methods). Consistent with our previous analysis of gene expression in adipose- and macrophage-specific genes, we found that the highest cell-type represented in our adipose biopsies was indeed adipocyte (Figure 4A), but we also found evidence of macrophage and neutrophil content.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Since highly expressed genes are often correlated with lower methylation levels in their promoters, we hypothesize that if our genes with significant associations with metabolic syndrome are expressed in adipose tissue, they will also have lower methylation levels in the cell types in which they are expressed. When we examine DNA methylation levels in CpG regions associated with traits in our EWAS, we find that adipocytes tend to have lower methylation levels at these loci, relative to other cell types, suggesting that many of them may be specific to adipocytes. However, two associated loci at *RXRA* and *RAP1GAP2* genes also show decreased methylation levels in macrophages and neutrophils, suggesting that DNA methylation at these loci may be derived from macrophages and/or neutrophils.

Finally, although the relative content of cell-types such as macrophages may be small, they can still contribute to expression and methylation levels, and to clinical phenotypes. We studied the correlation between the cell-type content and clinical traits across all individuals, and found that macrophage content was positively correlated with the clinical traits associated in our EWAS (Figure 4C). These results support the notion that both adipocytes and macrophages contribute to DNA methylation signatures, and to associations between DNA methylation and clinical traits. The correlation between neutrophil content and traits is minimal, suggesting that methylation levels derived from neutrophils are potentially derived from blood contamination during collection of biopsies.

## CHORMATIN STATES AT CANDIDATE LOCI

We used the Roadmap (37) and RegulomeDB databases to examine chromatin marks and chromatin states in adipocytes or adipose tissue in each of the EWAS loci. The chromatin marks found in each locus are summarized in Figure S4. The sites group into three clusters. The first represents regions of active transcription that contain H3K36me3. The third cluster likely contains enhancers, which are marked by H4K4me1 and H3K27ac. The second cluster is more heterogeneous and has generally fewer marks, with a few sites showing no marks at all.

## DNA METHYLATION BIOMARKER FOR TYPE II DIABETES

DNA methylation is a useful biomarker for assessing the age of an individual (38, 39), as well as BMI (4). We asked whether we could develop a biomarker for adipose tissue that could be used to assess a metabolic health outcome, type II diabetes. To this end, we first developed an aggregate measure of type II diabetes by combining multiple clinical traits measured in the METSIM cohort using principal component analysis. Briefly, traits were selected for inclusion into the aggregate measure of metabolic health using a greedy algorithm that considered combinations of features that produced the largest Welch's test statistic in the first principal component when comparing healthy individuals to individuals who had received a type II diabetes diagnosis at baseline examination in the METSIM cohort (n=10,172). Our final measure consists of a linear combination of six traits: two measurements of glucose at baseline and at two hours during an oral glucose tolerance test, a binary measure of elevated blood glucose, two measurements of urine albumin levels at the start and end of collection, and one measurement of LDL levels (Table 2).

This first principal component allows us to effectively segregate individuals by type II diabetes status as baseline (Figure 5A). A follow-up examination was conducted an average of 53.2 months (std=12.6) after the baseline examination. This allows us to identify a subpopulation that was healthy at baseline but develops type II diabetes at follow-up. We find that based on the PC1 score this group has baseline levels that are intermediate between the healthy and type 2 diabetes group (Figure 5A). This

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

suggest that our approach is also able to detect individuals at risk of developing type II diabetes. Additionally, the first principal component outperforms individual metabolic metrics commonly used for diagnosis of type II diabetes (40) in the classification of type II diabetes status at baseline or follow-up examination (Figure 5B). These results suggest that PC1 is a useful metric for assessing risk of developing type II diabetes.

Finally, we asked whether we could predict the value of PC1 using DNA methylation, in order to develop a biomarker to assess type II diabetes risk. To this end we used randomized lasso to select CpG sites used to generate a linear model that predicts the PC1 value of each individual. We used leave one out cross validation, fitting a separate model to test each sample independently, to measure the accuracy of this approach, and found that the R-squared between our predicted and measured PC1 values was 0.608 (Figure 5C). Furthermore, we were able to effectively classify (AUC=0.80) the type II diabetes status of an individual using the predicted PC1 value (Figure 5D). This suggests that using 1456 CpG sites measured in adipose tissue we are able to predict the risk of developing diabetes.

**DISCUSSION**

In this study, we utilized natural variation in DNA methylation in the adipose tissue of a human population to explore the relationship between DNA methylation and complex clinical traits associated with metabolic syndrome. We chose to focus our analysis on adipose tissue as it is believed to be the central tissue mediating metabolic syndrome traits. While metabolic syndrome surely involves a complex interplay between adipose tissue, liver and immune cells, it is likely that adipose tissue undergoes the most dramatic epigenetic changes during the advancement of metabolic syndrome. In fact, previous studies have shown that adipose tissue has significant epigenetic differences between lean and obese individuals (41).

Using epigenome-wide analysis we identified 21 novel associations for diabetes and obesity phenotypes, corresponding to 24 candidate genes. We further narrowed our candidates to 18 high confidence candidate genes based on presence of *cis*-eQTL for these genes in adipose tissue (Table 1). Our results demonstrate the power of EWAS to identify significant associations for metabolic traits in humans using only 201 individuals, and highlight how epigenetic factors such as DNA methylation should be considered in conjunction with genetic variation to elucidate the complex cellular mechanisms that ultimately lead to observable phenotypes.

We found three loci where multiple clinical traits mapped to the same methylation region and associated gene, *SLC1A4* on chromosome 2 (Figure 3), *CPEB4* on chromosome 5 (Table 1), and *FASN* on chromosome 17 (Figure 2). *FASN* is a known regulator of fatty acid metabolism (42), and its expression if associated with mature adipocytes that replete with the downstream products of FASN. We observed that DNA methylation in the *FASN* gene was correlated with multiple metabolic syndrome clinical traits. Remarkably, we show that the variation in DNA methylation levels of FASN capture approximately 16% of the variation of metabolic traits such as BMI (Figure 2D), a significant portion of the variation in the trait in our population.

The mechanisms by which metabolic syndrome traits affect methylation levels are still incompletely understood. While transcriptional levels of *FASN* and other genes likely respond quickly to insulin release, it is well established that, in contrast, DNA methylation levels are very stable, and change on a much slower timescale. We expect that methylation levels in the body of a gene change on the timescale of weeks, in response to the daily changes of insulin signaling. Thus, we hypothesize that DNA methylation levels at or near genes may reflect the history of insulin signaling during the previous weeks,

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and are thus robust markers for the average physiological state of the individuals. We observed that the expression of FASN decreases with increasing obesity, a result that confirms previous observations, (43) in agreement with previous multiple studies showing that obese individuals have lower insulin sensitivity, leading to lower FASN expression. The FASN regions we identified in our study are intragenic, and associated with several histone marks including H3k36me3, and H3k4me3 and H3k4me1. These marks are often associated with the boundaries of promoters and transcribed regions. Thus, we speculate that the reduced insulin sensitivity leads to a hypermethylation of the promoter, which is associated with a decrease in gene expression.

We also observed that the methylation levels near RXRA are associated with metabolic traits. RXRA is known to form a complex with PPARG, a master regulator of adipogenesis and adipocytes (44). In contrast to FASN and RXRA, the amino acid transporter *SLC1A4,* previously linked to metabolite levels and atherosclerosis (45), is a novel gene associated with both diabetes and obesity traits. The cytoplasmic polyadenylation element binding protein 4, *CPEB4* has been previously associated with obesity (46), and waist-to-hip ratio (47), but not with diabetes. Here we find that *CPEB4* is associated with measures of insulin sensitivity/ insulin resistance.

We found several other novel gene associations such as *LINC01317* which we found to be associated with insulin sensitivity (Matsuda index) and *TPCN1* which was associated with body weight in our EWAS. In addition, we found (strawberry notch homolog 2) *SBNO2* to be associated with BMI and body weight in our study, and with BMI in a previous EWAS (4). *SBNO2* regulates inflammatory responses (48), and a *Sbno2* mutant mouse model shows impaired osteoclast fusion, osteoblastogenesis, osteopetrosis, and increased bone mass (49).

As adipose tissue is a heterogeneous tissue that contains adipocytes, endothelial and immune cells, among others, we asked whether we could determine which cell types were most significant in our analyses. We found that most of our significantly associated loci were hypomethylated in adipocytes compared to other cell types. This suggested that most of the methylation variation we observe is likely occurring in adipocytes, which constitute the majority of cells in adipose tissue.

Using a DNA methylation based deconvolution approach we also estimated the abundance of each cell type in each individual. As expected, we found that adipocytes constituted around 80% of the cells in our samples. Intriguingly, however, we observed that the abundance of macrophages varied across individuals in a manner that was strongly correlated with metabolic traits. This suggests that obese individuals have higher macrophage counts in their adipose tissue compared to lean individuals, a result that supports pervious observations (36).

In previous studies DNA methylation has been used to develop robust biomarkers for multiple traits such as age (38) and BMI (4). We therefore asked whether we could develop an accurate biomarker to assess type II diabetes risk from our data. We first aggregated clinical traits to define a metric of metabolic health that is associated with the risk of developing type II diabetes, by combining measures of glucose, LDL and urine albumin. We showed that this metric capably stratifies the population into healthy and diabetic individuals. We also showed that high values of this metric strongly associated with the development of type 2 diabetes in follow-up measurements in the METSIM cohort. Finally, using a limited set of CpG sites, we developed a model that accurately predicts the values of this metric. This result suggesting that DNA methylation measurements in adipose tissue can be used to assess the risk of developing type II diabetes.

The current study outlines the usefulness of the studying a disease in the tissue type of interest, but is constrained by limited genetic variability of the METSIM cohort. The METSIM cohort is

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

composed of middle-aged Finnish men, and it is likely that some of our results will not extend to other ethnic populations or female cohorts. Global methylation patterns are known to differ between males and females in blood (50), and these sex specific methylation differences and their relationship with metabolic traits should be explored in future work. Furthermore, an open question is how the epigenetic profile varies between multiple tissues under the same physiological conditions. Comparing DNA methylation data from tissues such as liver, muscle, and visceral adipose may elucidate how these respond differently to metabolic syndrome. Finally, the composition of our study population was enriched for healthy individuals, and future work should focus on the epigenetic differences between diabetic/insulin insensitive individuals and healthy individuals."

In conclusion, our DNA methylation profiles of adipose tissue allowed us to identify loci that are likely reacting to the metabolic state of an individual, but whose modulation is also likely to affect the individual's metabolic profile. Our results demonstrate the usefulness of utilizing population variation in DNA methylation for identifying genes associated with complex clinical traits. Here, we identified 18 novel candidate genes for metabolic syndrome using the adipose tissue of 201 individuals. None of these loci could be found using GWAS in 152 individuals of the same cohort. Since DNA methylation in a fraction of CpGs is heritable and regulated by genetics in *cis* and in *trans* (3, 17, 51), EWAS and GWAS can be used in a complementary manner to uncover heritable factors contributing to the etiology of complex traits.

## MATERIALS AND METHODS

**Data access**. RRBS sequencing data and all EWAS association results can be obtained from GEO: GSE87893.

**Clinical phenotypes on human subjects**. Ethics Committee of the Northern Savo Hospital District approved the study. All participants gave written informed consent. Clinical trait phenotypes for the EWAS study were collected on 201 individuals from the METSIM cohort (19, 20, 35). The population-based METSIM study included 10,197 men, aged 45 to 73 years, from Kuopio, Finland. After 12 hours of fasting, a 2 hour oral 75 g glucose tolerance test was performed and the blood samples were drawn at 0, 30, and 120 min. Plasma glucose was measured by enzymatic hexokinase photometric assay (Konelab Systems reagents; Thermo Fischer Scientific, Vantaa, Finland), and insulin and pro-insulin were determined by immunoassay (ADVIA Centaur Insulin IRI no. 02230141; Siemens Medical Solutions Diagnostics, Tarrytown, NY, USA). Plasma levels of lipids were determined using enzymatic colorimetric methods (Konelab System reagents, Thermo Fisher Scientific, Vantaa, Finland). Plasma adiponectin was measured with Human Adiponectin Elisa Kit (Linco Research, St Charles, USA), C-reactive protein (CRP) with high sensitive assay (Roche Diagnostics GmbH, Mannheim, Germany), and interleukin 1 receptor agonist (IL1RA) with immunoassay (ELISA, Quantikine DRA00 Human IL-1RA, R&D Systems Inc., Minneapolis, USA). Serum creatinine was measured by the Jaffe kinetic method (Konelab System reagents, Thermo Fisher Scientific, Vantaa, Finland) and was used to calculate the glomerular filtration rate (GFR). Height and weight were measured to the nearest 0.5 cm and 0.1 kg, respectively. Waist circumference (at the midpoint between the lateral iliac crest and lowest rib) and hip circumference (at the level of the trochanter major) were measured to the nearest 0.5 cm. Body composition was determined by bioelectrical impedance (RJL Systems) in participants in the supine position. Summary statistics for each phenotype are shown in Table S1. We then transformed the

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

residuals using rank-based inverse-normal transformation for downstream analyses. This transformation involves ranking a given phenotype's values, transforming these ranks into quantiles and, converting the resulting quantiles into normal deviates. The goal of this transformation is to minimize spurious associations due deviations from the underlying assumption that data are normally distributed, it is common practice for GWAS of quantitative traits (32).

**Evaluation of insulin sensitivity.** We evaluated insulin sensitivity by the Matsuda index and insulin resistance by the HOMA-IR as previously described (20).

**RRBS libraries**. We prepared genomic DNA from adipose tissue biopsies with the DNeasy extraction kit (Qiagen, Valencia, CA, USA). We prepared RRBS libraries as previously described (17, 52), with minor modifications. Briefly, we isolated genomic DNA from flash frozen adipose biopsies using a phenol-chloroform extraction, digested 500ng of DNA with MspI restriction enzyme (NEB, Ipswich, MA, USA), carried out end-repair/adenylation (NEB) and ligation with TruSeq barcoded adapters (Illumina, San Diego, CA, USA). We selected DNA fragments of size range 200-300bp with AMPure magnetic beads (Beckman Coulter, Brea, CA, USA), followed by bisulfite treatment on the DNA (Millipore, Billerica, MA, USA), and PCR amplification (Bioline, Taunton, MA, USA). We sequenced the libraries by multiplexing 4 libraries per lane on the Illumina HiSeq2500 sequencer, with 100bp reads.

**Sequence alignment**. We aligned the reads with BSMAP to the hg19 human reference genome(27). We trimmed adapters with the Trim Galore! Tool (www.bioinformatics.babraham.ac.uk/projects/trim_galore/), allowed for up to 4 mismatches and selected uniquely aligned reads. We have previously shown that BSMAP performance is comparable to the BS-Seq aligners BS-Seeker2, and Bsmark in terms of accuracy and mappability (53).

**Methylation regions**. We filtered aligned data to keep only CpG cytosines with 10x coverage or more across all samples, and with data coverage in at least 75% of the samples. This resulted in 2,320,297 CpGs. From these cytosines, we defined methylation regions grouping nearby cytosines together in expanding windows using the following rules: 1) We treated each cytosine as a seed cytosine for a potential methylation region and in order to expand the region methylation levels were required to be correlated across the cohort at an R2>=0.9 for directly adjacent cytosines, 2) Extension of the methylation region from the seed cytosine was allowed to continue up to 500bp in either direction from the seed cytosine, 3) After region extension, only regions with greater than 2 cytosines were retained, 4) Overlapping or adjacent regions remaining from (3) were then merged, 5) A methylation region was limited to 3kb maximum. We chose these parameters since the majority of CpG islands are less than 3kb in size (54). Methylation for resulting regions was calculated as the average methylation of all included cytosines. The distribution of region size we observed was a minimum of 3bp, maximum of 3kb, and median 143bp.

**EWAS**. We used the linear mixed model package pyLMM (https://github.com/nickFurlotte/pylmm) to test for association and to account for potential population structure and relatedness among individuals. This method was previously described as EMMA (33), and we implemented the model in python to allow for continuous predictors, such as CpG methylation levels that vary between 0 and 1, as previously described (17). We applied the model: $y=\mu+x\beta+u+e$, where $\mu$=mean, x=CpG, $\beta$=CpG effect, and

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

u=random effects due to relatedness, with Var(u) = $\sigma_g^2$K and Var(e) = $\sigma_e^2$, where K=IBS (identity-by-state) matrix across all CpG methylation regions. We computed a restricted maximum likelihood estimate for $\sigma_g^2$K and $\sigma_e^2$, and we performed association based on the estimated variance component with an F-test to test that β does not equal 0. Associations were considered significant if the *p*-value for the association was below $1x10^{-7}$, based on the Bonferroni correction for the number of CpG regions tested.

**Inflation**. We calculated the inflation factor lambda by taking the chi-squared inverse cumulative distribution function for the median of the association *p*-values, with one degree of freedom, and divided this by the chi-squared probability distribution function of 0.5 (the median expected *p*-value by chance) with one degree of freedom. We plotted *qq*plots for representative phenotypes using the *qqplot* function in Matlab, with a theoretical uniform distribution with parameters 0,1.

**Adipose expression from human subjects**. Expression levels from adipose tissue biopsies were collected on 770 individuals of the METSIM cohort as previously described (19, 35), and 151 of these subjects were also represented in the current methylation dataset. Total RNA from METSIM participants was isolated from adipose tissue using the Qiagen miRNeasy kit, according to the manufacturer's instructions. RNA integrity number (RIN) values were assessed with the Agilent Bioanalyzer 2100 instrument and 770 samples with RIN >7.0 were used for transcriptional profiling. Expression profiling using Affymetrix U219 microarray was performed at the Department of Applied Genomics at Bristol-Myers Squibb according to manufacturer's protocols. The probe sequences were re-annotated to remove probes that mapped to multiple locations, contained variants with MAF > 0.01 in the 1,000 Genomes Project European samples, or did not map to known transcripts based on the RefSeq (version 59) and Ensembl (version 72) databases; 6,199 probesets were removed in this filtering step. For subsequent analyses, we used 43,145 probesets that represent 18,155 unique genes. The microarray image data were processed using the Affymetrix GCOS algorithm using the robust multiarray (RMA) method to determine the specific hybridizing signal for each gene.

**PEER factor analysis**. We corrected RMA-normalized expression levels for each gene using probabilistic estimation of expression residuals (PEER) factors (55). PEER factor correction is designed to detect the maximum number of *cis*-eQTL. We then transformed the residuals using rank-based inverse-normal transformation. We used the inverse normal-transformed PEER-processed residuals after accounting for 35 factors for downstream eQTL mapping.

***cis*-eQTL in adipose expression**. eQTL studies from the adipose biopsies of the METSIM cohort have been previously described (19). Briefly, gene expression in 770 adipose biopsy samples from the METSIM cohort was measured with Affymetrix U219 microarray. SNP genotyping was performed with Illumina OmniExpress genotyping chip and imputed based on the Haplotype Reference Consortium reference panel. Association of gene expression and SNPs were calculated with FaST-LMM. eQTL were defined as *cis* if the peak association had a p-value of $p < 2.46 \times 10^{-4}$ corresponding to 1% FDR, and if it was found within 1 Mb on either side of the exon boundaries of the gene, as previously described (32).

**GWAS for methylation loci**. We performed GWAS on the same 32 clinical traits, transformed using inverse normal transformation as described above, and 681,803 genotyped SNPs for 152 METSIM individuals where we had both genotypes and methylation data. We used a linear model and the R

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

package MatrixEQTL to perform the association, and selected associations where the *p*-value was below $1x10^{-7}$.

**Published histone marks**. We used the Roadmap ChIP-seq datasets to look for any histone marks in human adipocyte and adipose tissue samples. We used the RegulomeDB database to look for evidence of transcription factor footprinting, positional weight matrices (PWM), and active transcription. We accessed the public datasets at (http://www.roadmapepigenomics.org/data/) and (http://www.regulomedb.org/).

**Published GWAS hits**. Literature evidence for GWAS hits in candidate genes was obtained from the NHGRI GWA Catalog, using the PheGenI tool in NCBI.

**DNA methylation deconvolution**. To estimate the methylation contribution of different leukocytes to the adipose tissue, we used cell-specific methylation markers from DNA methylation signatures across different cell types. Cell-specific CpG methylation loci were identified from purified leukocyte (macrophages, neutrophils, B cells, CD4+ T cells, CD8+ T cells, NK cells) methylation profiles from the Blueprint epigenome project (56). Since there was only one purified adipocyte primary cell line reference available, we also included the average methylation profile across all adipose samples using in this study, which ostensibly consists primarily of adipocytes. The purified adipocyte cell reference was used as a filter to select cell type-specific CpG loci that are hypomethylated in both the adipocyte cell reference, and in the mean of the methylation levels for the 201 adipose samples. We filtered all cell methylomes to CpG loci that are common between the reference methylation profiles and the METSIM adipose tissues samples. To determine cell-specific methylation across all references, we first used a sliding window to aggregate the methylation profiles into regions of CpG loci with similar methylation (within 40% methylation difference across neighboring CpG within 500bp). Regions were selected that were uniquely hypomethylated for each cell types to provide 279 cell-specific hypomethylated regions. To estimate the proportion of each cell type within samples, we performed a non-negative least squares regression (57) on methylation at the cell-specific regions.

**Aggregate measure of metabolic health. The** METSIM cohort metabolic phenotype data included 10,197 individuals, and 484 traits. Individuals with a type 1 diabetes (n=25) diagnosis were dropped from further analysis, leaving 10,172 individuals. Numeric data was processed for downstream analysis by dropping traits with greater than 10% of data points missing. Missing values were then imputed using a k nearest neighbors (kNN) approach. KNN imputation of phenotype data occurred as follows; (1) Neighbors were ranked on Euclidean distance, (2) missing values were assigned the average value of the nearest neighbors (k=5). Following imputation, phenotype data were scored for normality (scipy.stats.normaltest)(58, 59) The threshold for a normally distributed trait was designated by randomly simulating normally distributed data of equal length as the METSIM phenotype data 1000 times, scoring the random distribution, and setting the threshold at the $90^{th}$ percentile score of the simulated distributions. Traits following a normal distribution were normalized (mean=0, STD=1) while traits from a non-normal distribution were rank based inverse normalized (mean=0, STD=1).

Traits directly predictive of type 2 diabetes status, such as family history or metformin consumption, were manually removed before feature selection. Features were initially screened for incorporation into the meta-trait using randomized logistic regression model (60). Selected features where then evaluated on their ability to distinguish between individuals with and without type 2 diabetes at

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

baseline using Welch's t-test. Starting with the feature that had the highest Welch's test statistic, combinations of traits were considered by iterating through all traits that passed the initial screen, incorporating the trait with the starting trait, performing PCA on the combined traits, scoring the 1st principal component using Welch's t-test, and returning the set of traits that produced the largest Welch's test statistic. The process was repeated until Welch's test statistic no longer increased. Six features were selected for incorporation into the meta-trait (Table 2). The combined trait matrix was decomposed using PCA.

Trait analysis pipelines were implemented in Python3.6.1, utilizing scikit-learn-0.18.1(61), numpy-1.13.1 (62), scipy-0.19.1 (63), pandas-0.20.3(64), seaborn-0.8.1(65), and matplotlib-2.0.1 (66) packages.

**Metabolic syndrome biomarker model.** The methylation matrix for model fitting was preprocessed by dropping all CpG sites with greater than 10% of data points missing. Imputing missing values using a kNN sliding window approach. Individual CpG sites with missing data were assigned the average value of the 5 nearest neighbors by Euclidean distance within a 6mb window. The resulting matrix contained 1,633,360 CpG sites. CpG sites selected using randomized lasso regression implemented in scikit-learn-0.18.1, 1456 CpG sites were selected. A leave on out cross validation strategy was utilized for model fitting, where independent models are fit for every test samples (n=228), to ensure robust model training. CpG sites were annotated with GREAT-3.0.0 (67) to generate a list of index genes. See supplemental file 1 for a list of CpG sites, average regression coefficients across the cross validated models, and index genes.

**Code Repository.** Custom code used in data processing and analysis can be found at https://github.com/NuttyLogic/METSIM_HMG_Code.

**ACKNOWLEDGEMENTS**

L.D.O. was supported by the Ruth L. Kirschstein National Research Service Award T32AR059033.
M.P and A.J.L were supported by National Institutes of Health (NIH) grant HL28481.
A.J.L. was supported by NIH grants HL30568 and 1P50 GM115318.
M.L. was supported by grants from Academy of Finland and Juselius Foundation.
K.L.M. was supported by NIH grant R01DK093757.
S.E.J. is an Investigator at the Howard Hughes Medical Institute.

**Conflicts of Interest**

The authors have no conflict of interest related to this manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## REFERENCES


1. Lusis, A.J., Attie, A.D., and Reue, K. (2008) Metabolic syndrome: from epidemiology to systems biology. *Nat. Rev. Genet.* , **9**, 819–830.
2. Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., *et al.* (2016) The genetic architecture of type 2 diabetes. *Nature* , **536**, 41–47.
3. Grundberg, E., Meduri, E., Sandling, J.K., Hedman, A.K., Keildson, S., Buil, A., Busche, S., Yuan, W., Nisbet, J., Sekowska, M., *et al.* (2013) Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* , **93**, 876–890.
4. Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., Tsai, P.C., Ried, J.S., Zhang, W., Yang, Y., *et al.* (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* , **541**, 81–86.
5. Horvath, S. (2013) DNA methylation age of human tissues and cell types. *Genome Biology* , **14**, R115.
6. Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology* , **12**, R10.
7. Chodavarapu, R.K., Feng, S., Ding, B., Simon, S.A., Lopez, D., Jia, Y., Wang, G.L., Meyers, B.C., Jacobsen, S.E., and Pellegrini, M. (2012) Transcriptome and methylome interactions in rice hybrids. *Proc. Natl. Acad. Sci. U.S.A.* , **109**, 12040–12045.
8. Orozco, L.D., Rubbi, L., Martin, L.J., Fang, F., Hormozdiari, F., Che, N., Smith, A.D., Lusis, A.J., and Pellegrini, M. (2014) Intergenerational genomic DNA methylation patterns in mouse hybrid strains. *Genome Biol.* , **15**, R68.
9. Quarta, C., Schneider, R., and Tschop, M.H. (2016) Epigenetic ON/OFF Switches for Obesity. *Cell* , **164**, 341–342.
10. Gluckman, P.D., Hanson, M.A., Buklijas, T., Low, F.M., and Beedle, A.S. (2009) Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat. Rev. Endocrinol.* , **5**, 401–408.
11. Wang, J., Wu, Z., Li, D., Li, N., Dindot, S. V, Satterfield, M.C., Bazer, F.W., and Wu, G. (2012) Nutrition, epigenetics, and metabolic syndrome. *Antioxid Redox Signal* , **17**, 282–301.
12. Chen, P.Y., Ganguly, A., Rubbi, L., Orozco, L.D., Morselli, M., Ashraf, D., Jaroszewicz, A., Feng, S., Jacobsen, S.E., Nakano, A., *et al.* (2013) Intrauterine calorie restriction affects placental DNA methylation and gene expression. *Physiol Genomics* , **45**, 565–576.
13. Aubert, D.F., Xu, H., Yang, J., Shi, X., Gao, W., Li, L., Bisaro, F., Chen, S., Valvano, M.A., and Shao, F. (2016) A Burkholderia Type VI Effector Deamidates Rho GTPases to Activate the Pyrin Inflammasome and Trigger Inflammation. *Cell Host Microbe* , **19**, 664–674.
14. Milagro, F.I., Campion, J., Garcia-Diaz, D.F., Goyenechea, E., Paternain, L., and Martinez, J.A. (2009) High fat diet-induced obesity modifies the methylation pattern of leptin promoter in rats. *J .Physiol. Biochem.* , **65**, 1–9.
15. Schwenk, R.W., Jonas, W., Ernst, S.B., Kammel, A., Jahnert, M., and Schurmann, A. (2013) Diet-dependent alterations of hepatic Scd1 expression are accompanied by differences in promoter methylation. *Horm. Metab. Res.* , **45**, 786–794.
16. Jiang, M., Zhang, Y., Liu, M., Lan, M.S., Fei, J., Fan, W., Gao, X., and Lu, D. (2011) Hypermethylation of hepatic glucokinase and L-type pyruvate kinase promoters in high-fat diet-induced obese rats. *Endocrinology* , **152**, 1284–1289.
17. Orozco, L.D., Morselli, M., Rubbi, L., Guo, W., Go, J., Shi, H., Lopez, D., Furlotte, N.A., Bennett, B.J., Farber, C.R., *et al.* (2015) Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metab.* , **21**, 905–917.
18. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., *et al.* (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature* , **518**, 187–196.
19. Civelek, M., Wu, Y., Pan, C., Raulerson, C.K., Ko, A., He, A., Tilford, C., Saleem, N.K., Stancakova, A., Scott, L.J., *et al.* (2017) Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am. J. Hum.*

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Genet.* , **100**, 428–443.

20. Laakso, M., Kuusisto, J., Stancakova, A., Kuulasmaa, T., Pajukanta, P., Lusis, A.J., Collins, F.S., Mohlke, K.L., and Boehnke, M. (2017) The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J. Lipid. Res.* , **58**, 481–493.
21. Lomba, A., Milagro, F.I., Garcia-Diaz, D.F., Campion, J., Marzo, F., and Martinez, J.A. (2009) A high-sucrose isocaloric pair-fed model induces obesity and impairs NDUFB6 gene function in rat adipose tissue. *J. Nutrigenet Nutrigenomics* , **2**, 267–272.
22. Schleinitz, D., Klöting, N., Körner, A., Berndt, J., Reichenbächer, M., Tönjes, A., Ruschke, K., Böttcher, Y., Dietrich, K., Enigk, B., *et al.* (2010) Effect of Genetic Variation in the Human Fatty Acid Synthase Gene (FASN) on Obesity and Fat Depot-Specific mRNA Expression. *Obesity* , **18**, 1218–1225.
23. Menendez, J.A., Vazquez-Martin, A., Ortega, F.J., and Fernandez-Real, J.M. (2009) Fatty acid synthase: Association with insulin resistance, type 2 diabetes, and cancer. *Clinical Chemistry* , **55**, 425–438.
24. Shi, H., Yu, X., Li, Q., Ye, X., Gao, Y., Ma, J., Cheng, J., Lu, Y., Du, W., Du, J., *et al.* (2012) Association Between PPAR-γ and RXR-α Gene Polymorphism and Metabolic Syndrome Risk: A Case-Control Study of a Chinese Han Population. *Archives of Medical Research* , **43**, 233–242.
25. Gr??n, F. and Blumberg, B. (2006) Environmental obesogens: Organotins and endocrine disruption via nuclear receptor signaling. *Endocrinology* , **147**, 50–55.
26. Lenhard, J.M. (2001) PPAR gamma/RXR as a molecular target for diabetes. *Receptors & channels* , **7**, 249–258.
27. Xi, Y. and Li, W. (2009) BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* , **10**, 232.
28. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* , **462**, 315–322.
29. Gertz, J., Varley, K.E., Reddy, T.E., Bowling, K.M., Pauli, F., Parker, S.L., Kucera, K.S., Willard, H.F., and Myers, R.M. (2011) Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genetics* , **7**, e1002228.
30. Carmona, J.J., Izzi, B., Just, A.C., Barupal, J., Binder, A.M., Hutchinson, J., Hofmann, O., Schwartz, J., Baccarelli, A., and Michels, K.B. (2013) Comparison of multiplexed reduced representation bisulfite sequencing ( mRRBS ) with the 450K Illumina Human BeadChip : from concordance to practical applications for methylomic profiling in epigenetic epidemiologic studies. *Epigenetics & Chromatin* , **6**, P36.
31. Illingworth, R.S. and Bird, A.P. (2009) CpG islands--'a rough guide'. *FEBS Lett* , **583**, 1713–1720.
32. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans (2015) *Science* , **348**, 648–660.
33. Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., and Eskin, E. (2008) Efficient control of population structure in model organism association mapping. *Genetics* , **178**, 1709–1723.
34. Zou, J., Lippert, C., Heckerman, D., Aryee, M., and Listgarten, J. (2014) Epigenome-wide association studies without the need for cell-type composition. *Nature Methods* , **11**, 309–311.
35. Civelek, M., Hagopian, R., Pan, C., Che, N., Yang, W.P., Kayne, P.S., Saleem, N.K., Cederberg, H., Kuusisto, J., Gargalovic, P.S., *et al.* (2013) Genetic regulation of human adipose microRNA expression and its consequences for metabolic traits. *Human Molecular Geneticso* , **22**, 3023–3037.
36. Weisberg, S.P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R.L., and Ferrante, A.W. (2003) Obesity is associated with macrophage accumulation in adipose tissue. *Journal of Clinical Investigation* , **112**, 1796–1808.
37. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., *et al.* (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology* , **28**, 1045–1048.
38. Horvath, S. (2015) Erratum to: DNA methylation age of human tissues and cell types. *Genome Biology* , **16**, 96.
39. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S.V., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., *et al.* (2013) Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Molecular Cell* , **49**, 359–367.

40. Sacks, D.B. (2011) A1C versus glucose testing: A comparison. *Diabetes Care* , **34**, 518–523.
41. Arner, P., Sinha, I., Thorell, A., Rydén, M., Dahlman-Wright, K., and Dahlman, I. (2015) The epigenetic signature of subcutaneous fat cells is linked to altered expression of genes implicated in lipid metabolism in obese women. *Clinical epigenetics* , **7**, 93.
42. Fernandez-Real, J.M., Menendez, J.A., Moreno-Navarrete, J.M., Bluher, M., Vazquez-Martin, A., Vazquez, M.J., Ortega, F., Dieguez, C., Fruhbeck, G., Ricart, W., *et al.* (2010) Extracellular fatty acid synthase: a possible surrogate biomarker of insulin resistance. *Diabetes* , **59**, 1506–1511.
43. Mayas, M.D., Ortega, F.J., Macías-González, M., Bernal, R., Gómez-Huelgas, R., Fernández-Real, J.M., and Tinahones, F.J. (2010) Inverse relation between FASN expression in human adipose tissue and the insulin resistance level. *Nutrition & metabolism* , **7**, 3.
44. Hendriks, W.H., O'conner, S., Thomas, D. V., Rutherfurd, S.M., Taylor, G.A., and Guilford, W.G. (2000) Structure of the intact ppar-γ-rxr-α nuclear receptor complex on dna. *Journal of the Royal Society of New Zealand* , **30**, 105–111.
45. Inouye, M., Ripatti, S., Kettunen, J., Lyytikainen, L.P., Oksala, N., Laurila, P.P., Kangas, A.J., Soininen, P., Savolainen, M.J., Viikari, J., *et al.* (2012) Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet* , **8**, e1002907.
46. Comuzzie, A.G., Cole, S.A., Laston, S.L., Voruganti, V.S., Haack, K., Gibbs, R.A., and Butte, N.F. (2012) Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* , **7**, e51954.
47. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Magi, R., *et al.* (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genetics* , **42**, 949–960.
48. El Kasmi, K.C., Smith, A.M., Williams, L., Neale, G., Panopoulos, A.D., Watowich, S.S., Hacker, H., Foxwell, B.M., and Murray, P.J. (2007) Cutting edge: A transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway. *J. Immunol.* , **179**, 7215–7219.
49. Maruyama, K., Uematsu, S., Kondo, T., Takeuchi, O., Martino, M.M., Kawasaki, T., and Akira, S. (2013) Strawberry notch homologue 2 regulates osteoclast fusion by enhancing the expression of DC-STAMP. *The Journal of Experimental Medicine* , **210**, 1947–1960.
50. Zhang, F.F., Cardarelli, R., Carroll, J., Fulda, K.G., Kaur, M., Gonzalez, K., Vishwanatha, J.K., Santella, R.M., and Morabia, A. (2011) Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. *Epigenetics* , **6**, 623–629.
51. McRae, A.F., Powell, J.E., Henders, A.K., Bowdler, L., Hemani, G., Shah, S., Painter, J.N., Martin, N.G., Visscher, P.M., and Montgomery, G.W. (2014) Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biology* , **15**, R73.
52. Smith, Z.D., Gu, H., Bock, C., Gnirke, A., and Meissner, A. (2009) High-throughput bisulfite sequencing in mammalian genomes. *Methods* , **48**, 226–232.
53. Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.Y., and Pellegrini, M. (2013) BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* , **14**, 774.
54. Akalin, A., Fredman, D., Arner, E., Dong, X., Bryne, J.C., Suzuki, H., Daub, C.O., Hayashizaki, Y., and Lenhard, B. (2009) Transcriptional features of genomic regulatory blocks. *Genome Biology* , **10**, R38.
55. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* , **7**, 500–507.
56. Martens, J.H. and Stunnenberg, H.G. (2013) BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* , **98**, 1487–1489.
57. Lawson, C.L. and Hanson, R.J. (1995) Solving least squares problems SIAM.
58. D'Agostino, R.B. (1971) An omnibus test of normality for moderate and large sample size. *Biometrika* , **58**, 341–348.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59. D'Agostino E. S, R. and P. (1973) Testing for departures from normality. *Biometrika* , **60**, 613–622.
60. Meinshausen, N. and Bühlmann, P. (2009) Stability selection. *Journal of the Royal Statistical Society* , **72**, 1–30.
61. Fabian Pedregosa Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duches, G.V. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* .
62. Van Der Walt, S., Colbert, S.C., and Varoquaux, G. (2011) The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering* , **13**, 22–30.
63. Aivazis, K.J.M.M. (2011) Python for Scientists and Engineers. *Computing in Science & Engineering* , **13**, 9–11.
64. McKinney, W. (2011) pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing* .
65. Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Lukauskas, S., Gemperline, D.C., Augspurger, T., Halchenko, Y., Cole, J.B., Warmenhoven, J., *et al.* (2017) mwaskom/seaborn: v0.8.1 (September 2017). 10.5281/zenodo.883859.
66. Hunter, J.D. (2007) Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* , **9**, 99–104.
67. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* , **28**, 495–501.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Table 1. Clincial Trait EWAS**

| EWAS locus | Chr | Bp start | Bp end | Candidate Genes(s) | *cis*-eQTL | Intra/inter-genic | EWAS clinical trait | EWAS Pval | EWAS beta (effect size) | EWAS Inflation factor | Methylation range |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Met1 | 10 | 34545726 | 34545903 | *LINC01317, LINC01320* | | Intergenic, 23kb, 356kb | MATSUDA insulin sensitivity index | 7.53E-19 | 39.3 | 0.99 | 16.10% |
| Met2 | 10 | 47979865 | 47980000 | *MSH2, MSH6* | *MSH2 cis*-eQTL (p=3.3e-94, rs2303425 ), and *MSH6 cis*-eQTL (p=1.1e-13, rs2134056) | Intergenic, 350kb, 30kb | Waist circumference | 3.26E-08 | -4.74 | 0.99 | 68.52% |
| Met2 | | | | | | | Fat mass (%) | 6.08E-10 | -5.27 | 1 | |
| Met2 | | | | | | | Fat free mass (%) | 6.11E-10 | 5.27 | 0.99 | |
| Met2 | | | | | | | Body mass index | 7.75E-09 | -5.04 | 1.01 | |
| Met3 | 11 | 65064793 | 65064867 | *AFTPH, SLC1A4* | *SLC1A4* cis- eQTL (p=1.6e-10, chr2:65252385) | Intergenic, 244kb, 150kb | Waist circumference | 4.92E-08 | -4.44 | 0.99 | 58.28% |
| Met3 | | | | | | | Fat mass (%) | 3.89E-09 | -4.82 | 1 | |
| Met3 | | | | | | | Fat free mass (%) | 3.57E-09 | 4.83 | 0.99 | |
| Met3 | | | | | | | OGTT fasting plasma insulin | 2.25E-09 | -4.85 | 0.97 | |
| Met3 | | | | | | | Body mass index | 5.71E-10 | -5.05 | 1.01 | |
| Met3 | | | | | | | MATSUDA insulin sensitivity index | 8.30E-09 | 4.76 | 0.99 | |
| Met3 | | | | | | | HOMAIR Insulin resistance index based on HOMA | 2.19E-09 | -4.85 | 0.97 | |
| Met4 | 12 | 7774215 | 7774379 | *AFAP1* | *AFAP1 cis*- eQTL (p=1.16e-16, rs34072960) | Intragenic | Fat mass (%) | 4.43E-08 | 6.59 | 1 | 78.69% |
| Met4 | | | | | | | Fat free mass (%) | 4.42E-08 | -6.59 | 0.99 | |
| Met4 | | | | | | | Body mass index | 5.57E-08 | 6.65 | 1.01 | |
| Met5 | 12 | 173132627 | 173132883 | *BOD1, CPEB4* | *BOD cis*-eQTL (p=1.3e-6,rs60748211). *CPEB4 cis*-eQTL (p=2.4e-174, rs72812818) | Intergenic, 88kb, 182kb | OGTT fasting plasma insulin | 1.97E-08 | 5.95 | 0.97 | 77.70% |
| Met5 | | | | | | | OGTT 120 min plasma insulin | 2.45E-09 | 6.07 | 0.98 | |
| Met5 | | | | | | | MATSUDA insulin sensitivity index | 4.42E-10 | -6.81 | 0.99 | |
| Met5 | | | | | | | HOMAIR Insulin resistance index based on HOMA | 1.04E-08 | 6.1 | 0.97 | |
| Met6 | 17 | 3132000 | 3132119 | *BPHL* | *BPHL cis*-eQTL (p=5.0e-81, rs7765391) | Intragenic | Weight | 9.22E-08 | 11.16 | 0.99 | 19.17% |
| Met7 | 17 | 137263485 | 137263573 | *RXRA* | *RXRA cis*-eQTL (p=1.0e-09, rs62576325) | Intragenic | Waist circumference | 6.08E-08 | 3.41 | 0.99 | 85% |
| Met8 | 17 | 75600604 | 75600822 | *CAMK2G* | *AMK2G cis*- eQTL (p=3.9e-28, rs2675671) | Intragenic | OGTT 120 min plasma proinsulin | 9.54E-08 | 18.43 | 0.96 | 10.40% |
| Met9 | 17 | 100183346 | 100183441 | *HPS1* | *HPS1 cis*-eQTL (p=1.1e-53, rs701801) | Intragenic | Waist circumference | 1.55E-11 | 10.02 | 0.99 | 30% |
| Met10 | 17 | 65248602 | 65248737 | *FRMD8, SCYL1* | | Intergenic, 67kb, 43kb | OGTT fasting plasma insulin | 6.84E-08 | -16.14 | 0.97 | 30% |
| Met10 | | | | | | | HOMAIR Insulin resistance index based on HOMA | 5.45E-08 | -16.25 | 0.97 | |
| Met11 | 17 | 42883183 | 42883320 | *PRICKLE1* | | Intragenic | OGTT 30 min plasma insulin | 1.70E-19 | -17.2 | 0.81 | 17.64% |
| Met12 | 19 | 113734359 | 113734579 | *TPCN1* | *TPCN1 cis*- eQTL (p=1.4e-6, rs2004720) | Intragenic | Weight | 7.24E-08 | 5.72 | 0.99 | 13.24% |
| Met12 | | | | | | | Glomerular filtration rate (eGFRcockrof) | 6.24E-08 | 5.61 | 0.97 | 71.39% |
| Met13 | 21 | 2893657 | 2893829 | *RAP1GAP2* | | Intragenic | OGTT fasting plasma insulin | 2.93E-08 | -11.03 | 0.97 | 30.25% |
| Met13 | | | | | | | HOMAIR Insulin resistance index based on HOMA | 1.96E-08 | -11.16 | 0.97 | |
| Met14 | 22 | 40913548 | 40913600 | *RAMP2* | | Intragenic | OGTT insulin area under the curve | 6.51E-29 | 12.06 | 0.99 | 32.53% |
| Met15 | 2 | 80041514 | 80042067 | *FASN* | *FASN cis*-eQTL (p=9.3e-10, rs4239015 ) | Intragenic | Weight | 4.07E-09 | 19.42 | 0.99 | 12.18% |
| Met15 | | | | | | | Waist circumference | 8.54E-08 | 17.31 | 0.99 | |
| Met15 | | | | | | | Fat mass (%) | 1.36E-08 | 18.29 | 1 | |
| Met15 | | | | | | | Fat free mass (%) | 1.40E-08 | -18.27 | 0.99 | |
| Met15 | | | | | | | Body mass index | 3.05E-10 | 21.57 | 1.01 | |
| Met16 | 2 | 80046324 | 80047195 | *FASN* | | Intragenic | Body mass index | 1.18E-08 | 16.61 | 1.01 | 20.05% |
| Met17 | 2 | 80050198 | 80050365 | *FASN* | | Intragenic | Body mass index | 2.48E-10 | 8.01 | 1.01 | 50.31% |
| Met18 | 4 | 80051500 | 80053080 | *FASN* | | Intragenic | Weight | 4.08E-09 | 7.95 | 0.99 | 67.62% |
| Met18 | | | | | | | Waist circumference | 4.29E-10 | 8.33 | 0.99 | |
| Met18 | | | | | | | Fat mass (%) | 1.03E-08 | 7.85 | 1 | |
| Met18 | | | | | | | Fat free mass (%) | 9.81E-09 | -7.86 | 0.99 | |
| Met18 | | | | | | | OGTT fasting plasma insulin | 1.08E-08 | 7.66 | 0.97 | |
| Met18 | | | | | | | OGTT 120 min plasma insulin | 5.42E-09 | 8.03 | 0.98 | |
| Met18 | | | | | | | Body mass index | 1.37E-11 | 9.44 | 1.01 | |
| Met18 | | | | | | | MATSUDA insulin sensitivity index | 1.79E-08 | -8.05 | 0.99 | |
| Met18 | | | | | | | HOMAIR Insulin resistance index based on HOMA | 3.57E-09 | 7.94 | 0.97 | |
| Met19 | 5 | 1148096 | 1148232 | *SBNO2* | | Intragenic | Weight | 3.20E-08 | 8.54 | 0.99 | 40.68% |
| Met19 | | | | | | | Body mass index | 7.58E-08 | 8.49 | 1.01 | |
| Met20 | 6 | 16661536 | 16661714 | *NRIP1, USP25* | *NRIP1 cis*-eQTL (p=6.6e-85, rs2178895) | Intergenic, 224kb, 440kb | Body mass index | 3.27E-08 | 8.14 | 1.01 | 86.15% |
| Met21 | 9 | 28151964 | 28152032 | *MN1* | | Intragenic | OGTT 120 min plasma insulin | 9.16E-08 | 7.64 | 0.98 | 30.95% |

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Table 2. PC1 Feature Contribution**
Variance Explained = 0.293

| Feature | Contribution |
|---|---|
| Glucose Baseline | 0.666 |
| Glucose 120min | 0.032 |
| Elevated Blood Glucose | 0.175 |
| LDL | -0.462 |
| Urine Albumin Baseline | 0.417 |
| Urine Albumin 60min | -0.371 |

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Figure 1. Epigenome-wide association of metabolic clinical traits

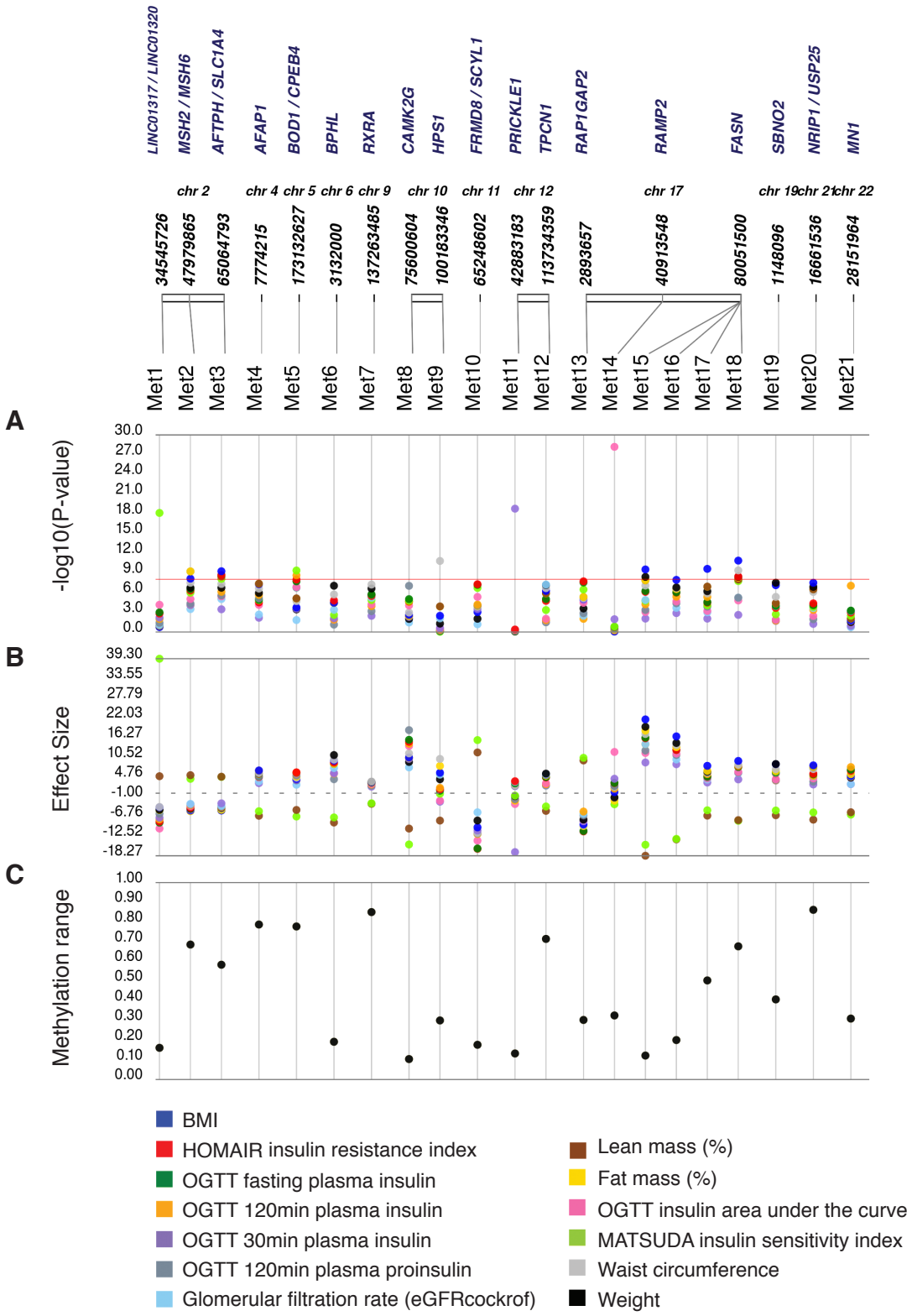



**Figure 1**. Epigenome-wide association of metabolic clinical traits. Association between DNA CpG methylation and clinical traits. (A) "PheWAS" plot showing association of each of the methylation loci (Met 1-21) and the clinical traits. The genomic location of the CpG is on the x-axis and the association significance is on the y-axis. Different colors represent different traits. (B) The effect size for each association shown in (A). (C) For each methylation locus on the x-axis, the range of methylation across individuals in the population is shown on the y-axis.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Figure 2. *FASN* is associated with multiple clinical traits

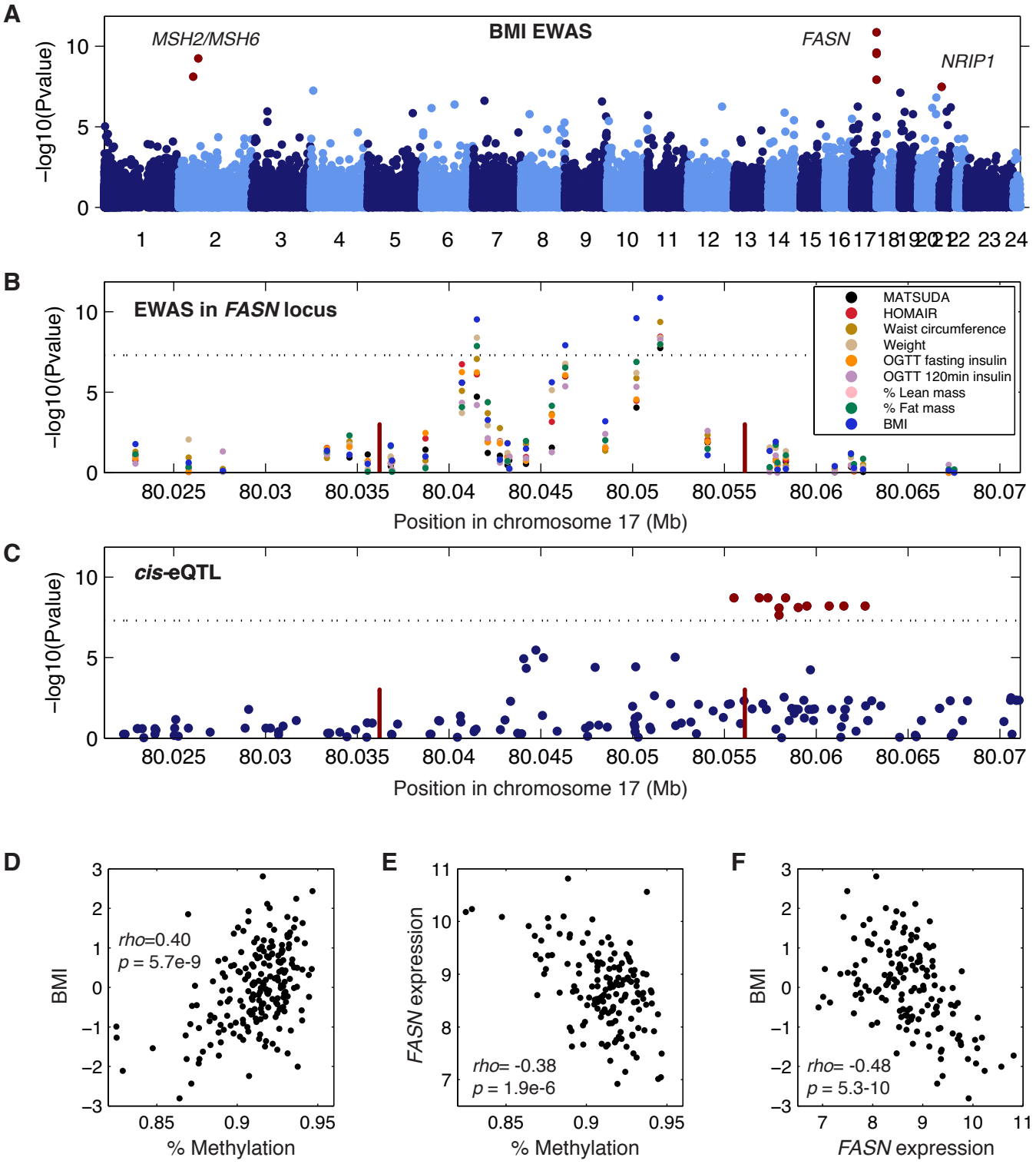



**Figure 2.** FASN is associated with multiple clinical traits. (A) Manhattan plot showing EWAS results for BMI. Each dot represents a CpG region with the genomic location of each CpG region on the x-axis and chromosomes shown in alternating colors. The association significance is on the y-axis and significant hits are shown as red dots. (B) Association results for multiple phenotypes near the FASN locus. Each dot represents a different association to a CpG region and different colored points represent distinct clinical traits. The genomic location of each CpG region is on the x-axis and the association significance is on the y-axis. Red vertical bars denote the transcription start and end of FASN. The dotted significance threshold line is drawn at 5x10-8. (C) cis-eQTL results for FASN expression in adipose tissue biopsies. Each dot represents a SNP. the genomic location of each SNP is on the x-axis and the association significance is on the y-axis. Significant SNPs are shown as red dots. Red vertical bars denote the transcription start and end of FASN. The dotted significance threshold line is drawn at 5x10-8. (D)-(F) Each point represents an individual in the cohort, showing correlation between (D) methylation levels for the peak associated CpG region and BMI, (E) methylation levels for the peak CpG region and expression of FASN, and (F) expression of FASN and BMI.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Figure 3. *SLC1A4* is associated with multiple clinical traits

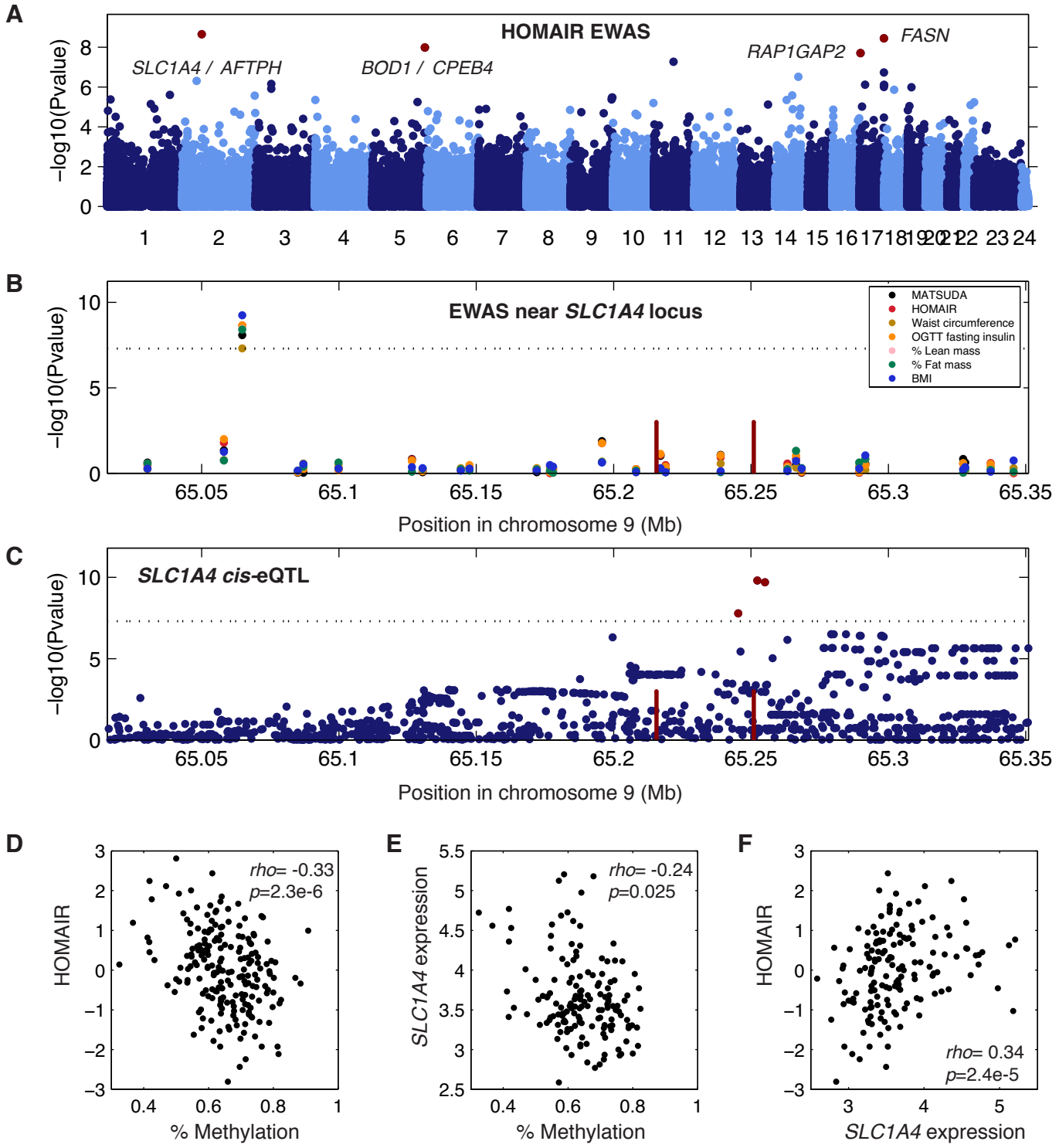



**Figure 3**. SLC1A4 is associated with multiple clinical traits. (A) Manhattan plot showing EWAS results for insulin resistance index HOMAIR. Each dot is a CpG region, the genomic location of each CpG region is on the x-axis with chromosomes shown in alternating colors, the association significance is on the y-axis, significant hits are shown as red dots. (B) Association results for multiple phenotypes near the SLC1A4 locus. Each dot represents a different association to a CpG region, different colored points represent distinct clinical traits, the genomic location of each CpG region is on the x-axis, the association significance is on the y-axis. Red vertical bars denote the transcription start and end of SLC1A4. The dotted significance threshold line is drawn at 5x10-8. (C) cis-eQTL results for SLC1A4 expression in adipose tissue biopsies. Each dot represents a SNP, the genomic location of each SNP is on the x-axis, the association significance is on the y-axis, significant SNPs are shown as red dots. Red vertical bars denote the transcription start and end of SLC1A4. The dotted significance threshold line is drawn at 5x10-8. (D)-(F) Each point represents an individual in the cohort, showing correlation between (D) methylation levels for the peak associated CpG region and HOMAIR, (E) methylation levels for the peak CpG region and expression of SLC1A4, and (F) expression of SLC1A4 and HOMAIR.

bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Figure 4. Cell-type deconvolution

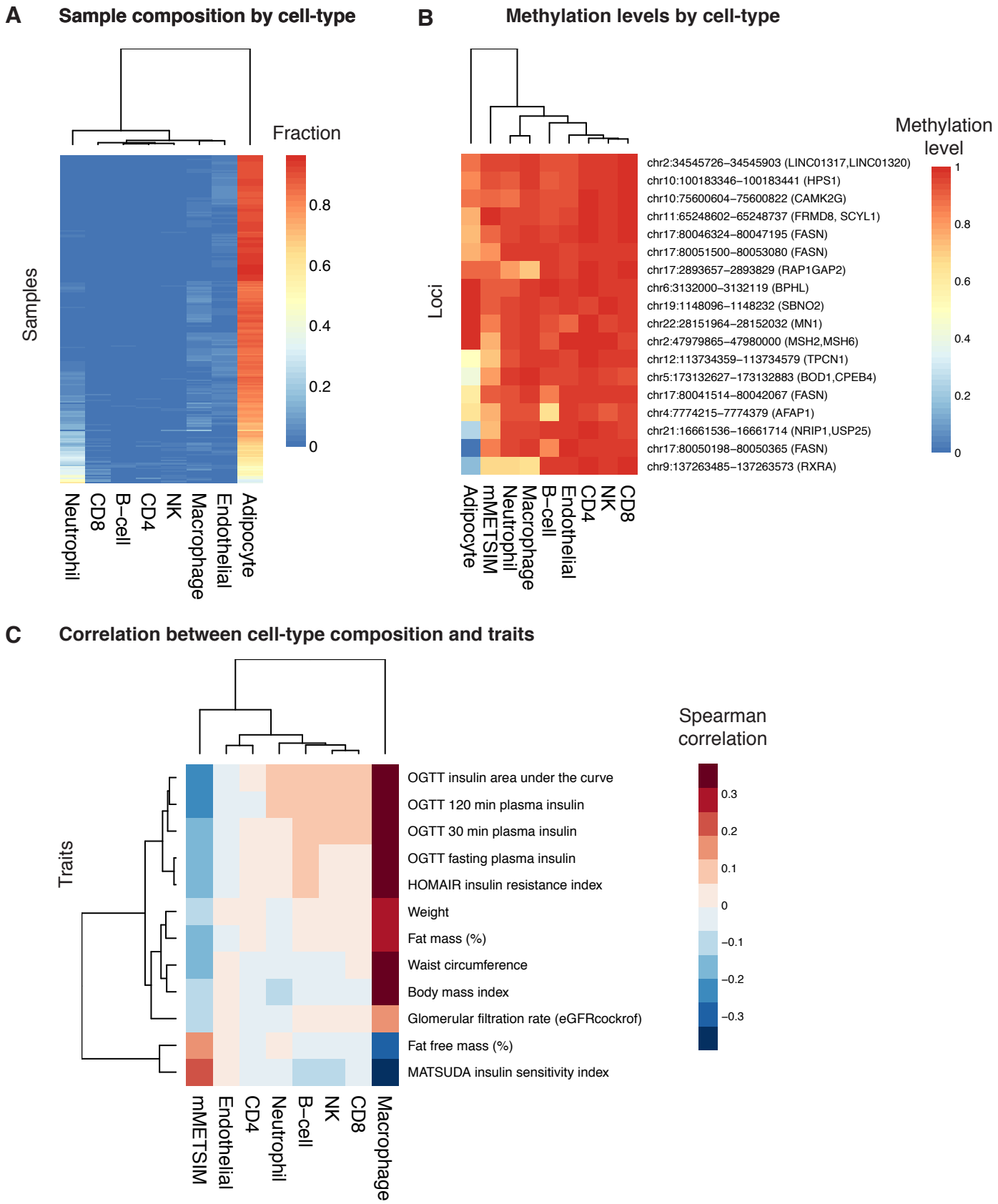
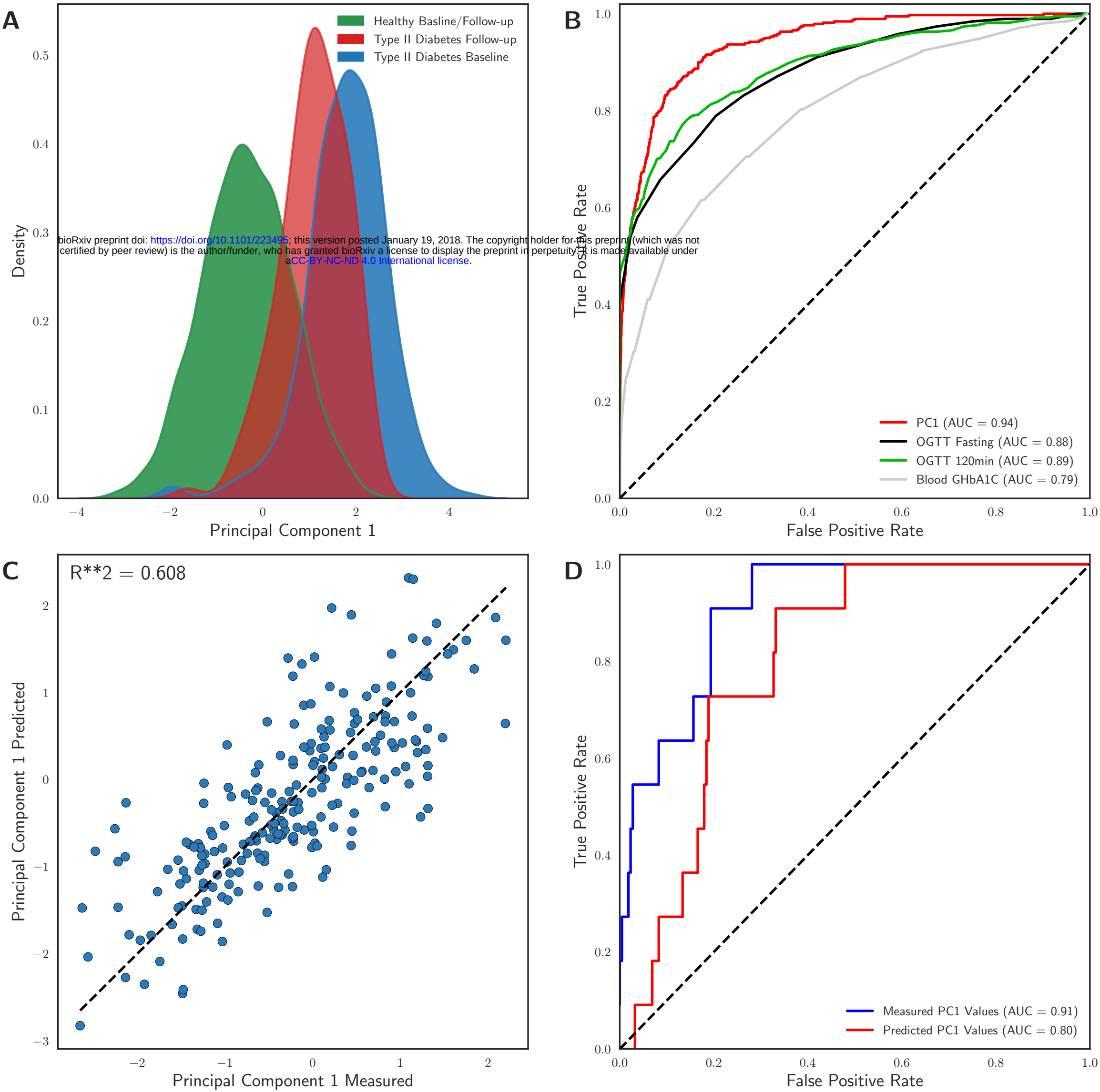



**Figure 4**. Cell-type deconvolution. (A) Sample composition by cell-type is shown for different cell types (columns), across all METSIM samples (rows). The color in the heatmap represents the relative fraction that each cell-type contributes to the total in each sample. (B) For each methylation locus (rows), the methylation levels in METSIM samples or for different cell-types (columns) are shown in the heatmap, the color represents the methylation levels. (C) Correlation between cell-type composition and clinical trait. Traits are plotted in rows, and cell-types are plotted in columns. The color in the heatmap represents spearman correlation between the fraction derived from each cell-type and a clinical trait for and an individual, across all individuals in the METSIM samples.

# Figure 5. A Methylation Biomarker to Assess Type II Diabetes




bioRxiv preprint doi: https://doi.org/10.1101/223495; this version posted January 19, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Figure 5. Methylation Biomarker to Assess Type II Diabetes.**(A) Kernel density estimates of principal component 1 by Type II diabetes status at baseline and follow-up examination. Healthy individuals had not been diagnosed with type II diabetes at baseline or follow-up examination, Type II Diabetes Follow-up individuals had not received a type II diabetes diagnosis at baseline but were diagnosed by follow-up examination, and Type II Diabetes Baseline individuals received a type II diabetes diagnosis before or at baseline examination. (B) A combined type II diabetes feature, PC1, outperforms individual features for classification of diabetes at baseline or follow-up examination among all METSIM samples (n= 10,172). (C) Measured and predicted PC1 values for cross-validated regression models fit to 1456 CpG sites (n=228). (D) ROC curves for the measured and predicated PC1 values for classification of type II diabetes status at baseline or follow-up examination (n=228).