



**HAL**  
open science

## Personalized Epigenomic Signatures That Are Stable Over Time and Covary with Body Mass Index

Delphine Fradin, Rafael A Irizarry, Martin J Aryee, Peter Murakami, Thor Aspelund, Gudny Eiriksdottir, Tamara B. Harris, Lenore Launer, Vilmundur Gudnason, M. Daniele Fallin, et al.

► **To cite this version:**

Delphine Fradin, Rafael A Irizarry, Martin J Aryee, Peter Murakami, Thor Aspelund, et al.. Personalized Epigenomic Signatures That Are Stable Over Time and Covary with Body Mass Index. *Science Translational Medicine*, 2010, 2 (49), pp.49ra67-49ra67. 10.1126/scitranslmed.3001262 . hal-02329984

**HAL Id: hal-02329984**

**<https://hal.science/hal-02329984v1>**

Submitted on 29 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Personalized Epigenomic Signatures Stable Over Time and Covarying with Body Mass Index**

Delphine Fradin<sup>1\*</sup>, Rafael A. Irizarry<sup>2,3\*</sup>, Martin J. Aryee<sup>2,8</sup>, Peter Murakami<sup>3</sup>, Thor Aspelund<sup>4,5</sup>, Gudny Eiriksdottir<sup>4</sup>, Tamara B. Harris<sup>6</sup>, Lenore Launer<sup>6</sup>, Vilmundur Gudnason<sup>4,5</sup>, Andrew P. Feinberg<sup>3†</sup>, and M. Daniele Fallin<sup>3,7†</sup>

\* These authors contributed equally

† These authors contributed equally

<sup>1</sup> Inserm, Paris, France; <sup>2</sup> Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore; <sup>3</sup> Center for Epigenetics, Johns Hopkins School of Medicine, Baltimore; <sup>4</sup> Icelandic Heart Association, Kopavogur, Iceland; <sup>5</sup> University of Iceland, Reykjavik Iceland; <sup>6</sup> Intramural Research Program, National Institute of Aging, Bethesda; <sup>7</sup> Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore; <sup>8</sup> Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore

Corresponding authors:

M. Daniele Fallin, Ph.D.  
Associate Professor of Epidemiology,  
Biostatistics, Medicine  
Director, Genetic Epidemiology  
Johns Hopkins Bloomberg School of Public  
Health  
615 N. Wolfe St., W6509  
Baltimore, MD 21205  
dfallin@jhsph.edu

Andrew P. Feinberg, M.D., M.P.H.  
King Fahd Professor of Molecular Medicine  
Director, Center for Epigenetics  
Chief, Division of Molecular Medicine,  
Department of Medicine  
Johns Hopkins University School of  
Medicine  
855 N. Wolfe St., Rangos 570  
Baltimore, MD 21205  
afeinberg@jhu.edu

## **Abstract**

**The epigenome represents non-sequence-based modifications heritable during cell division, such as DNA methylation, that may affect normal phenotypes and predisposition to disease. Here we have performed unbiased genome-scale analysis of ~4 million CpG sites in 74 individuals using comprehensive array-based relative methylation (CHARM) analysis. We find 227 regions with extreme inter-individual variability (variably methylated regions (VMRs)) across the genome, which are enriched for developmental genes based on Gene Ontology analysis. Furthermore, half of these VMRs are stable within individuals over an average of 11 years, and these VMRs define a personalized epigenomic signature. Four of these VMRs show covariation with body mass index consistently at two study visits and are located in or near genes previously implicated in regulating body weight or diabetes. This work suggests a novel epigenetic strategy for identifying patients at risk of common disease.**

## Introduction

A role for epigenetics in common disease has long been suspected [1], and a strong relationship has previously been shown for cancer [2]. We have argued that common disease involves both genetic and epigenetic factors, and that epigenetic modification could both mark environmental effects as well as mediate genetic effects [1]. In addition to particular studies of exposure-epigenetic relationships, evidence of epigenetic changes with aging support an environmental component to epigenetic variation [3,4,5,6]. Studies of identical twins show greater differences in global DNA methylation in older than in younger twins, supporting an age-dependent progression of epigenetic change [4,6]. Previously we showed global methylation changes over an 11 year span in participants of an Icelandic cohort from the Age, Gene/Environment Susceptibility (AGES)-Reykjavik Study [7]. Recently, Christensen showed age- and tissue-related alterations in some CpG islands based on an array of 1,413 arbitrarily chosen CpG sites near gene promoters [5], further supporting the evidence for dynamic methylation patterns over time. Other work, however, has suggested that epigenetic marks, or their maintenance, are themselves controlled by genes, and are thus heritable in the traditional sense and associated with particular DNA variants [8]. This would imply stability of methylation marks, rather than variation controlled by changing environments.

The work we will describe represents the first genome-scale, gene-specific analysis of DNA methylation in the same individuals over time. The results offer a solution to the conundrum of stable vs. time-sensitive differences in DNA methylation, as we will show both stable and dynamic sites of variable DNA methylation. Furthermore, the more stable sites of variation can identify a personalized epigenomic signature that may correlate with common genetic disease.

## Methods

**Samples.** Non-immortalized lymphocyte samples were taken from participants of the AGES Reykjavik Study, which is described in detail elsewhere [9]. 74 samples contributed to these analyses. These reflect samples meeting our high quality array data filter among a randomly chosen set of 100 samples from the 638 AGES participants that had ample DNA from two visits. CHARM data were only considered in analyses if they passed our internal quality assessment. For cross-sectional analyses of the most recent collection (visit 7), 64 samples contributed data, while 48 contributed to cross-sectional analyses of the earlier visit 6 data. For identification of dynamic VMRs, a subset of 38 samples had quality CHARM data at both time points. For the

analyses with BMI presented here, BMI was calculated as the body weight in kilograms (kg) divided by the height in meters (m) squared.

**Genome-wide methylation assay.** We performed comprehensive high-throughput array-based relative methylation (CHARM) analysis, which is a microarray-based method agnostic to preconceptions about methylation, including location relative to genes and CpG content [10,11]. The resulting quantitative measurements of methylation, denoted with  $M$ , are log ratios of intensities from total (Cy3) and McrBC-fractionated DNA (Cy5): positive and negative  $M$  values are quantitatively associated with methylated and unmethylated sites, respectively. For each sample we analyze ~4.5 million CpG sites across the genome using a custom designed NimbleGen HD2 microarray, including all of the classically defined CpG islands as well as all non-repetitive lower CpG density genomic regions of the genome. We include 4,500 control probes to standardize these  $M$  values so that unmethylated regions are associated, on average, with values of 0. CHARM is 100% specific at 90% sensitive for known methylation marks identified by other methods (e.g., in promoters), while including the more than half of the genome not identified by conventional region pre-selection. The CHARM results have also been extensively corroborated by quantitative bisulfite pyrosequencing analysis [10].

**Identification of VMRs.** We first screened the methylome for regions where methylation varied substantially across individuals. We term these variably methylated regions “VMRs”, to distinguish them from regions identified for their discrimination of groups, such as tissue types or cases versus controls, which we and others have previously called DMRs. Our use of the term “VMR” can be considered a specific type of “metastable epi-allele” introduced by Rakyan in 2002 to denote variable expression of imprinted loci or variable methylation of an agouti methylation variant [12].

To identify VMRs from our data, the raw CHARM data were first processed using the statistical procedure described by Aryee et al [13]. This statistical procedure produced quality metrics (percent between 0-100) for each sample and, for those that passed our quality test (>80%), a vector of methylation percentage estimates for each feature on the array. These were then smoothed to reduce measurement error using the standard CHARM approach [10,11]. We denote the resulting methylation percentages for subject  $i$  at microarray feature  $j$  for time  $t$  as  $M_{ijt}$ .

We used cross-sectional analysis of visit 7 data to identify polymorphic variably methylated regions (VMRs) based on extreme inter-individual variance across consecutive probes.

Specifically, we estimated between subject variability using the median absolute deviation (MAD), a robust estimate of the standard deviation [14]. We computed the median of  $|M_{ijt}-m_{jt}|$  across subjects, with  $m_{jt}$ , the median  $M_{ijt}$  across subjects  $i$ , and referred to it as  $s_{jt}$  [14]. To avoid false positives in subsequent analysis of correlations with covariates, we required a very stringent definition for designating a polymorphic VMR: a region of 10 or more consecutive probes attaining values of  $s_{jt}$  above the 99<sup>th</sup> percentile of all the  $s_{jt}$  and an average  $s_{jt} > 0.125$ . We chose these cut-off values using permutation tests. Specifically, we randomized the genomic order of the CHARM probes and applied the above algorithm to find VMRs (including the smoothing step) for each permuted data set. Using our criteria, we obtained 0 false positives. Lowering either the number of consecutive probes or the average  $s_{jt}$  thresholds produced false positives.

These VMRs were then annotated for genomic location and gene proximity. Genes within 3kb of VMRs were considered in a GO analysis of biological process categories. For each GO category, we performed a hypergeometric test [15], with corresponding nominal p value, to determine enrichment of genes near VMRs. We also calculated the false discovery rate for each category statistic, to account for the multiple comparisons.

**Identification of stable versus dynamic VMRs.** We generated methylation profiles for each sample using the average  $M_{ijt}$  within the range of each VMR. This included a vector of  $k$  VMR values for each subject  $i$  and time point  $t$ . We calculated  $D_{ik}$ , the median absolute within-person difference between methylation profiles from visit 6 to visit 7 for each VMR  $k$ . We then fit a two component Gaussian mixture model [16] to these values and used the resulting estimated posterior distributions to classify VMRs into three groups: “stable”: those with posterior probability of membership in the lower distribution  $> 0.99$ , reflecting little intra-individual change over time; “dynamic”: those with posterior probability of membership in the higher distribution  $> 0.99$ , reflecting those with high intra-individual change over time; and “ambiguous”: those not meeting either criteria, and thus in the overlap between the two distributions. (Note: Among the stable VMRs, there is some change over time observed in both directions, and when one takes the absolute value of this difference, the result is a small positive number, and thus the central tendency of  $D_k$  for stable VMRs is not zero.)

To evaluate discrimination of individuals based on patterns, we applied hierarchical clustering to the vectors of methylation values for the VMRs and graphed individuals into a dendrogram

based on similarity of VMRs. We then selected only those VMRs designated as “stable” in the analysis above and repeated the hierarchical clustering and dendrogram graphic.

**Identification of BMI-related methylated regions.** We performed cross-sectional analyses for data at each visit separately. For each stable VMR, we fit a linear regression model to summarize the relationship between BMI and methylation. Specifically, for each VMR  $k$ , we fit the following model:

$$Y_i = a_k + b_k M_{ik} + e_{ik}$$

with  $Y_i$  is BMI for individual  $i$ ,  $M_{ik}$  the methylation level for individual  $i$  in the  $k$ -th VMR, and  $e$  representing unexplained variability. Here  $b_k$  represents the parameter of interest that summarizes the correlation between BMI and methylation. This produced one Wald-statistic for each VMR. We fit this model to the data from visit 7 and to account for the multiple comparisons due to multiple VMRs, we reported a list of regions with a false discovery rate of 0.30. To confirm these results, we independently applied the same regression approach to visit 6 and obtained estimates of  $b$  along with p-values.

## Results

We report CHARM analyses on samples of the AGES study, assessing 4.5 million CpG sites genome-wide. In brief, the AGES study constitutes visit 7 (in 2002-2005) of the Reykjavik Study, which began with 18,000 residents of Reykjavik recruited in 1967. The AGES study recruited 5758 of the surviving members, who were aged 69–96 years in 2002. Of these, 638 gave a DNA sample in 1991 as part of the sixth Reykjavik Study visit, and therefore have DNA from two time points, about 11 years apart, available for methylation analysis. We present data for 74 samples, a random set of those who had ample DNA remaining for both study visits. Descriptive statistics for these samples are given in Table 1.

We performed Comprehensive High-Throughput Array-based Relative Methylation (CHARM) analysis, which has been shown to identify differential DNA methylation without assumptions regarding where such changes would be, and uses arrays tiled through regions based on their relative CpG content, including all CpG islands, as well as CpG island “shores” which have been shown to be enriched in differential methylation {Feinberg, #47}.

CHARM analysis in the visit 7 sample identified 227 regions meeting our criteria for polymorphic methylation patterns across individuals (variably methylated regions, VMRs). These represent

regions of extreme variability across individuals defined by 10 or more consecutive probes with an average standard deviation  $> 0.125$  (Supplementary Table 1). These VMRs showed enrichment for development and morphogenesis categories (Table 2), including genes from all four HOX clusters. The appearance of developmental genes is predicted by our model that epigenetic variation would involve developmental genes, and this variability itself increases evolutionary fitness in an environmentally changing world [14].

We then considered whether methylation at these regions changed within individuals over time. The distribution of the absolute value of average within-person change in methylation over time per VMR is shown in Figure 1, which implies two underlying distributions. This fit a two-component mixture model, with 41 VMRs easily classified into the higher intra-individual difference group (probability of membership in orange distribution  $> 0.99$ , Figure 1), which we label “dynamic VMRs”, 119 VMRs easily classified into the lower distribution (probability of green distribution  $> 0.99$ ), which we label “stable VMRs”, and 67 residing in the overlapping region and thus labeled “ambiguous” with respect to intra-individual change over time. Thus, approximately half the regions that are variably methylated across individuals appear to be stable over time within individuals.

These VMRs appear to act as an epigenetic signature. Figure 2A shows a dendrogram based on clustering of similarity in vectors of methylation values at VMRs. Use of only stable VMRs in the clustering algorithm uniquely identifies each individual (Figure 2B). These stable VMRs may represent polymorphic methylated regions that are not particularly susceptible to exposure modifications or that do not naturally change with age. One can postulate that these stable VMRs, sites that are variable across individuals but not due to exposures or age, may be the methylomic sites with variation most likely controlled by inherited mechanisms such as DNA variation or stochastic methylation reassignment in development.

Our results help focus epidemiologic hypotheses about how methylation may play a role in disease risk by allowing focused attention to particular VMRs in the methylome. In this vein, we explored the relationship between methylation and an easily measured phenotype, BMI, that is known to have many disease correlates. We identified 13 VMRs that met a false discovery rate (FDR) criteria of  $<25\%$  in cross-sectional analyses of visit 7 (Table 3). Of these, 4 had a  $P < 0.10$  and the same strength and direction of correlation with BMI at the earlier visit 6. These VMRs are in or near genes *PM20D1*, *MMP9*, *PRKG1*, and *RFC5*. To illustrate our findings, the methylation curves among obese ( $BMI \geq 30$ ) and normal ( $BMI < 25$ ) subjects for the VMR at

*PM20D1* are shown in Figure 3. Scatter plots for the relationship between methylation and BMI for all four VMRs showing significant correlations at both visits are shown in Figure 4.

## Discussion

We previously showed that global DNA methylation changes within individuals over time [7], and we have now confirmed and identified the locations of site-specific changes at “dynamic VMRs” using a genome-wide approach. In addition, we identified a separate set of “stable VMRs” that can be used to uniquely identify individuals, in an epigenetic signature akin to genetic fingerprinting. This signature may be correlated with disease status, implying an epigenetic signature can mark disease risk or disease states. In particular, we show stable VMRs that correlate with BMI at two separate visits a decade apart.

Some have argued that DNA methylation changes over time and is an important biological mediator of environmental effects on human disease, while others support the concept of inherited DNA methylation patterns, implying they are potentially variable across individuals but less likely to be dynamic over time. This has been a conundrum, since these appear to be opposing ideas. However, we show that both ideas have merit. It is important to identify these regions in the context of disease consequences, since those that are particularly labile may be the sites relevant when considering epigenetic marks as mediators of environmental effects, while those that are stable may be relevant as mediators or moderators of genetic effects. Further, those that do not change over time can be used as an epigenetic signature for an individual, similar to genotype. These regions can then be considered as candidates for assessment of methylation associations with disease or health-related phenotypes under specific risk models.

Our results help focus the integration of methylation measurement into epidemiologic studies of disease risk by providing specific genomic sites for inquiry. Our exploration of possible correlations between methylation at these VMRs and an easily measured disease-related phenotype, BMI, identified 13 genes, 4 of which were consistently correlated with BMI across two separate study visits. Remarkably, many of these have been previously implicated in obesity or diabetes. *MMP9*, as well as another member of this family, *MMP3*, encodes a metalloproteinase that have been shown to be upregulated in obese individuals [17]. Several MMPs, including *MMP9*, are upregulated in human adipocytes [18]. Matrix metalloproteinases have also been previously associated with obesity in rodent models [19,20]. Interestingly,

*PM20D1* is also a metalloproteinase, and, although not yet well-characterized, may have similar implications for obesity. *PRKG1*, a cGMP-dependent protein kinase, plays an important role in foraging behavior, food acquisition and energy balance [21]. *RFC5* is an intriguing gene as it encodes a metabolism-linked DNA replication complex loading protein, dysfunction of which leads to DNA repair defects. It might thus play a role in well-known but poorly understood DNA damage related complications of diabetes.

In an obese mouse model, *SORCS1* has been located on a type 2 diabetes quantitative trait locus (QTL) [22], and this has been confirmed in humans, where *SORCS1* SNPs and haplotypes were associated with fasting insulin secretion [23]. *IL1RAPL2* is located at a region on chromosome X that is associated with Prader-Willi like syndrome, while *DACH2* is also an X-linked gene associated with Wilson-Turner syndrome, both of which are Mendelian disorders with obesity features. *TTC13* is part of a family containing another tetratricopeptide repeat gene, *TTC8*, that has been directly linked to Bardet-Biedl syndrome, which includes obesity as a primary feature. *APCDD1* is a positional candidate gene associated with QTL that affects fat deposition in pigs [24], and is located at a region on chromosome 18 that is linked to body fat (%) in men [25].

Our identification of VMRs is of course limited by the number of individuals contributing to this genome-wide CHARM analysis. It is likely that increased sample sizes will improve detection of additional VMRs. Further, the dynamic VMRs defined here are based on an eleven year window among elderly participants. It is important to also identify methylomic regions that show intra-individual changes at early segments of the lifespan, and to connect these changes to particular environmental exposures. Also, our analyses were based on methylation patterns in DNA derived from blood, thus containing a mixture of cell types that could confound our results. However, in our previous study of global DNA methylation (i.e. non-site-specific) in these samples, we found no relationship between lymphocyte count and methylation [7]. A recent paper by Heijmans (2010) also showed that cellular heterogeneity was not associated with DNA methylation amounts for the majority of sites they studied [26]. Our use of blood as a DNA source may also limit the interpretations of these results, given the tissue specificity of DNA methylation. However, there is growing precedent for lymphoid tissues serving as a good surrogate tissue for changes in other target tissues. For example, loss of imprinting (LOI) of *IGF2*, one of the best studied disease-related epigenetic mutations, is found in both lymphocytes and colon, and changes of either are associated with increased colorectal cancer risk (Cui et al. Science 2003). Finally, our exploration of the correlation between BMI and

methylation was based on availability of quantitative data and relevance to human disease. We were unable to assess the relationship of VMRs to categorical outcomes in this sample that is, although larger than previous genome-wide site-specific methylation reports, is limited for categorical phenotypes. This study supports further examination of other measures of obesity, diabetes, and related health consequences with respect to the particular VMRs identified here.

These results have important implications. An individual epigenetic signature that is stable over time has not previously been described. Such a signature could be driven by underlying sequence variation, by early environmental exposure, e.g. prenatally, or both. Even if in part or completely genetically driven, this epigenotype may be more proximate to the ultimate phenotype, in this case body mass index, and thus have considerable value for disease risk assessment. While the sample size is larger than previous genome-scale gene-specific methylation studies, it is still relatively small compared to classical sequence-driven approaches such as GWAS. Even so, the data suggest that this epigenomic approach to disease phenotype will be an important complement to such studies. Even with these numbers, we could identify four genes with VMRs related to BMI. In addition, the identification of stable VMRs may have long term consequences for developing personalized epigenomics in medicine, that might reflect both one's genome and early (e.g. *in utero*) environment.

### **Acknowledgments**

This work was funded by the NIH grants 1R01ES015211 awarded for Dr. Fallin and 2P50HG003323 awarded to Dr. Feinberg. The AGES Reykjavik study is funded by NIH contract, N01-AG-12100, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament).

### **Author Contributions**

M.D.F., V.G., A.P.F. designed the study; D.F. performed CHARM assays; D.F., R.I., M.J.A., P.M., T.A., performed statistical analyses; V.G., G.E., T.H., L.L. designed the AGES study and recruited participants; D.F., R.I., M.D.F., A.P.F. interpreted the data and wrote the paper.

**Table 1. Descriptive Information (Mean (standard error)) for Samples Used in CHARM Analyses at Each Time Point**

|                     | <b>Visit 6<br/>(1991)</b> | <b>Visit 7<br/>(2002-2005)</b> |
|---------------------|---------------------------|--------------------------------|
| Age                 | 74.08 (3.49)              | 82.80 (3.45)                   |
| Sex (% male)        | 0.33                      | 0.31                           |
| BMI                 | 26.56 (3.81)              | 26.01 (4.10)                   |
| Glucose             | 0.08 (0.28)               | 0.11 (0.32)                    |
| Type 2 diabetes (%) | 5.90                      | 5.79                           |
| Coronary events (%) | 0.10                      | 0.14                           |
| Waist/hip ratio     | -                         | 0.66 (0.10)                    |
| Fat percent         | -                         | 29.31 (7.89)                   |
| Hemoglobin A1C      | -                         | 0.47 (0.07)                    |
|                     | N=48                      | N=64                           |

**Table 2. Gene Ontology Results with P<0.01 for 227 VMRs Identified**

| <b>Pvalue</b> | <b>FDR</b> | <b>Odds Ratio</b> | <b>Obs Count</b> | <b>Expected Count</b> | <b>GO Term</b>                         | <b>Genes</b>                            |
|---------------|------------|-------------------|------------------|-----------------------|--|---|
| 0.0011        | 0.222      | 7.04              | 5                | 0.79                  | Ant./post. pattern formation           | HOXA5; HOXB6; HOXD8; HOXC10; HOXA1      |
| 0.0019        | 0.222      | 43.31             | 2                | 0.07                  | blastoderm segmentation                | HOXB6; HOXD8                            |
| 0.0019        | 0.222      | 43.31             | 2                | 0.07                  | determ. anterior/post. axis, embryo    | HOXB6; HOXD8                            |
| 0.0082        | 0.256      | 17.31             | 2                | 0.14                  | neuron recognition                     | FOXG1; NTM                              |
| 0.0086        | 0.256      | 3.63              | 6                | 1.77                  | pattern specification process          | HOXA5; FOXG1; LEF1; HOXC10; MYF6; HOXA1 |
| 0.0096        | 0.256      | 7.47              | 3                | 0.44                  | placenta development                   | ESX1; LEF1; CDX4                        |
| 0.0096        | 0.256      | 15.74             | 2                | 0.15                  | intra-Golgi vesicle-mediated transport | COPZ1; GABARAPL2                        |

**Table 3. Stable VMRs Associated with BMI**

| Chr          | Nearest Gene  | Visit 7      |                |                     | Visit 6         |                     |
|--------------|---------------|--------------|----------------|---------------------|-----------------|---------------------|
|              |               | Qval         | Pval           | Regression Estimate | Pval            | Regression Estimate |
| chrX         | IL1RAPL2      | 0.114        | 0.00304        | -20.3               | 0.266           | -8.9                |
| <b>chr1</b>  | <b>PM2OD1</b> | <b>0.114</b> | <b>0.00332</b> | <b>7.6</b>          | <b>0.00824</b>  | <b>7.7</b>          |
| chr6         | NEDD9         | 0.114        | 0.00351        | 12.1                | 0.38            | 5.2                 |
| <b>chr20</b> | <b>MMP9</b>   | <b>0.160</b> | <b>0.00658</b> | <b>11.6</b>         | <b>0.0605</b>   | <b>8.9</b>          |
| chr10        | SORCS1        | 0.215        | 0.0128         | -13.6               | 0.112           | -9.4                |
| <b>chr10</b> | <b>PRKG1</b>  | <b>0.215</b> | <b>0.0132</b>  | <b>11.8</b>         | <b>0.000711</b> | <b>18.9</b>         |
| <b>chr12</b> | <b>RFC5</b>   | <b>0.243</b> | <b>0.0175</b>  | <b>-11.8</b>        | <b>0.0653</b>   | <b>-8.8</b>         |
| chr1         | TTC13         | 0.249        | 0.022          | 9.27                | 0.523           | 3.3                 |
| chrX         | DACH2         | 0.249        | 0.0311         | -15.1               | 0.539           | 4.1                 |
| chr5         | TRIM36        | 0.249        | 0.0326         | 11.3                | 0.0781          | -14.1               |
| chr14        | FLRT2         | 0.249        | 0.0278         | -9.5                | 0.19            | -5.8                |
| chr1         | C1orf57       | 0.249        | 0.0253         | -10.6               | 0.282           | -6.5                |
| chr18        | APCDD1        | 0.249        | 0.0332         | -10.7               | 0.901           | 0.7                 |

**Bold** values indicate confirmation in visit 6 analysis ( $p < 0.1$  and consistent regression parameter estimates); *italics* indicate conflicting directions of correlation with BMI

## Figure Legends

**Figure 1: Distribution of intra-individual change over time at VMRs.** Mixture distribution analysis shows  $D_k$ , the average absolute value of intra-individual differences in methylation over time, fits two underlying curves: little change, “stable” in green and larger changes, “dynamic” in orange.

**Figure 2: Similarity between individuals based on VMR methylation. Panel A:** Dendrogram based on clustering applied to methylation profiles at all 227 VMRs. **Panel B:** Dendrogram based on clustering applied to methylation profiles using only the 119 stable VMRs. Numbers represent individual IDs.

**Figure 3: Methylation Curves at PM20D1.** Methylation curves for visit 7 and visit 6 data. Dashed lines are individual methylation curves. Solid lines are average curves by obese (blue) and normal (red) groups. The green line indicates the boundaries of the VMR. CpG density is shown below with CpG islands marked in orange. Gene location shown at bottom.

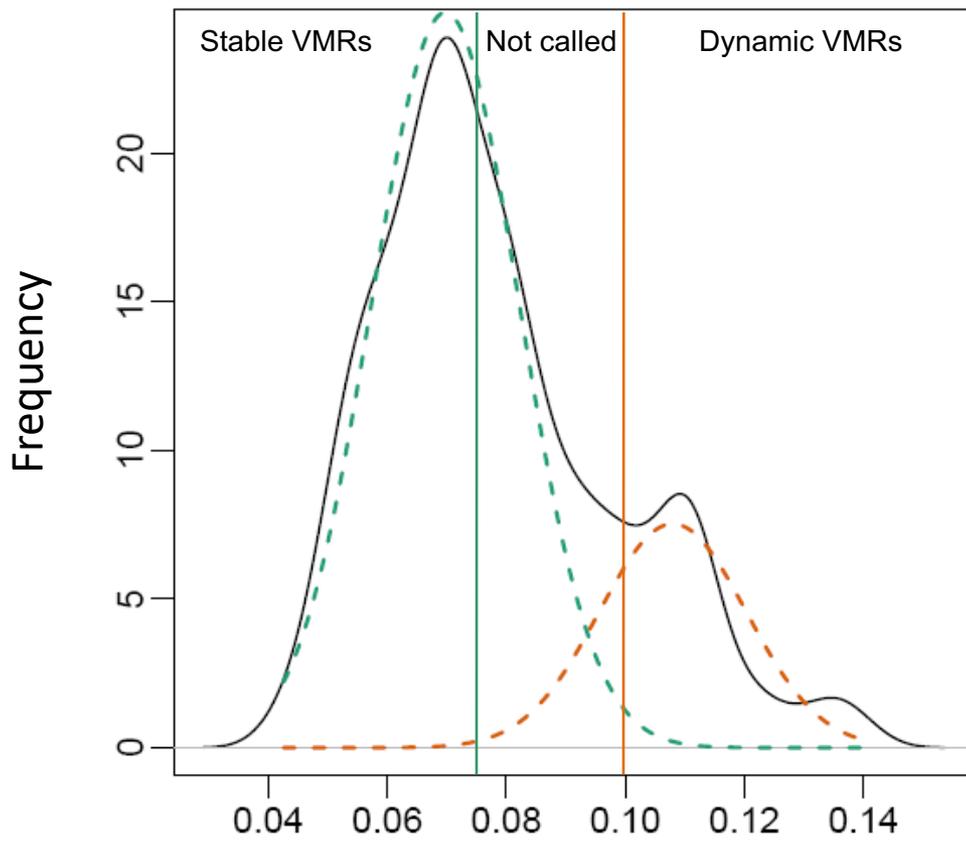
**Figure 4: Correlations Between Methylation and BMI at Six BMI-Related VMRs.** Points are individual IDs, blue indicates visit 7, red indicates visit 6.

## References

1. Bjornsson HT, Fallin MD, Feinberg AP (2004) An integrated epigenetic and genetic approach to common human disease. *Trends Genet* 20: 350-358.
2. Feinberg AP, Tycko B (2004) The history of cancer epigenetics. *Nat Rev Cancer* 4: 143-153.
3. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, et al. (2009) The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One* 4: e6767.
4. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 102: 10604-10609.
5. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, et al. (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 5: e1000602.
6. Poulsen P, Esteller M, Vaag A, Fraga MF (2007) The epigenetic basis of twin discordance in age-related diseases. *Pediatr Res* 61: 38R-42R.
7. Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, et al. (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 299: 2877-2883.
8. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE (2007) Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum Mol Genet* 16: 547-554.
9. Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, et al. (2007) Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* 165: 1076-1087.
10. Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, et al. (2008) Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* 18: 780-790.
11. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41: 178-186.
12. Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E (2002) Metastable epialleles in mammals. *Trends Genet* 18: 348-351.
13. Ayree M, Wu Z, Ladd-Acosta C, Herb B, Feinberg A, et al. (2010) Accurate genome-scale percentage DNA methylation estimates from microarray data. *Biostatistics*.
14. Feinberg AP, Irizarry RA (2009) Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* 107 Suppl 1: 1757-1764.
15. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23: 257-258.
16. Banfield JD, Raftery AE (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics* 49: 803-821.
17. Nair S, Lee YH, Rousseau E, Cam M, Tataranni PA, et al. (2005) Increased expression of inflammation-related genes in cultured preadipocytes/stromal vascular cells from obese compared with non-obese Pima Indians. *Diabetologia* 48: 1784-1788.
18. O'Hara A, Lim FL, Mazzatti DJ, Trayhurn P (2009) Microarray analysis identifies matrix metalloproteinases (MMPs) as key genes whose expression is up-regulated in human adipocytes by macrophage-conditioned medium. *Pflugers Arch* 458: 1103-1114.
19. Uemura S, Matsushita H, Li W, Glassford AJ, Asagami T, et al. (2001) Diabetes mellitus enhances vascular matrix metalloproteinase activity: role of oxidative stress. *Circ Res* 88: 1291-1298.

20. Chavey C, Mari B, Monthouel MN, Bonnafous S, Anglard P, et al. (2003) Matrix metalloproteinases are differentially expressed in adipose tissue during obesity and modulate adipocyte differentiation. *J Biol Chem* 278: 11888-11896.
21. Kaun KR, Sokolowski MB (2009) cGMP-dependent protein kinase: linking foraging to energy homeostasis. *Genome* 52: 1-7.
22. Clee SM, Yandell BS, Schueler KM, Rabaglia ME, Richards OC, et al. (2006) Positional cloning of *Sorcs1*, a type 2 diabetes quantitative trait locus. *Nat Genet* 38: 688-693.
23. Goodarzi MO, Lehman DM, Taylor KD, Guo X, Cui J, et al. (2007) *SORCS1*: a novel human type 2 diabetes susceptibility gene suggested by the mouse. *Diabetes* 56: 1922-1929.
24. Lee KT, Park EW, Moon S, Park HS, Kim HY, et al. (2006) Genomic sequence analysis of a potential QTL region for fat trait on pig chromosome 6. *Genomics* 87: 218-224.
25. Lewis CE, North KE, Arnett D, Borecki IB, Coon H, et al. (2005) Sex-specific findings from a genome-wide linkage analysis of human fatness in non-Hispanic whites and African Americans: the HyperGEN study. *Int J Obes (Lond)* 29: 639-649.
26. Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, et al. Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J*.

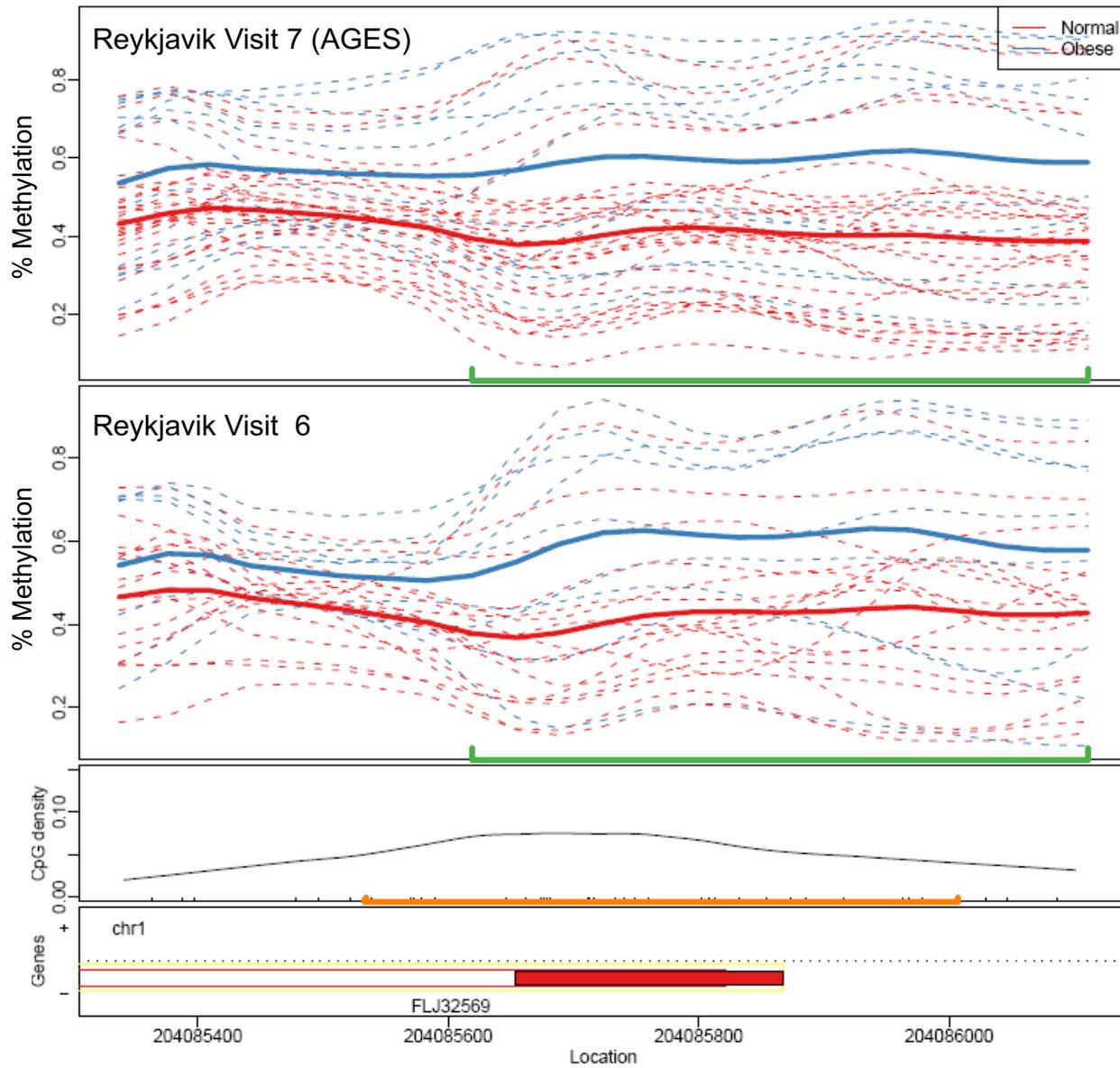
**Figure 1**



$D_k$  = Average within-person change over time for VMR k (absolute value)



# Figure 3



# Figure 4

