Supplementary Material for

Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci

Vitor Onuchic, Eugene Lurie, Ivenise Carrero, Piotr Pawliczek, Ronak Y. Patel, Joel Rozowsky, Timur Galeev, Zhuoyi Huang, Robert C. Altshuler, Zhizhuo Zhang, R. Alan Harris, Cristian Coarfa, Lillian Ashmore, Jessica W. Bertol, Walid D. Fakhouri, Fuli Yu, Manolis Kellis, Mark Gerstein, Aleksandar Milosavljevic*

*Corresponding author. Email: amilosav@bcm.edu

Published 23 August 2018 as *Science* First Release
DOI: 10.1126/science.aar3146

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/science.aar3146/DC1)

Table S3 as a separate Excel file

Materials and Methods

Variant calling from WGS data

The Roadmap Epigenomics Consortium generated WGS data on 11 tissue donors and 2 cell lines from the Roadmap Epigenomics Project, which we utilized for genome-wide identification of SNPs, insertions and deletions (INDELS), and copy number variants (CNVs). FastQC (v0.10.1) was used to trim adapter contigs and perform fastq quality control. The sequence data was aligned against human reference assembly GRCh37 using BWA mem *(56)* (v0.7.8). Picard (1.111) was used to remove PCR duplicate reads, after which INDEL realignment and base quality recalibration was performed using GATK (v3.2.2) *(57)*. The SNP and INDEL calling was done using goSNAP *(58)*, an ensemble joint calling pipeline, in which we employed three callers, Atlas-SNP (v1.4.3) *(59)*, GATK (v3.2.2) HaplotypeCaller (HC) *(60)*, and GATK UnifiedGenotyper (UG) for SNP calling. Four callers were utilized for INDEL calling: Atlas-Indel (v1.4.3), GATK-HC, GATK-UG, and Platypus (v0.7.9.1) *(61)*. With each caller we enforced the joint calling mode, across all 13 WGS sources, during both the SNP and INDEL discovery. Only the variants passing the default filter of each caller were retained. To achieve high sensitivity and specificity, as well as to reduce caller specific bias, we further applied a consensus strategy using goSNAP to generate a high quality SNP call set via 3-out-of-3 consensus. Similarly for INDELS, we generated a call set by taking the intersection of INDELS called by both Atlas-Indel and GATK-UG and then merged with INDELS called by GATK-HC. Rescoring and filtering was then performed on this set using Platypus. Genotype likelihood was then calculated for SNPs and INDELS using SNPTools BBMM algorithm *(62)* and Platypus, respectively.

We were able to assess the SNP call set quality of 5 (H9, STL001,STL002,STL003, skin02) of 13 of our WGS sources by comparing against their genotyping array data available on the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/?search=Genotyping&display=500). The average per sample SNP calling sensitivity, specificity, and false discovery rates were 94.63%, 99.87%, and 0.29%, respectively. The average genotype discordance rate was 0.14%, with a heterozygous genotype discordance rate of 0.27%. On average, we called 3.74 million SNPs per sample (98% in dbSNPv141), with a Ti/Tv ratio of 2.02, and 589.5 thousand INDELS per sample (92% in dbSNPv141).

CNVs may have significant impact on the set of detected heterozygosities as every SNP appearing in one of the copies of a duplicated region can be mistakenly called as heterozygous SNP when reads obtained from different copies of the same duplicated region are stacked together during their alignment to the reference genome. This situation occurs when the length of duplicated region is significantly larger than the length of reads. We detected false heterozygous SNPs, introduced by CNVs, by analyzing the value of read depth. Duplicated genome regions should have significantly larger coverage than “standard” regions, where reads originate only from single copies on two chromosomes. Furthermore, the coverage of “standard” regions should be twice as large as a coverage of regions with simple deletion, where sequence on one of the chromosomes is deleted. To detect duplicated regions, as well as simple deletions, we utilized a tool called CNVnator with the embedded GC-correction algorithm *(63)* (version 0.3). Analysis with the use of CNVnator was performed for every donor

separately based on available WGS data. The pipeline prepared by us was based on the description contained in the README file from the tarball with CNVnator sources. During computations, the whole reference genome was divided into sequences of adjacent regions (bins) with lengths equal to a user-specified "bin size" of 1000 bp. 1000 bp was chosen as all input WGS data was composed of paired-end 100 bp reads and the average distance between reads from a single pair equaled ~300 bp, we therefore assumed that a single read pair covers ~500 bp and that the value of "bin size" parameter should be set to 1000 bp (two times as long as "read size"). Average read depths were calculated for each bin separately, followed by the merging of bins with similar coverages. As a result, CNVnator returned a list of regions with coverage significantly different than the "standard" (i.e. coverage of standard regions). In the obtained results we noticed that some regions that contained unknown sequences (many base-pairs marked as "N") were called by CNVnator as deletions. To remove these false calls we calculated a list of regions containing significant amount of "N" base pairs and subtracted them from CNV results. Every returned region had assigned a special parameter RD, which equals the estimated ratio between measured coverage of this region and the "standard" coverage. To examine the influence of regions called by CNVnator as under- or over-covered on the set of found heterozygous SNPs, we calculated the density of occurrences of heterozygous SNPs in these regions (**fig. S12**).

Allele-specific methylation detection

Reads from WGBS experiments were mapped using Pash *(64)*, deduplication was performed using Picard, and local realignment around regions with potential INDELS and quality score recalibration were performed using Bis-SNP *(65)*. In the process of identification of allele-specific methylation, WGS variant calls were used both to determine the positions of heterozygous SNPs and to identify all homozygous CpG positions in the genome of each donor. With such information, and with the fully processed WGBS aligned reads, an in house script was then used to identify positions exhibiting significant allelic differences in CpG methylation.

Our script counted the number of times a methylated or unmethylated homozygous CpG occurred in the same read as each of the two possible alleles in the heterozygous SNP position for autosomal chromosomes. If the same read overlapped multiple CpGs, they were each considered as independent observations. Reads with a low quality score (Phred < 20) on the SNP position, or with a base call that did not match either of the two alleles expected in that position based on the WGS calls, were discarded. Due to the nature of bisulfite sequencing data, where cytosines may be observed as thymines due to bisulfite conversion, it was not possible to determine which allele the read came from in several cases. In such cases, the read was also discarded. If low quality score, or an unexpected base call was observed on a CpG position for a particular read, that observation did not contribute to the final counts. The significance of the association between the allele at the heterozygous SNP position and the methylation state of the CpGs in the 200bp surrounding region was assessed using Fisher's exact test. 200bp windows surrounding the heterozygous SNP position were chosen as the WGBS dataset was composed of single-end 100 bp reads. The test was only performed for heterozygous SNP positions that showed a minimum of 6 observations of either a methylated or unmethylated CpG position for both alleles. Once the p-values were computed for all

such heterozygous SNP positions, the Benjamini-Hochberg procedure was used to control the false discovery rate for such associations. The difference in the level of methylation between alleles was also computed for each heterozygous SNP. Any positions overlapping CNVs, or small INDELS, were discarded from further analyses. Additionally, any positions falling within ENCODE blacklisted regions were removed *(66)*. Finally, ASM calls were made by identifying the heterozygous SNP positions with false discovery rate (FDR) below a specified threshold (10%), and absolute difference in methylation between alleles above a minimum threshold (either 20 or 30%). Of the 49 samples with WGBS, 3 were discarded from individual analyses due to low coverage ($\leq$ 15X mean read depth coverage for the sample).

Allelic imbalance detection for histone marks and transcription

The identification of autosomal allelic imbalances for histone marks and transcription were performed using the previously published AlleleSeq pipeline *(19)*. Briefly, personal diploid genomes were built based on the WGS variant calls for each individual. Each read from ChIP-seq or RNA-seq datasets was aligned to both versions of the reference genome, and the best scoring mapping was kept. Duplicate reads were removed. For each heterozygous SNP position with at least 6 reads mapping to it, an even binomial test was used to determine whether there was a significant difference between alleles in the number of ChIP-seq or RNA-seq reads mapping to that position. Any positions that overlapped CNVs, INDELS, had low mappability, fell within ENCODE blacklisted regions *(66)*, or overlapped repetitive regions (defined using the RepeatMasker software package) were removed from further analyses. After multiple hypothesis adjustment using the Benjamini-Hochberg procedure, allelic imbalances in expression or in specific histone marks were called by identifying the positions with an FDR below a specified threshold (10%) and with the most prevalent allele representing at least 70% of the reads. Patterns of co-occurrence of allelic imbalances in DNA methylation and other epigenomic marks were examined within the same samples.

Detection of allelic imbalances within imprinted regions

To evaluate the accuracy of our imbalance calling method, we measured the enrichment of allelic imbalances around imprinted genes. The annotation of imprinting control regions and imprinted gene regions used were those previously defined *(67)*. For each methylome in our dataset, we used a Chi-squared test to determine whether ASM was significantly enriched around heterozygous SNPs falling within imprinting control regions compared to heterozygous SNPs falling outside of those regions. Only heterozygous SNPs accessible for detection (those having at least 6 counts per allele) were included. For all other types of epigenomic data (including RNA-seq), enrichment for allelic imbalances was investigated within imprinted gene clusters with an additional 10kb extension up- and downstream and using heterozygous SNPs falling outside of such regions as controls. Enrichment for three histone marks (H3K4me1, H3K36me3, and H3K27me3) could not be reliably ascertained due to limited numbers of heterozygous loci with sufficient read coverage in these datasets.

Assaying ASM co-occurrence and directionality consistency

To address the issue of whether ASM events tend to co-occur across samples, we identified in each pair of methylomes from our dataset, all heterozygous SNPs that were present in both samples, and had at least 20 homozygous CpG observations per allele in both of them. To further account for differences in detection power associated with differences in coverage, we down sampled the allelic methylation counts in the sample with the highest number of observations to match the number of observations in the other sample. ASM was then called in each sample using a threshold of 10% FDR and a minimum of 30% difference in methylation.

For each pair of samples, we then determined the degree of co-occurrence of ASM by computing the proportion of all loci with ASM in at least one of the two samples that also had ASM in the other sample. Expected values of ASM co-occurrence for each pair of samples were computed based on the number of heterozygous loci found in both samples passing the coverage thresholds, the numbers of ASMs observed in each sample over those loci, and on the assumption that the distributions of ASM events over those loci was independent between the two samples.

Using this set of ASM calls we were also able to determine whether the difference in methylation between alleles for a heterozygous SNP found in both samples was in the same (same allele having higher or lower methylation in both samples) or in different directions (**fig. S6C**).

Gaussian mixture modeling of sub-threshold ASM

For heterozygous SNPs around which ASM was observed in one methylome but not in the other, the distribution of methylation differences between alleles in the sample without ASM had one peak near zero and another around 30%, with the difference in the same direction as the sample with ASM. To quantitate the degree of sub-threshold ASM, we fitted a Gaussian Mixture Model with two components using expectation maximization (mixtools package in R) to those values of differences in methylation between alleles. We then assigned each of those co-occurring heterozygous loci to one of the two Gaussian component models based on their posterior probabilities for each locus.

Cross-methylome SD-ASM calling

For every autosomal locus displaying exactly two alleles across all the individuals in our dataset, we took all WGBS reads mapping to that position in any sample and counted the number of co-occurrences of either methylated or unmethylated homozygous CpGs in the same read as each of the possible alleles across all 49 methylomes. Note that samples with a homozygous locus were included as long as the total number of alleles across all individuals present at the locus was two. Fisher's exact test was then used to determine whether the allele at that position was significantly associated with the methylation status of the surrounding CpGs. The test was only performed for heterozygous SNP positions that showed a minimum of 6 observations of either a methylated or unmethylated CpG position for both alleles and 6 mapped reads. Once a p-value was computed for all such heterozygous SNP positions, the Benjamini-Hochberg procedure was used to control the FDR for such associations. The difference in the level of methylation between alleles was also computed for each heterozygous SNP. As for the per methylome ASM calling, any

positions overlapping CNVs or small INDELS in any individual were discarded. Similarly, any positions falling within ENCODE blacklisted regions were removed *(66)*.

Assaying SD-ASM occurrence within regulatory regions

The combined set of methylomes was utilized to assess the degree of SD-ASM occurrence over various regulatory regions. The distribution of SD-ASM loci around regulatory regions was determined by calculating the proportion of heterozygous loci exhibiting SD-ASM over the sum of loci with SD-ASM and control loci (methylation difference between alleles $\leq$ 5%) in 200bp bins starting from the center of the regulatory element. Promoters with CpG islands were defined as the regions surrounding the transcription start site of protein-coding genes (Gencode v19 annotation) that were defined as promoters in Broad Institute's Reg2Map HoneyBadger2 annotation and were within 5Kb of CpG islands. These regions were detected by DNase I data from 53 epigenomes and were consistently classified as a promoter by ChromHMM across 127 epigenomes using the 15 state model. Enhancers were defined based on the aforementioned dataset obtained from the Broad Institute's Reg2Map HoneyBadger2 annotation. These regions were detected by DNase I data from 53 epigenomes which were consistently classified as enhancers by ChromHMM across 127 epigenomes both with the 15 state and 25 state models, yielding a total of 474,004 enhancer regions (https://personal.broadinstitute.org/meuleman/reg2map/HoneyBadger2_release/).

Calculation of sequence-dependent epiallele frequencies, Shannon entropy, and Coefficient of Constraint

The set of loci retained for the cross-methylome analyses were further filtered to retain only those loci that exhibited at least 4 homozygous CpGs across all individuals in the 200 bp window (100 bp each side) surrounding the variants of interest. The loci displaying SD-ASM in the combined set, defined as falling below threshold of 10% FDR and a exhibiting a minimum of 30% difference in methylation, were chosen for analysis. A control set for comparison was generated by taking the closest variants without SD-ASM in the combined set, defined as not having significant difference in methylation between alleles and a maximum of 5% difference in methylation between alleles. For the chosen loci, the patterns of methylation status of the 4 CpGs nearest to the variant were compiled for all reads across all samples. Only reads that contained both the complete observations of the methylation status of all 4 CpGs and the heterozygous SNP were retained. In order to compare the epiallele frequency spectra between alleles at the loci of interest, another filter was applied to select only those loci that had at least 10 complete reads (containing methylation status of all 4 CpGs) for each allele when combining across samples. In the following, X denotes allele random variable that takes values a_1 or a_2, the two alleles at the heterozygous SNP locus.

Overall Shannon entropy H(Y) was calculated based on the epiallele frequency profile $p(y_i|X)$, for each of i= 1,...,16 ($16 = 2^4$) possible epialleles, y_i over 4 CpG. Frequencies $p(y_i|X)$ were calculated by dividing the number of reads with an allele and a particular epiallele by the total number of reads with the same allele meeting the aforementioned criteria. The overall entropy (in bits) was calculated under assumption that a_1 and a_2 are equiprobable from a simple average of the two profiles as follows:

$$H(Y) = -\sum_{i=1}^{16} 0.5(p(y_i \mid X = a_1) + p(y_i \mid X = a_2)) \log_2(0.5(p(y_i \mid X = a_1) + p(y_i \mid X = a_2)))$$

Entropy conditional on equiprobable allele X was calculated by the following formula:

$$H(Y \mid X) = -\sum_{i=1}^{16} 0.5p(y_i \mid X = a_1) \log_2(p(y_i \mid X = a_1)) + 0.5p(y_i \mid X = a_2) \log_2(p(y_i \mid X = a_2))$$

and mutual information I(X;Y) was calculated as: $I(X;Y) = H(Y) - H(Y|X)$.

An epiallele was defined to be "frequent" at a locus of interest if it belonged to the minimal-size set of epialleles with combined frequency of 60%. The minimal-size set was computed by sorting the 16 epialleles from most frequent to least frequent and adding epialleles starting from the most frequent into the set until the combined frequency of epialleles in the set reached 60%. For calculation of frequent epialleles, equal weight was given to each allele for the observed epialleles at heterozygous loci.

For those set of loci exhibiting SD-ASM and containing exactly 2 frequent epialleles, which corresponded to either all methylated or all unmethylated CpGs, we measured the degree genetic variation determines ("constrains") epigenetic variation using Coefficient of Constraint C_{XY} that was calculated as follows: $C_{XY} = I(X;Y)/H(Y)$.

Validation of stochasticity of epialleles at SD-ASM loci using ENCODE WGBS dataset

Alignment files for WGBS datasets of smooth muscle cells (ENCSR076OHG) were downloaded from the ENCODE project website for both isogenic replicates (https://www.encodeproject.org). Duplicate removal, sorting, and indexing of BAM files was performed using Samtools (*68*). Alignments with MAPQ smaller than 20 were discarded from downstream use. Reads from both processed BAM files were then combined for quantification of epiallelic patterns. The same 200bp windows containing a minimum of 4 CpGs examined in the combined dataset, were examined in this dataset, if the allele which the window was centered on was in a homozygous state. Homozygosity in this ENCODE dataset was required to mitigate any confounding due to differential read mapping between two alleles and to show that stochasticity is not related to homozygous/heterozygous status. Coordinates from the combined dataset were converted from hg19 to GRCh38 using UCSC's online liftOver tool (*69*). Those windows with a minimum of 30 reads containing complete observations for all 4 of the closest CpGs to the window center were retained for downstream calculations. 864 of the SD-ASM loci were homozygous and had sufficient coverage in the external ENCODE dataset. A similar number of control loci were chosen genome-wide from 200bp windows meeting the same criteria at random. Shannon entropy and quantification of number of frequent epialleles per locus were calculated in a similar fashion as described above. Tanghulu plots were generated by using CgmapTools (*70*) and utilizing only WGBS reads from replicate one from the aforementioned ENCODE dataset.

Calculating enrichment and directionality of SD-ASM in transcription factor motifs

The R package motifbreakR *(71)* was utilized to determine whether a given polymorphism, in the combined DNA methylation dataset, significantly overlapped a sequence match against a library of human transcription factor motifs previously generated via high-throughput SELEX *(36)*. Each allele was scored for a given motif to allow inference of how much information is affected by a change in sequence due to a polymorphism. The scoring parameters chosen included a threshold of 1e-4, background probabilities of 0.25, and using the log-probabilities scoring algorithm option. To filter for motifs more likely to be bound, tested heterozygous SNPs were restricted to motifs that overlapped DNAse I hypersensitive regions defined based on the complete set of epigenomes from the Roadmap Epigenomics Project, as described above. For each given motif, enrichment was performed by testing whether the proportion of heterozygous SNPs with SD-ASM was significantly higher within the motif relative to 500bp windows flanking the motif and using heterozygous SNPs without SD-ASM (methylation difference less than 5%) as controls by Fisher's exact test. Directionality was assessed by testing whether, for a given motif, the allele with higher methylation was significantly associated with the higher scoring allele for heterozygous SNPs falling within the motif using a binomial test. A motif was required to contain at least 10 heterozygous SNPs falling within in order for directionality to be assayed. Only heterozygous SNPs passing a threshold of difference in scores greater than 1.5 between alleles were retained for directionality analyses. All p-values had multiple hypothesis adjustment using the Benjamini-Hochberg procedure, for the number of transcription factor motifs assayed.

Testing relation of allelic looping and CTCF binding with allelic imbalances in DNA methylation

A list of previously identified heterozygous SNPs, exhibiting both allelic looping and CTCF binding in the GM12878 cell line, were referenced and those that were present in our combined DNA methylation dataset were retained for further analyses *(35)*. Of the 48 heterozygous SNPs that were common between the datasets, 44 fell within a DNAse I hypersensitive region as defined based on the complete set of epigenomes from the Roadmap Epigenomics Project. To assay whether the presence SD-ASM at an heterozygous SNPs position falling within an CTCF motif was a better predictor of allelic looping than difference in scores, based on the log-probabilities scoring algorithm, a threshold of "disruptiveness" was set to compare the specificity of allelic imbalances in methylation compared to score differences at a similar sensitivity. Utilizing the motifbreakR_motif library of motifs in the motifbreakR package *(71)*, heterozygous SNPs falling within CTCF motifs present in accessible regions (within the aforementioned DNAse hypersensitive region) within the combined DNA methylation dataset were identified. If a heterozygous SNP fell within several different CTCF motifs, the score for each allele was averaged across all possible CTCF motif scores for that allele. A similar sensitivity was achieved by setting a threshold for score differences, where a similar number of heterozygous SNPs were deemed to have a "disruptive" difference between alleles as heterozygous SNPs showing SD-ASM. CTCF motif-containing loci genome-wide were ranked by both motif disruption scores, as well as by allelic methylation imbalance scores, and the two top-lists were examined for the loci that showed allelic looping by ChIA-PET. The specificity of each metric to predict allelic

looping was then determined by comparing how many heterozygous SNPs had a "disruptive" difference between alleles, based on the calculated threshold, to how many exhibited SD-ASM for the 44 overlapping heterozygous SNPs that previously exhibited allelic looping.

Selection of genomic regions with allelic imbalance DNA variations and cis-overlapping motifs for P53 and cMYC for luciferase assays

Regions associated with SD-ASM in the combined dataset were selected for functional analysis to determine the effect of the associated DNA sequence variation on target gene expression. A threshold of methylation difference of more than 20% and FDR < 0.1 were chosen as cutoffs to select variants with SD-ASM. 10 heterozygous SNP loci were selected on the basis of the following criteria: (1) $\geq$ 20% SD-ASM; and (2) allelic differences in predicted transcription factor binding affinities; and one or more of the following: (3a) being in strong ($r^2 \geq 0.9$) linkage disequilibrium with GWAS SNPs; (3b) falling near eQTLs whose associated genes showed ASE; or (3c) falling within CisOMs (**table S5**). Those loci near GTEx *(72)* eQTLs were further restricted to sites defined as active enhancers from the aforementioned Broad Institute's Reg2Map HoneyBadger2 annotation. All selected loci were predicted to harbor drastic differences in predicted transcription factor binding affinities between alleles based on the R package motifBreakR and further visual inspection in the Jaspar database *(73)*. DNA conservation and epigenetic markers for these regions were considered for prioritization purposes. Of these 10 loci, four constructs were successfully generated and tested.

Allelic enhancer activity measurement using luciferase assay

HEK293 cells were used for testing genomic regions for enhancer activity using luciferase assay. Cells were grown in DMEM, 10% FBS without antibiotics medium at 37°C as previously described *(44)*. After trypsin digestion, the cells were grown for 2 hours then transfected using lipofectamine 3000 (Life Technology, CA) with pGL3-basic-Luc and pGL3-enh-Luc as negative and positive controls, respectively. The pGL3-SV40-Renilla plasmid served as an internal control for normalization and transfection efficiency. Site directed mutagenesis was used to introduce the identified DNA variations (SNPs) within regulatory regions that show enhancer activity in cell-based assay as previously described *(74)*. Allelic differences in luciferase expression with 5 replicates for each allele.

Testing enrichment of heterozygous loci with ASM associated with ASE

Tissue types from the DNA methylation datasets were matched with the corresponding RNA-seq dataset from the same sample to determine if there was an enrichment of ASM in promoters of genes that exhibited ASE. Forty-three samples had both RNA-seq and DNA methylation datasets available. In order to call ASM, a difference in methylation of more than 30% between alleles at an FDR of less than 0.10 was used. For ASE calling in the RNA-seq dataset, a difference in reads aligned of more than 70% between alleles and a FDR less than 0.10 was used. ASM was screened in protein coding promoters matched with their downstream gene from the RNA-seq dataset. The promoters were defined based on their overlap with the transcription start sites (500bp upstream and 1000bp downstream). Enrichment was obtained by testing

whether heterozygous variants with ASM in promoters, whose associated genes exhibited ASE, were significantly more enriched than heterozygous variants without ASM by Chi-square analysis. Similarly, enrichment was performed for aforementioned enhancers regions defined from the Broad Institute's Reg2Map HoneyBadger2 annotation.

Testing enrichment of ASM in proximity to previously published eQTLs

Relevant tissue-specific eQTLs from the GTEx *(72)* dataset were matched with similar tissue types to our DNA methylation datasets to determine if there was an enrichment of heterozygous variants with ASM in proximity to eQTLs whose associated gene exhibited ASE in our RNA-seq dataset from the same sample as well. Thirty-four samples were present with matching tissues. A window of 1000 base pairs surrounding an eQTL was set to test nearby heterozygous variants in the methylation dataset. Enrichment was determined by testing whether the number of heterozygous variants with ASM was significantly higher than heterozygous variants without ASM near eQTLs whose associated gene had ASE via Chi-square analysis.

Testing enrichment of allelic imbalances in promoters marked by H3K4me3 or H3K27ac and enhancers marked by H3K27ac

ChIP-seq for histone marks and RNA-seq datasets were utilized from the same sample, of which 40 samples had both H3K27ac and RNA-seq data generated and 35 samples that had both H3K4me3 and RNA-seq datasets. The promoters were defined based on their overlap with the transcription start sites as previously mentioned. Enhancers were defined based on the aforementioned dataset obtained from the Broad Institute's Reg2Map HoneyBadger2 annotation. Only enhancers at a distance of less than 100,000 base pairs from the gene were used. Histone allelic imbalance was defined as a difference between alleles of more than 70% of aligned reads and an FDR of less than 0.10. Enrichment of histone allelic imbalances in enhancers and promoters was determined by testing whether the number of heterozygous variants with histone allelic imbalance in promoters or enhancers, whose genes showed ASE, was significantly higher than for the heterozygous variants with no histone allelic imbalance using a Chi-square test.

Testing enrichment of SD-ASM near reported GWAS loci

The utilized GWAS dataset was obtained from both the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI) websites. The data was divided between the SNPs that have been replicated in a second study and those SNPs that haven't been replicated. A window of 1000 base pairs was set surrounding GWAS SNPs to screen for nearby heterozygous variants exhibiting SD-ASM. For these analyses, the combined methylation dataset was used. Enrichment near GWAS SNPs was tested by comparing whether the number of heterozygous variants with SD-ASM was significantly higher than the number of variants without SD-ASM in proximity to a GWAS SNP using Chi-square analysis. These analyses were done genome-wide. Similar analyses were performed focusing on GWAS SNPs and variants with SD-ASM falling solely in enhancer elements. An enhancer was defined similarly as mentioned above. Chi-square analysis was performed to test whether the number of

variants with SD-ASM was significantly higher than variants without SD-ASM near GWAS SNPs falling in enhancer regions.

Analysis of derived allele frequency of variants with ASM

For each of the 46 individual samples with WGBS and adequate coverage, estimates of DAF for variants exhibiting ASM were obtained from the 1000 genomes project *(48)*. Variants overlapping regions with low accessibility to variant calling were excluded. For each methylome in our dataset, we used a Chi-squared test to determine whether ASM was significantly enriched for heterozygous SNPs with a low DAF (< 1%) compared to those without ASM (methylation difference between alleles < 5%). This test was performed both genome wide and in aforementioned enhancer regions. A similar test was performed on the combined set to test enrichment of low DAF (< 1%) with SD-ASM. To minimize any locus-dependent confounding, we chose variants without SD-ASM (methylation differences between alleles < 5%), occurring within 1kb of a variant with SD-ASM, as controls. To minimize confounding due to ascertainment biases, we tracked the number of individuals where each ASM and control "index hets" could be ascertained and matched the distributions of ASM and controls.

Allelic Epigenome Atlas

The Allelic Epigenome Atlas was developed using the GenboreeKB document modeling system and is accessible at https://genboree.org/genboreeKB/projects/allelic-epigenome. The annotations are downloadable in JSON document format and include allelic epigenome states and additional metadata. The JSON document collections are divided by chromosome. JSON documents provide information about a particular allele both at the level of individual samples and for the combined dataset. Allelic imbalance p-values are provided both before and after correction for multiple hypothesis testing. If the allele is associated with a given transcription factor motif, the name of the transcription factor is included, along with the associated motif scores for both the reference and alternate alleles. Proximity to known GWAS variants, promoter, and/or enhancer regions is flagged. Lastly, all variants have been registered within the ClinGen Allele Registry (https://reg.clinicalgenome.org/) and assigned canonical allele IDs. Additional metadata describes experiments performed and the tissues/cell types analyzed. Many descriptors in the metadata document collections for Experiment and Tissue use ontology terms from the Experimental Factor Ontology (EFO) *(75)*, the Sequence Ontology (SO) *(76)*, and/or the UBERON ontology *(77)*.

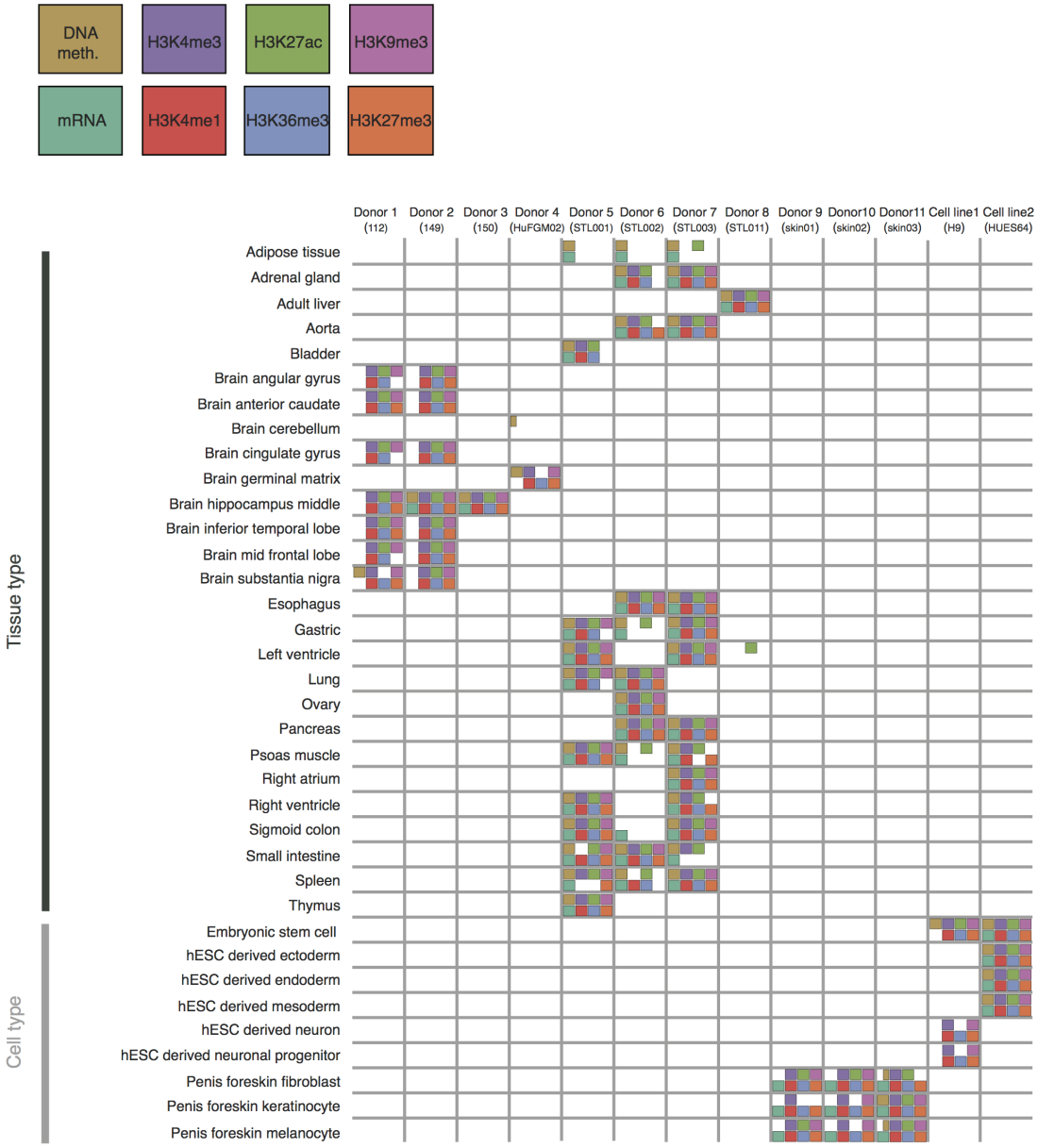

Fig. S1: Description of utilized datasets for detection of allelic imbalances. Each colored box represents which of the 8 assays were available for each of the 71 samples. For DNA methylation, a half mark indicates samples that were excluded from individual analyses due to low coverage, but were included in the combined dataset. Noted in parentheses for each donor is the donor ID code used in the Roadmap Epigenomics project.

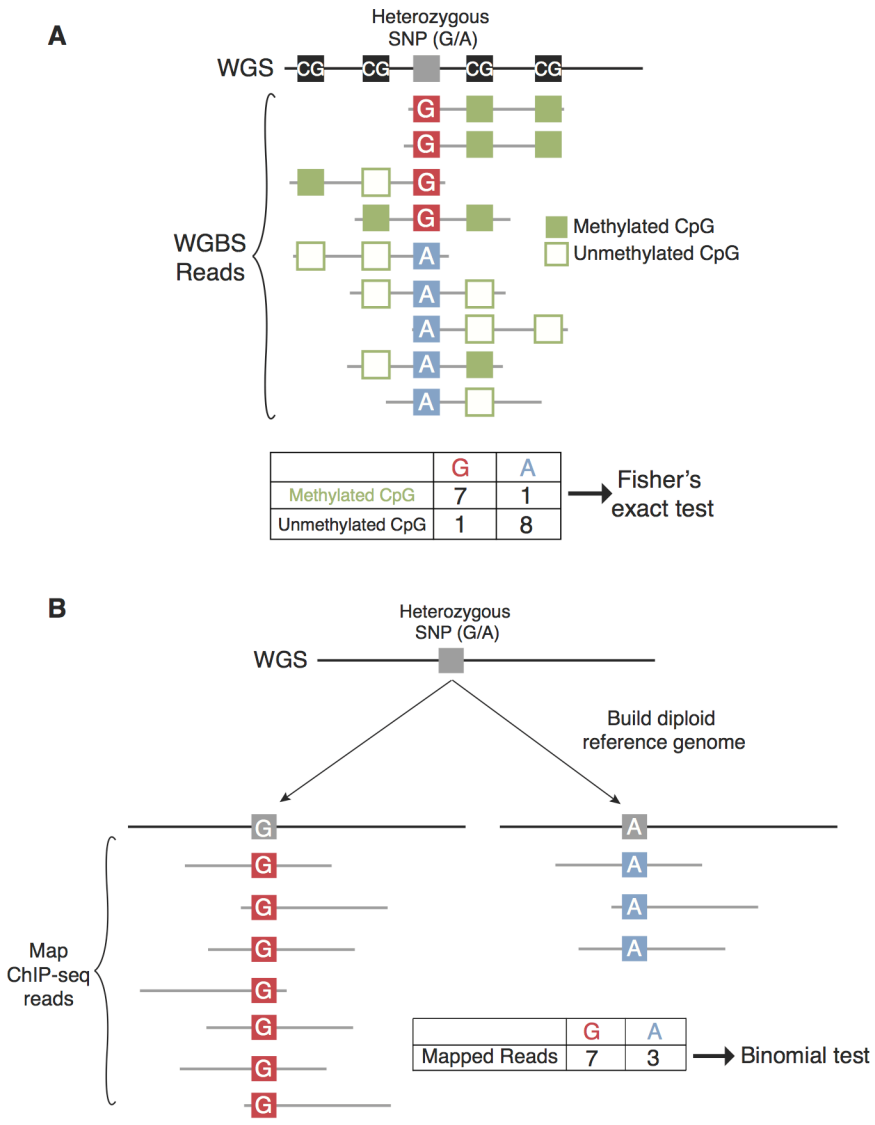

Fig. S2: Calling allelic imbalance. (**A**) Detection of allele-specific methylation. WGS variant calls are used both to determine the positions of heterozygous SNPs and to identify all homozygous CpG positions in the genome of each individual. For every heterozygous SNP position, WGBS is used data to count the number of times a methylated or unmethylated homozygous CpG occurs in the same read as each of the two possible alleles in the SNP position. The significance of the association between the allele at the heterozygous SNP position and the methylation state of the CpGs in the surrounding region is assessed using Fisher's exact test. (**B**) Detection of allelic imbalances in ChIP-seq and RNA-seq datasets. Personal diploid genomes are built based on WGS variant calls for each individual. Each read, from ChIP-seq or RNA-seq datasets, is aligned to both versions of the reference genome, and the best scoring mapping is kept. For each heterozygous SNP an even binomial test is used to determine whether there is a significant difference between alleles in the number of ChIP-seq or RNA-seq reads mapping to that position.

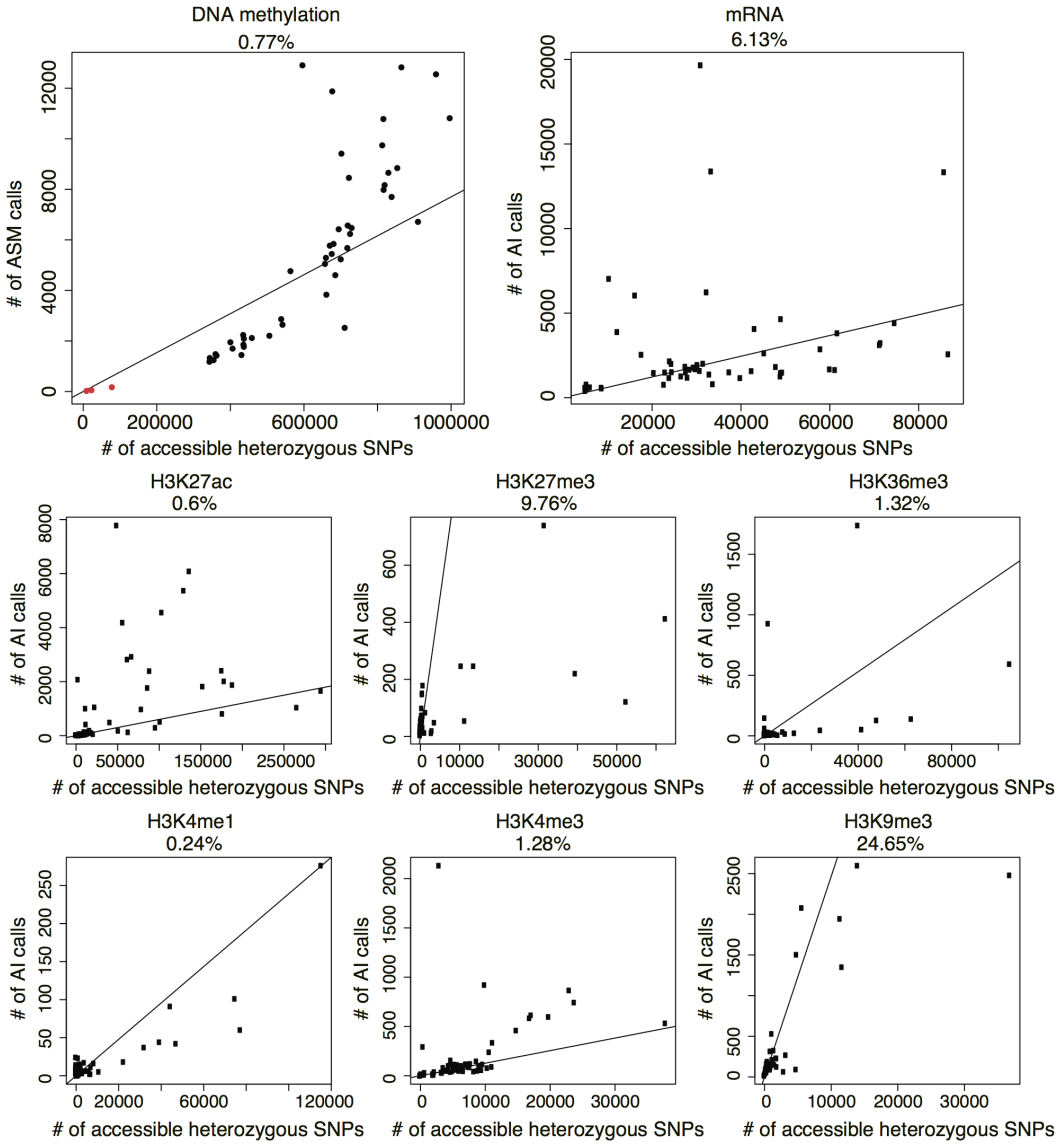

Fig. S3: Extent of loci showing allelic imbalance per sample. Each dot in the plots displayed in this figure represents one sample in the dataset. The x-axis represents the number of heterozygous loci that were accessible to allelic imbalance calling in each sample. The y-axis represents the number of allelic imbalance calls made in each sample. The slope of the line in each plot, as well as the number above each plot, represent the median percentage of accessible loci where allelic imbalances were found. For DNA methylation the three samples with insufficient coverage are noted in red.

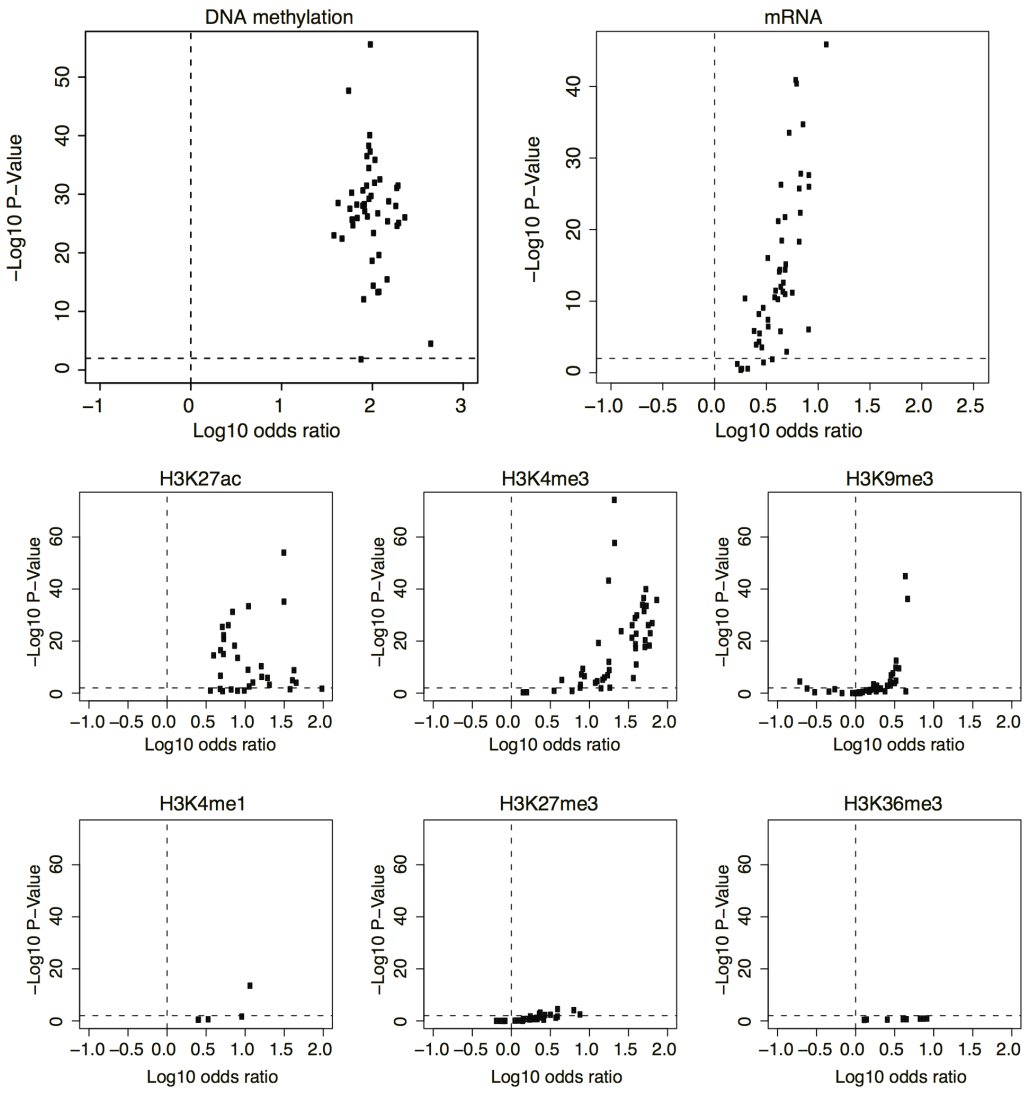

Fig. S4: Enrichment of allelic imbalances in imprinted regions. Each dot in the plots represents one sample in the dataset. For each of those samples the number of AI loci were compared to other heterozygous loci within and around imprinted regions of the genome using a Chi-square test. For DNA methylation, the enrichment was tested over known imprinting control regions. For all other marks, the enrichment was tested over imprinted genes with 10kb windows added around them.

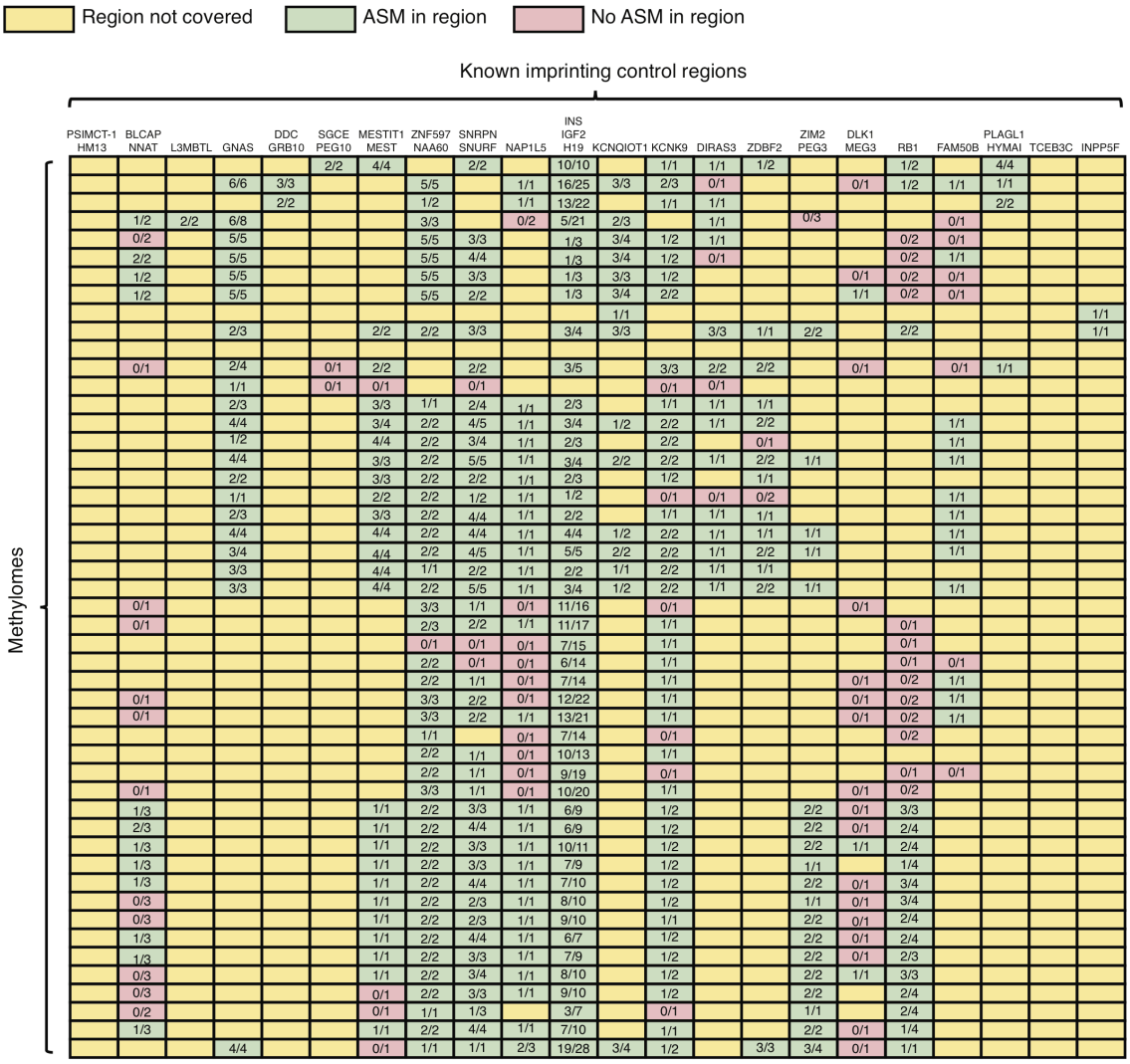

Fig. S5: ASM calls per imprinting control region across 49 methylomes. Shown in each cell is the number of ASM calls made out of the total number of accessible loci that overlapped an imprinting control region (ICR) for each methylome. Green cells represent ICRs overlapping at least one ASM call in that methylome. Red cells represent ICRs that overlap at least one accessible heterozygous SNP, but no ASM is observed. Yellow cells represent ICRs that did not overlap any accessible heterozygous SNPs in the methylome. A total of 1,259 distinct ascertained heterozygous SNPs overlapped a known imprinting control region in at least one of the methylomes. Of those, 892 (70.8%) displayed ASM.

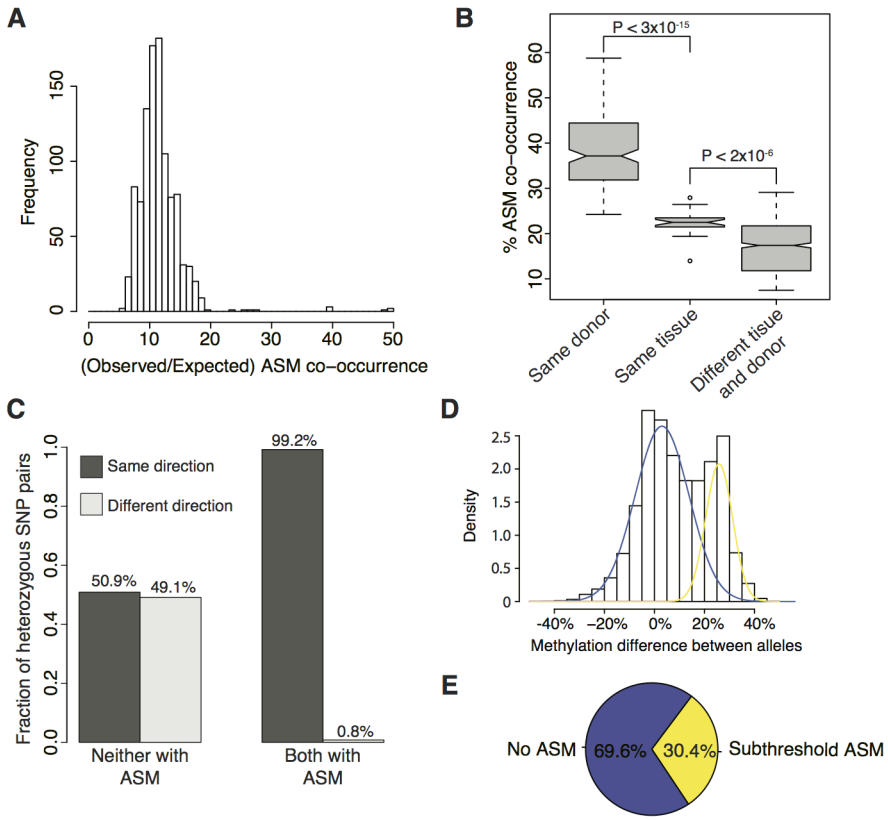

Fig. S6: ASM co-occurs across samples towards the same allele. (A) Histogram of observed versus expected number of ASM co-occurrences across all pairs of methylomes. (**B**) Boxplots of co-occurrence of ASM across all pairs of samples for pairs of samples from different tissues of the same individual, from pairs of the same tissue but different individuals, and from pairs of samples without matching tissue or individual of origin. Significance calculated using two-sided t-test. (**C**) Consistency of directionality of difference in methylation between alleles either in loci where both samples in a pair display ASM or in loci where neither sample displays ASM. (**D**) Distribution of methylation differences between alleles for heterozygous SNPs around which ASM was observed in one of the methylomes but not in the other in a pair. Plotted is the distribution of methylation differences between alleles for the sample that did not display ASM, relative to the direction of methylation difference in the other sample. Overlaid is the Gaussian Mixture Model fitting showing the distribution as a mixture of two normal distributions (**E**) The posterior probability of a heterozygous SNP having subthreshold ASM, or not having ASM, as determined by fitting a Gaussian Mixture Model (from subpanel **D**).

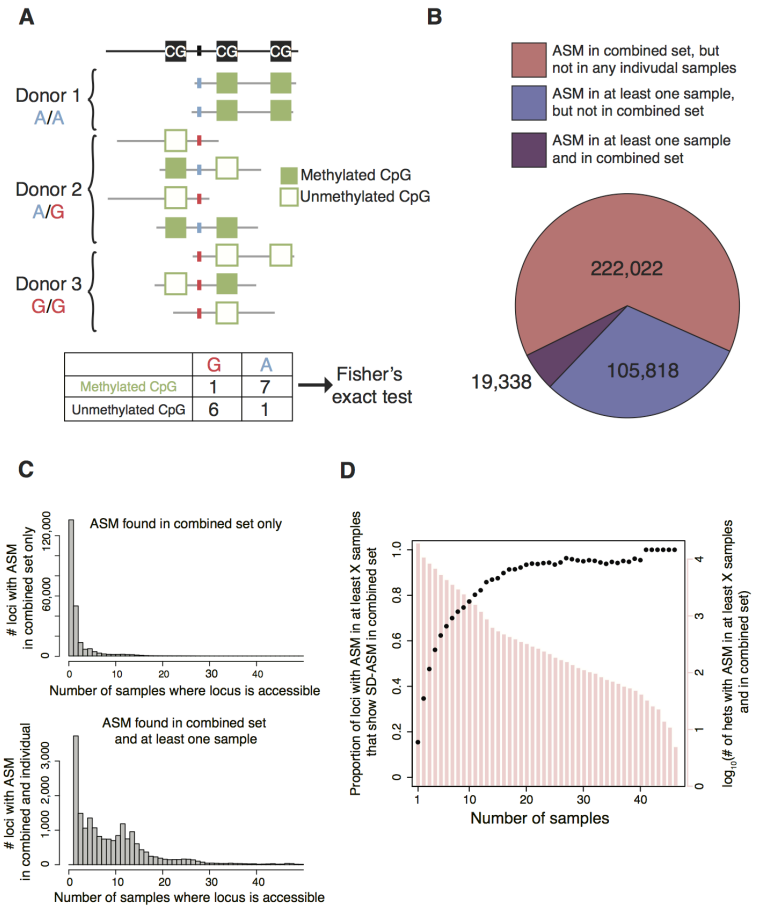

Fig. S7: Pooling methylomes increases power to detect SD-ASM. (**A**) Calling SD-ASM by combining observations from all methylomes. For every SNP position with a total of two alleles across all samples, we use WGBS data to count the number of times a methylated or unmethylated homozygous CpG occurred in the same read as each of the two possible alleles in the SNP position. (**B**) Number of loci with SD-ASM in the combined set, in at least one individual sample, or in both. Of all the "index hets" accessible in the combined set only 84% (4,129,430) were accessible (by the same criterion of at least 6 counts per allele) in at least one individual sample. Of all the loci accessible in the combined set, about 5% (241,360 out of 4,913,361) displayed SD-ASM at an FDR ≤ 0.1 and with methylation difference between alleles of at least 30%, while only 125,156 displayed SD-ASM in at least one individual sample under the same detection thresholds. The overlap between the loci with SD-ASM called in the combined and individual sets was 19,338. (**C**) Increase in detection power is responsible for increase in SD-ASM called in combined set (top), but not in single samples (bottom). Specifically, 57% of the loci where SD-ASM was detected in the combined set, but not in individual samples, were not accessible for detection (having < 6 counts per allele) in any individual sample. (**D**) Proportion of variants with ASM in at least X samples that also displayed SD-ASM in the combined set of imbalances. Notice that if a locus displayed ASM in multiple samples, it was more likely to also display SD-ASM in the combined set.

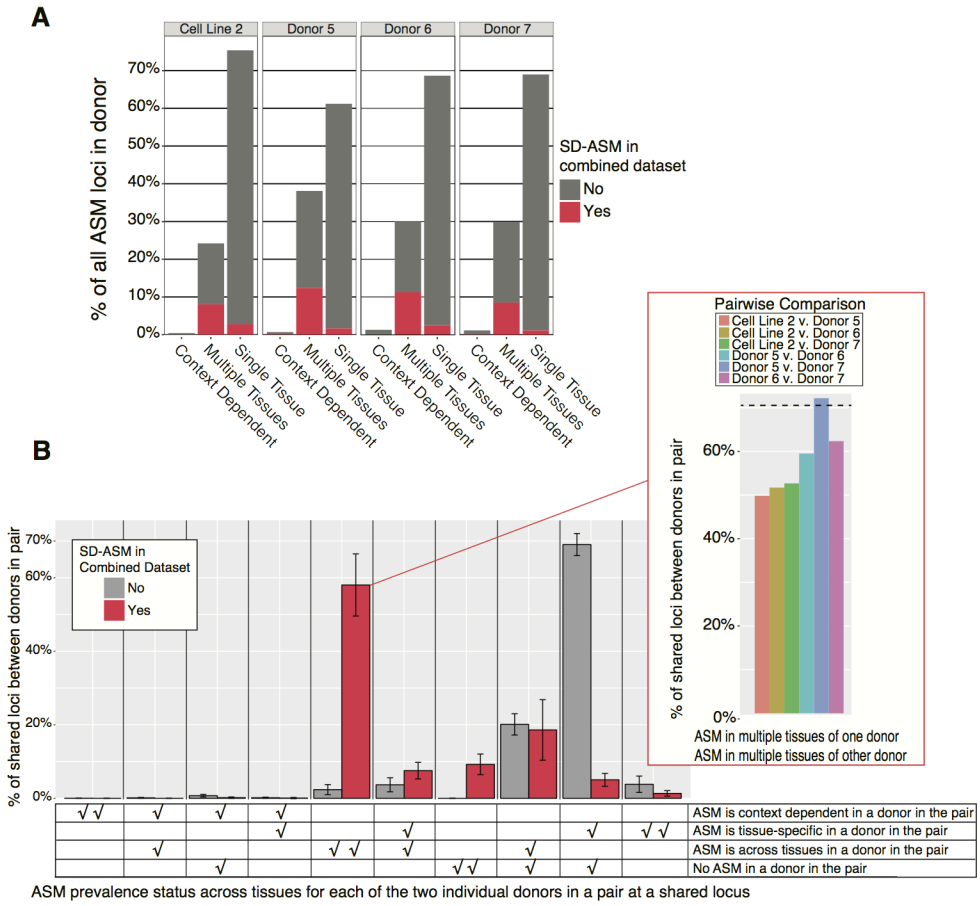

Fig. S8: Loci showing ASM in multiple tissues across donors are over-represented within the list of loci showing SD-ASM in the combined dataset. (**A**) The amount of ASM loci per donor, for 4 donors (3 individuals and 1 cell line), where total ASM per donor is defined as all ASM loci that show ASM in at least 1 tissue of that donor. The ASM signal was defined to be context-dependent if ASM was observed in multiple tissues, but directionality of ASM across these tissues was inconsistent. (**B**) Pairwise comparisons of ASM prevalence across tissues between two donors at shared heterozygous loci. Average values across all possible pairwise comparisons of donors from (A) of the percent of shared loci exhibiting possible combinations of ASM prevalence across tissues for each donor in a pair. The callout shows an example of the six possible pairwise donor comparisons comprising an average percentage. Shared loci profiled in pairwise comparisons of donors were split into two groups based on their SD-ASM status in the combined dataset: those loci that showed SD-ASM in the combined dataset (red) or those that did not show SD-ASM in the combined dataset, but showed ASM in at least one tissue of an individual donor (gray). The two check marks under each group of bars denote the possible status of ASM prevalence across the tissues of a donor, at a shared locus, for each donor in the pair.

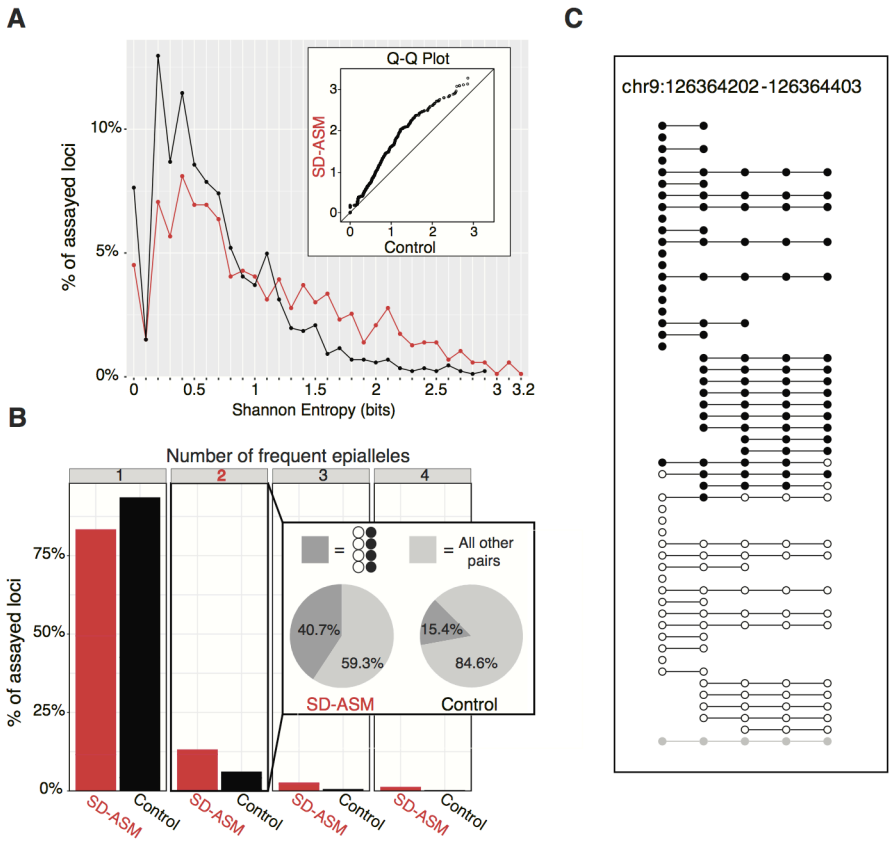

Fig. S9: SD-ASM loci display higher frequency of multiple epialleles compared to genome-wide control loci in an external dataset. (**A**) A larger percentage of SD-ASM loci (red), which were examined for their epiallele frequency spectra in the combined dataset, show a shift to higher Shannon entropy (in bits) compared to randomly chosen control loci (black) when checked in an external WGBS dataset. A Quantile-Quantile plot is included showing entropy distribution between control (x-axis) and SD-ASM loci (y-axis). (**B**) Histogram of the number of frequent epialleles for those loci with SD-ASM (red) in our combined dataset or control loci (black). The callout indicates that among the SD-ASM loci with two frequent epialleles, more than twice the amount show predominately the completely methylated and unmethylated epialleles compared to controls in this external dataset. (**C**) Tanghulu plot of a bistable SD-ASM locus. Individual CpGs, represented as circles on each WGBS read, are either methylated (black) or unmethylated (white).

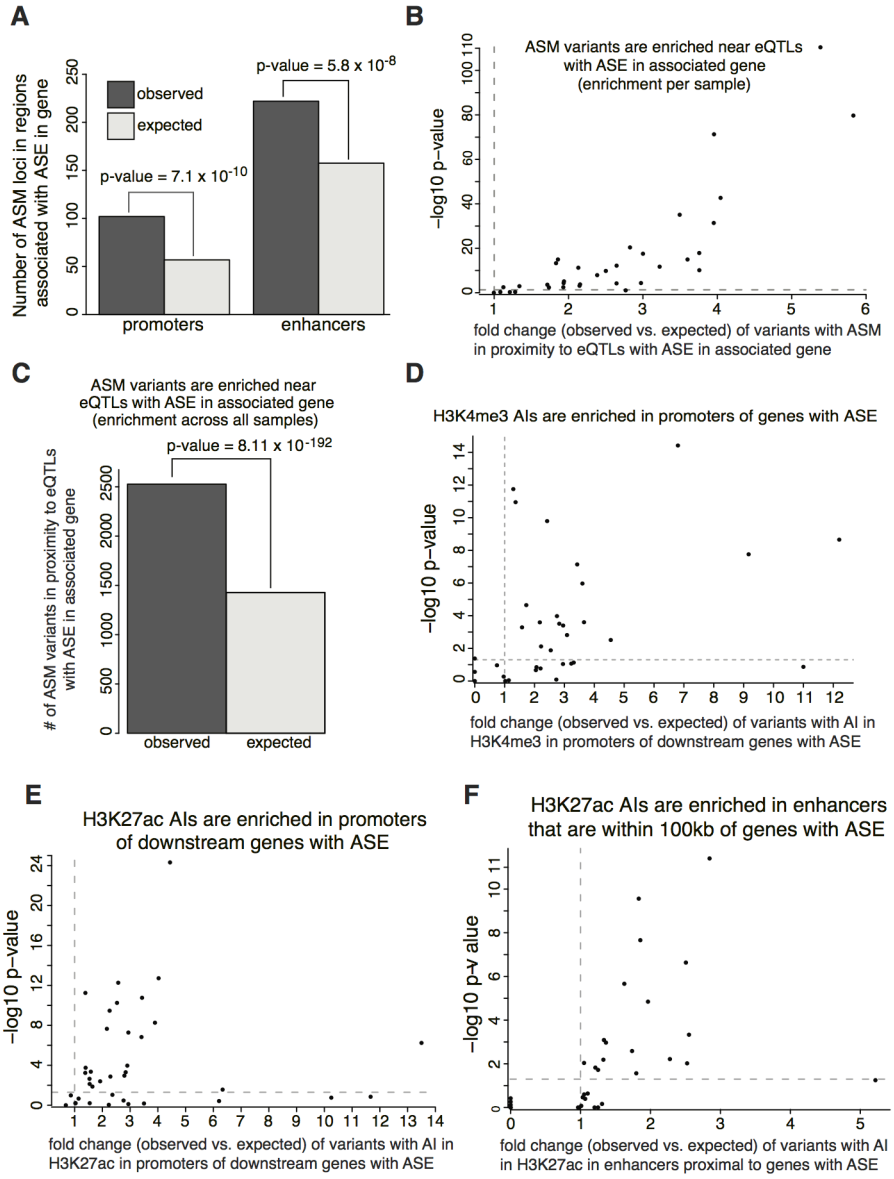

Fig. S10: Allelic imbalances in regulatory regions are associated with ASE of nearby genes. (A) Comparison of ASM to non-ASM heterozygous variants in promoters, and enhancers within 100kb, of genes with ASE. All occurrences of ASE and ASM across all matched pairs of RNA-seq and methylomes in our dataset were utilized. (**B**) Comparison of ASM in the neighborhood of tissue specific eQTLs (from GTEx) displaying ASE of associated genes compared to nearby eQTLs without such ASE events. Enrichments were performed for each matched pair of methylome and RNA-seq experiments in our dataset with a matched tissue type in GTEx. (**C**) ASM events are more prevalent in the neighborhood of tissue specific eQTLs (GTEx), displaying ASE of associated genes, than in nearby eQTLs without such ASE events after combining across all pairs of methylome and RNA-seq experiments in our dataset with a matched tissue in GTEx. (**D-F**) Promoters upstream of genes displaying ASE are enriched for AI in H3K4me3 (**D**) and H3K27ac (**E**) and H3K27ac AIs are enriched in enhancers within 100 kb of genes with ASE (**F**) compared to non-AI controls. All enrichments tested using Chi-square test.

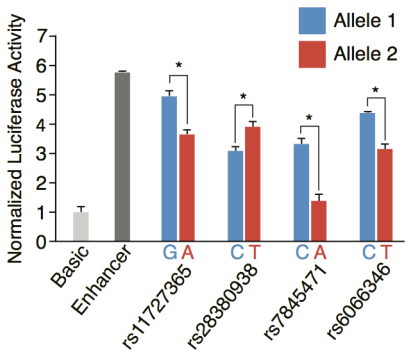

Fig. S11: Differential enhancer activity between alleles at variants with SD-ASM at putative transcription factor binding sites. Comparison of enhancer activity on luciferase expression between alleles (marked blue for the reference allele and red for the alternate allele) at 4 variants with SD-ASM with allelic differences in predicted transcription factor binding. All four SD-ASM variants show significant (p-value < 0.05, one-way analysis of variance) differences in downstream luciferase expression as noted by an asterisk. Nucleotides for both alleles are noted for each of the 4 SD-ASM variants along with their associated rsID. Positive (dark gray) and negative (light gray) controls are included for comparison.

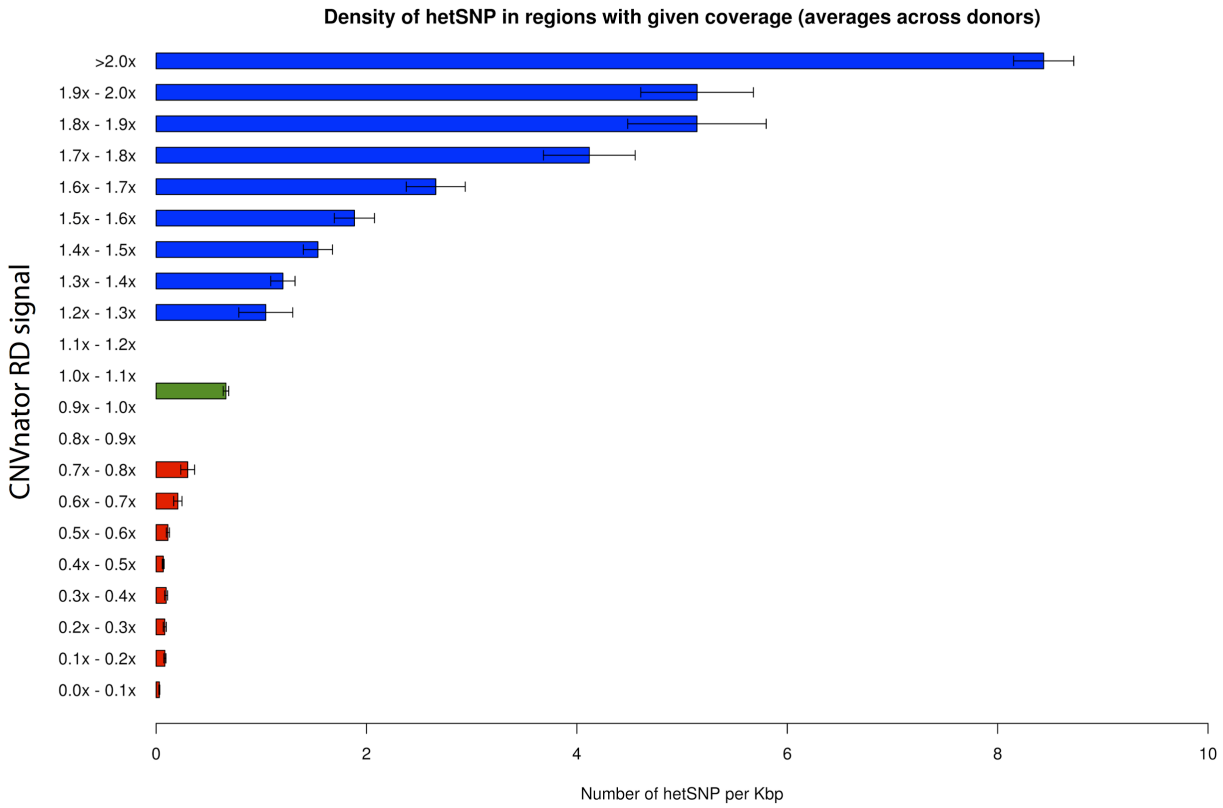

Fig. S12: Density of heterozygous SNPs in regions returned by CNVnator depending on detected Read Depth (RD value). Each bar corresponds to the average number of heterozygous SNPs per 1 kilobase (Kb) window detected in regions with RD values within a given range. Regions called by CNVnator as deletions are marked by the red color, while bars corresponding to regions called as duplications are marked in blue. The green bar in the middle of the histogram corresponds to the average number of heterozygous SNPs per 1 Kb calculated for regions omitted by CNVnator (regions with standard coverage). For each donor these values were calculated separately. Bars presented on the histogram are averages with standard deviations calculated across donors.

Table S1: Statistics on number of allelic imbalance calls.

	25th Percentile	Median	75th Percentile	AI in at least one sample
DNA methylation	1,668	4,600	7,237	125,156
mRNA	592	912	1,666	76,729
H3K27ac	5	28	776	36,481
H3K27me3	0	3	7	521
H3K36me3	0	2	5	2,000
H3K4me1	0	1	3	133
H3K4me3	8	25	52	4,535
H3K9me3	10	31	60	1,273

Table S2: Sequencing read depth coverage metrics for 49 individual and the combined WGBS datasets. Samples highlighted in red did not have sufficient coverage for individual analyses, but were included in the combined dataset.

WGBS Dataset	*25th Percentile*	*50th Percentile*	*75th Percentile*	*Mean*
Combined	1409	1870	2180	1691.3
Donor 1: Brain Substantia Nigra	3	13	29	18
Donor 2: Brain Hippocampus Middle	5	18	34	21.3
Donor 3: Brain Hippocampus Middle	6	21	40	25.5
Donor 4: Brain Cerebellum	2	4	6	4
Donor 4: Brain Germinal Matrix	1	7	25	16.9
Donor 5: Adipose Tissue	13	21	29	20.7
Donor 5: Bladder	35	54	70	50.9
Donor 5: Gastric	16	25	33	23.8
Donor 5: Left Ventricle	36	55	71	51.4
Donor 5:Lung	14	22	30	21.7
Donor 5: Psoas Muscle	13	22	30	21.4
Donor 5: Right Ventricle	17	26	34	24.4
Donor 5: Sigmoid Colon	35	49	61	46.9
Donor 5: Small Intestine	35	49	61	46.7
Donor 5: Spleen	16	26	34	24.7
Donor 5: Thymus	33	50	64	46.8
Donor 6: Adipose Tissue	16	25	32	23.5
Donor 6: Adrenal Gland	17	26	34	24.7
Donor 6: Aorta	14	22	29	20.9
Donor 6: Esophagus	15	24	32	23.1
Donor 6: Gastric	17	27	35	25.3
Donor 6: Lung	34	48	59	44.1
Donor 6: Ovary	38	54	67	49.7
Donor 6: Pancreas	14	23	30	21.7
Donor 6: Psoas Muscle	14	23	30	21.7
Donor 6: Small Intestine	14	21	28	20.2
Donor 6: Spleen	17	27	35	25.2
Donor 7: Adipose Tissue	31	50	66	47.3
Donor 7: Adrenal Gland	29	46	62	44.6
Donor 7: Aorta	56	81	100	74.9
Donor 7: Esophagus	30	43	54	40.4
Donor 7: Gastric	34	52	67	48.8
Donor 7: Left Ventricle	31	46	57	42.8
Donor 7: Pancreas	32	47	59	43.7
Donor 7: Psoas Muscle	35	52	66	48.4
Donor 7: Right Atrium	34	52	67	48.5

Donor 7: Right Ventricle	35	54	70	50.8
Donor 7: Sigmoid Colon	37	51	61	46.3
Donor 7: Small Intestine	14	23	32	22.6
Donor 7: Spleen	31	45	57	43
Donor 8: Adult Liver	38	55	68	50.6
Donor 11: Penis Foreskin Fibroblast	4	8	14	11
Donor 11: Penis Foreskin Keratinocyte	17	29	41	28.6
Donor 11: Penis Foreskin Melanocyte	6	11	17	12.3
Cell Line 1: Embryonic Stem Cell	15	28	39	27.6
Cell Line 2: hESC Derived CD184+ Endoderm	24	55	84	55.4
Cell Line 2: hESC Derived CD56+ Ectoderm	15	49	95	59.5
Cell Line 2: hESC Derived CD56+ Mesoderm	11	33	62	39
Cell Line 2: Embryonic Stem Cell	7	29	65	39.8

Table S4: An association between TFs with enrichment of SD-ASM within their binding motifs that that did not show directionality, while another group of TFs showed neither SD-ASM enrichment within their binding motifs nor directionality.

SD-ASM Enrichment					No SD-ASM Enrichment				
AR	ETS1	HSF2	NR3C1	TFAP2A	ALX3	EVX2	HOXD8	PAX3	SHOX2
ATF7	ETV3	HSF4	NR3C2	TFAP2B	ALX4	FOXK1	HSF1	PAX4	SMAD3
BATF3	ETV6	ID4	NRL	TFAP2C	ARX	GRHL1	ISX	PAX5	TEF
BCL6B	EVX1	IRF3	PAX6	TFE3	BHLHE22	GSC	LEF1	PAX7	TFCP2
CPEB1	FEV	IRF8	PRDM1	THRA	BHLHE23	GSC2	LHX2	PAX9	UNCX
CREB3L1	FLI1	IRF9	PROX1	THRB	CART1	GSX1	LMX1A	PHOX2A	VAX1
DBP	FOXI1	KLF13	RARG	USF1	CUX1	GSX2	MEOX1	PHOX2B	VAX2
DPRX	GABPA	KLF14	RORA	VDR	CUX2	HLF	MIXL1	PITX1	VSX1
EBF1	GBX1	LBX2	RUNX2	ZBTB49	DLX1	HMX1	NEUROD2	PITX3	VSX2
EHF	GCM1	LHX6	RUNX3	ZBTB7A	DLX2	HMX2	NFKB1	PRDM4	ZNF306
ELF1	GLI2	MESP1	RXRB	ZBTB7B	DLX5	HMX3	NFKB2	PROP1	ZNF784
ELF3	GLIS1	MGA	SCRT1	ZBTB7C	DLX6	HOXA2	NKX3-1	PRRX1	
ELF4	GLIS2	MNT	SNAI2	ZNF282	DMBX1	HOXB2	NKX3-2	PRRX2	
ELK4	GLIS3	MTF1	SPDEF	ZNF524	DRGX	HOXB5	NOTO	RAX	
EN2	HES5	NFATC1	SREBF2	ZNF713	DUXA	HOXC12	OLIG1	RAXL1	
ERF	HES7	NFIL3	TBX19		EMX1	HOXC13	ONECUT2	RHOXF1	
ERG	HEY1	NHLH1	TCF3		EMX2	HOXD12	OTX2	SCRT2	
ESR1	HEY2	NR2C2	TCF4		ESX1	HOXD13	PAX1	SHOX	

Table S5: A group of 10 heterozygous SNP loci with SD-ASM selected for testing allelic differences in luciferase expression. The first column contains the rsID of the 10 selected variants with SD-ASM. Columns 2-6 contain criteria used for selection of variants, with highlighted cells in green showing the criteria that was met. Column 2 details the percent of methylation difference between alleles. Column 3 contains the transcription factor motifs that were deemed to have drastic differences in predicted binding affinity between alleles. Column 4 shows the rsID of the GWAS variant that was in linkage disequilibrium with the SD-ASM locus *(78)*. Column 5 details the tissue and associated gene that exhibited ASE for the eQTL that was nearby the SD-ASM variants. Variants that were successfully tested for effect on luciferase expression are highlighted in yellow.

	Criteria for selection				
rsID	1: >= 20% SD-ASM	2: Allelic difference in TF binding affinity	3a: Linkage disequilibrium (r2 >= 0.9) with GWAS SNP (from Tewhey et al.)	3b: <10kb from GTEx eQTL with ASE in associated gene in matched tissue type	3c: Within CisOM for P53 & cMYC
rs536803360	42.8%	HLTF/cMYC/TWIST1,P53			
rs6066346	55.6%	cMYC/MAX,P53			
rs11727365	26.1%	cMYC,P53			
rs7845471	29.3%	RFX5			*
rs28380938	**	E2F4		MICA (Lung)	
rs3819652	**	MEF2C		PI4KA(Aorta)	
rs658524	29.6%	KLF1	rs2231884		
rs67688344	35.9%	KLF5	rs7206971		
rs7576384	38.9%	ELK1	rs11123170		
rs2259928	95.6%	RUNX2	rs7267979		

* Within overlapping ChIP-seq peaks of P53 and cMYC, but not within the overlapping motifs themselves
** ASM in individual samples that exhibited ASE

Additional Data:

Table S3 (separate Excel file)

A list of correlations between coefficient of constraint and absolute differences in predicted binding likelihood of 203 transcription factors for SD-ASM loci with 2 frequent epialleles.

References and Notes

1. K. Kerkel, A. Spadola, E. Yuan, J. Kosek, L. Jiang, E. Hod, K. Li, V. V. Murty, N. Schupf, E. Vilain, M. Morris, F. Haghighi, B. Tycko, Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* **40**, 904–908 (2008). doi:10.1038/ng.174 Medline

2. L. C. Schalkwyk, E. L. Meaburn, R. Smith, E. L. Dempster, A. R. Jeffries, M. N. Davies, R. Plomin, J. Mill, Allelic skewing of DNA methylation is widespread across the genome. *Am. J. Hum. Genet.* **86**, 196–212 (2010). doi:10.1016/j.ajhg.2010.01.014 Medline

3. Y. Zhang, C. Rohde, R. Reinhardt, C. Voelcker-Rehage, A. Jeltsch, Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol.* **10**, R138 (2009). doi:10.1186/gb-2009-10-12-r138 Medline

4. J. Gertz, K. E. Varley, T. E. Reddy, K. M. Bowling, F. Pauli, S. L. Parker, K. S. Kucera, H. F. Willard, R. M. Myers, Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLOS Genet.* **7**, e1002228 (2011). doi:10.1371/journal.pgen.1002228 Medline

5. A. Hellman, A. Chess, Extensive sequence-influenced DNA methylation polymorphism in the human genome. *Epigenetics Chromatin* **3**, 11 (2010). doi:10.1186/1756-8935-3-11 Medline

6. C. G. Bell, S. Finer, C. M. Lindgren, G. A. Wilson, V. K. Rakyan, A. E. Teschendorff, P. Akan, E. Stupka, T. A. Down, I. Prokopenko, I. M. Morison, J. Mill, R. Pidsley, P. Deloukas, T. M. Frayling, A. T. Hattersley, M. I. McCarthy, S. Beck, G. A. Hitman; International Type 2 Diabetes 1q Consortium, Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLOS ONE* **5**, e14040 (2010). doi:10.1371/journal.pone.0014040 Medline

7. B. Tycko, Allele-specific DNA methylation: Beyond imprinting. *Hum. Mol. Genet.* **19** (R2), R210–R220 (2010). doi:10.1093/hmg/ddq376 Medline

8. S. M. Waszak, O. Delaneau, A. R. Gschwind, H. Kilpinen, S. K. Raghav, R. M. Witwicki, A. Orioli, M. Wiederkehr, N. I. Panousis, A. Yurovsky, L. Romano-Palumbo, A. Planchon, D. Bielser, I. Padioleau, G. Udin, S. Thurnheer, D. Hacker, N. Hernandez, A. Reymond, B. Deplancke, E. T. Dermitzakis, Population variation and genetic control of modular chromatin architecture in humans. *Cell* **162**, 1039–1050 (2015). doi:10.1016/j.cell.2015.08.001 Medline

9. G. McVicker, B. van de Geijn, J. F. Degner, C. E. Cain, N. E. Banovich, A. Raj, N. Lewellen, M. Myrthil, Y. Gilad, J. K. Pritchard, Identification of genetic variants that affect histone modifications in human cells. *Science* **342**, 747–749 (2013). doi:10.1126/science.1242429 Medline

10. W. Sun, J. Poschmann, R. Cruz-Herrera Del Rosario, N. N. Parikshak, H. S. Hajan, V. Kumar, R. Ramasamy, T. G. Belgard, B. Elanggovan, C. C. Y. Wong, J. Mill, D. H. Geschwind, S. Prabhakar, Histone acetylome-wide association study of autism spectrum disorder. *Cell* **167**, 1385–1397.e11 (2016). doi:10.1016/j.cell.2016.10.031 Medline

11. M. F. Lyon, Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**, 372–373 (1961). doi:10.1038/190372a0 Medline

12. Z. Shipony, Z. Mukamel, N. M. Cohen, G. Landan, E. Chomsky, S. R. Zeliger, Y. C. Fried, E. Ainbinder, N. Friedman, A. Tanay, Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115–119 (2014). doi:10.1038/nature13458 Medline

13. G. Landan, N. M. Cohen, Z. Mukamel, A. Bar, A. Molchadsky, R. Brosh, S. Horn-Saban, D. A. Zalcenstein, N. Goldfinger, A. Zundelevich, E. N. Gal-Yam, V. Rotter, A. Tanay, Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**, 1207–1214 (2012). doi:10.1038/ng.2442 Medline

14. E. Florio, S. Keller, L. Coretti, O. Affinito, G. Scala, F. Errico, A. Fico, F. Boscia, M. J. Sisalli, M. G. Reccia, G. Miele, A. Monticelli, A. Scorziello, F. Lembo, L. Colucci-D'Amato, G. Minchiotti, V. E. Avvedimento, A. Usiello, S. Cocozza, L. Chiariotti, Tracking the evolution of epialleles during neural differentiation and brain development: D-Aspartate oxidase as a model gene. *Epigenetics* **12**, 41–54 (2017). doi:10.1080/15592294.2016.1260211 Medline

15. P. Ginart, J. M. Kalish, C. L. Jiang, A. C. Yu, M. S. Bartolomei, A. Raj, Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. *Genes Dev.* **30**, 567–578 (2016). doi:10.1101/gad.275958.115 Medline

16. K. D. Siegmund, P. Marjoram, Y. J. Woo, S. Tavaré, D. Shibata, Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4828–4833 (2009). doi:10.1073/pnas.0810276106 Medline

17. A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis; Roadmap Epigenomics Consortium, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi:10.1038/nature14248 Medline

18. Materials and methods are available as supplementary materials.

19. J. Rozowsky, A. Abyzov, J. Wang, P. Alves, D. Raha, A. Harmanci, J. Leng, R. Bjornson, Y. Kong, N. Kitabayashi, N. Bhardwaj, M. Rubin, M. Snyder, M. Gerstein, AlleleSeq: Analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* **7**, 522 (2011). doi:10.1038/msb.2011.54 Medline

20. M. D. Schultz, Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang, J. R. Ecker, Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015). doi:10.1038/nature14465 Medline

21. D. Leung, I. Jung, N. Rajagopal, A. Schmitt, S. Selvaraj, A. Y. Lee, C.-A. Yen, S. Lin, Y. Lin, Y. Qiu, W. Xie, F. Yue, M. Hariharan, P. Ray, S. Kuan, L. Edsall, H. Yang, N. C. Chi, M. Q. Zhang, J. R. Ecker, B. Ren, Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518**, 350–354 (2015). doi:10.1038/nature14217 Medline

22. J. N. Hutchinson, T. Raj, J. Fagerness, E. Stahl, F. T. Viloria, A. Gimelbrant, J. Seddon, M. Daly, A. Chess, R. Plenge, Allele-specific methylation occurs at genetic variants associated with complex disease. *PLOS ONE* **9**, e98464 (2014). doi:10.1371/journal.pone.0098464 Medline

23. C. Do, C. F. Lang, J. Lin, H. Darbary, I. Krupska, A. Gaba, L. Petukhova, J.-P. Vonsattel, M. P. Gallagher, R. S. Goland, R. A. Clynes, A. Dwork, J. G. Kral, C. Monk, A. M. Christiano, B. Tycko, Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *Am. J. Hum. Genet.* **98**, 934–955 (2016). doi:10.1016/j.ajhg.2016.03.027 Medline

24. C. G. Bell, F. Gao, W. Yuan, L. Roos, R. J. Acton, Y. Xia, J. Bell, K. Ward, M. Mangino, P. G. Hysi, J. Wang, T. D. Spector, Obligatory and facilitative allelic variation in the DNA methylome within common disease-associated loci. *Nat. Commun.* **9**, 8 (2018). doi:10.1038/s41467-017-01586-1 Medline

25. M. Gutierrez-Arcelus, T. Lappalainen, S. B. Montgomery, A. Buil, H. Ongen, A. Yurovsky, J. Bryois, T. Giger, L. Romano, A. Planchon, E. Falconnet, D. Bielser, M. Gagnebin, I. Padioleau, C. Borel, A. Letourneau, P. Makrythanasis, M. Guipponi, C. Gehrig, S. E. Antonarakis, E. T. Dermitzakis, Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2**, e00523 (2013). Medline

26. M. Gutierrez-Arcelus, H. Ongen, T. Lappalainen, S. B. Montgomery, A. Buil, A. Yurovsky, J. Bryois, I. Padioleau, L. Romano, A. Planchon, E. Falconnet, D. Bielser, M. Gagnebin, T. Giger, C. Borel, A. Letourneau, P. Makrythanasis, M. Guipponi, C. Gehrig, S. E. Antonarakis, E. T. Dermitzakis, Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLOS Genet.* **11**, e1004958 (2015). doi:10.1371/journal.pgen.1004958 Medline

27. A. Hellman, A. Chess, Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007). doi:10.1126/science.1136352 Medline

28. W. A. Cheung, X. Shao, A. Morin, V. Siroux, T. Kwan, B. Ge, D. Aïssi, L. Chen, L. Vasquez, F. Allum, F. Guénard, E. Bouzigon, M.-M. Simon, E. Boulier, A. Redensek, S. Watt, A. Datta, L. Clarke, P. Flicek, D. Mead, D. S. Paul, S. Beck, G. Bourque, M.

Lathrop, A. Tchernof, M.-C. Vohl, F. Demenais, I. Pin, K. Downes, H. G. Stunnenberg, N. Soranzo, T. Pastinen, E. Grundberg, Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol.* **18**, 50 (2017). doi:10.1186/s13059-017-1173-7 Medline

29. S. G. Park, S. Hannenhalli, S. S. Choi, Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. *BMC Genomics* **15**, 526 (2014). doi:10.1186/1471-2164-15-526 Medline

30. V. K. Rakyan, M. E. Blewitt, R. Druker, J. I. Preis, E. Whitelaw, Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002). doi:10.1016/S0168-9525(02)02709-9 Medline

31. S. N. Martos, T. Li, R. B. Ramos, D. Lou, H. Dai, J.-C. Xu, G. Gao, Y. Gao, Q. Wang, C. An, X. Zhang, Y. Jia, V. L. Dawson, T. M. Dawson, H. Ji, Z. Wang, Two approaches reveal a new paradigm of 'switchable or genetics-influenced allele-specific DNA methylation' with potential in human disease. *Cell Discov.* **3**, 17038 (2017). doi:10.1038/celldisc.2017.38 Medline

32. J. Davila-Velderrain, J. C. Martinez-Garcia, E. R. Alvarez-Buylla, Modeling the epigenetic attractors landscape: Toward a post-genomic mechanistic understanding of development. *Front. Genet.* **6**, 160 (2015). doi:10.3389/fgene.2015.00160 Medline

33. M. T. Maurano, H. Wang, S. John, A. Shafer, T. Canfield, K. Lee, J. A. Stamatoyannopoulos, Role of DNA methylation in modulating transcription factor occupancy. *Cell Reports* **12**, 1184–1195 (2015). doi:10.1016/j.celrep.2015.07.024 Medline

34. Y. Guo, Q. Xu, D. Canzio, J. Shou, J. Li, D. U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M. Q. Zhang, B. Ren, A. R. Krainer, T. Maniatis, Q. Wu, CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015). doi:10.1016/j.cell.2015.07.038 Medline

35. Z. Tang, O. J. Luo, X. Li, M. Zheng, J. J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Ruszczycki, P. Michalski, E. Piecuch, P. Wang, D. Wang, S. Z. Tian, M. Penrad-Mobayed, L. M. Sachs, X. Ruan, C.-L. Wei, E. T. Liu, G. M. Wilczynski, D. Plewczynski, G. Li, Y. Ruan, CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015). doi:10.1016/j.cell.2015.11.024 Medline

36. A. Jolma, J. Yan, T. Whitington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, J. Taipale, DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013). doi:10.1016/j.cell.2012.12.009 Medline

37. R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E.

Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, J. A. Stamatoyannopoulos, The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012). doi:10.1038/nature11232 Medline

38. A. Feldmann, R. Ivanek, R. Murr, D. Gaidatzis, L. Burger, D. Schübeler, Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLOS Genet.* **9**, e1003994 (2013). doi:10.1371/journal.pgen.1003994 Medline

39. E. Hervouet, F. M. Vallette, P. F. Cartron, Dnmt1/Transcription factor interactions: An alternative mechanism of DNA methylation inheritance. *Genes Cancer* **1**, 434–443 (2010). doi:10.1177/1947601910373794 Medline

40. E. Hervouet, F. M. Vallette, P. F. Cartron, Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics* **4**, 487–499 (2009). doi:10.4161/epi.4.7.9883 Medline

41. A. Nayak, J. Glöckner-Pagel, M. Vaeth, J. E. Schumann, M. Buttmann, T. Bopp, E. Schmitt, E. Serfling, F. Berberich-Siebelt, Sumoylation of the transcription factor NFATc1 leads to its subnuclear relocalization and interleukin-2 repression by histone deacetylase. *J. Biol. Chem.* **284**, 10935–10946 (2009). doi:10.1074/jbc.M900465200 Medline

42. H. Zeng, D. K. Gifford, Predicting the impact of non-coding variants on DNA methylation. *Nucleic Acids Res.* **45**, e99 (2017). doi:10.1093/nar/gkx177 Medline

43. N. E. Banovich, X. Lan, G. McVicker, B. van de Geijn, J. F. Degner, J. D. Blischak, J. Roux, J. K. Pritchard, Y. Gilad, Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLOS Genet.* **10**, e1004663 (2014). doi:10.1371/journal.pgen.1004663 Medline

44. K. Kin, X. Chen, M. Gonzalez-Garay, W. D. Fakhouri, The effect of non-coding DNA variations on P53 and cMYC competitive inhibition at cis-overlapping motifs. *Hum. Mol. Genet.* **25**, 1517–1527 (2016). doi:10.1093/hmg/ddw030 Medline

45. D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, H. Parkinson, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42** (D1), D1001–D1006 (2014). doi:10.1093/nar/gkt1229 Medline

46. D. R. De Silva, R. Nichols, G. Elgar, Purifying selection in deeply conserved human enhancers is more consistent than in coding sequences. *PLOS ONE* **9**, e103357 (2014). doi:10.1371/journal.pone.0103357 Medline

47. L. D. Ward, M. Kellis, Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* **337**, 1675–1678 (2012). doi:10.1126/science.1225057 Medline

48. A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis; 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi:10.1038/nature15393 Medline

49. J. Zhu, F. He, S. Hu, J. Yu, On the nature of human housekeeping genes. *Trends Genet.* **24**, 481–484 (2008). doi:10.1016/j.tig.2008.08.004 Medline

50. M. T. Maurano, E. Haugen, R. Sandstrom, J. Vierstra, A. Shafer, R. Kaul, J. A. Stamatoyannopoulos, Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47**, 1393–1401 (2015). doi:10.1038/ng.3432 Medline

51. S. Gravina, X. Dong, B. Yu, J. Vijg, Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* **17**, 150 (2016). doi:10.1186/s13059-016-1011-3 Medline

52. G. Elliott, C. Hong, X. Xing, X. Zhou, D. Li, C. Coarfa, R. J. A. Bell, C. L. Maire, K. L. Ligon, M. Sigaroudinia, P. Gascard, T. D. Tlsty, R. A. Harris, L. C. Schalkwyk, M. Bilenky, J. Mill, P. J. Farnham, M. Kellis, M. A. Marra, A. Milosavljevic, M. Hirst, G. D. Stormo, T. Wang, J. F. Costello, Intermediate DNA methylation is a conserved signature of genome regulation. *Nat. Commun.* **6**, 6363 (2015). doi:10.1038/ncomms7363 Medline

53. G. Jenkinson, E. Pujadas, J. Goutsias, A. P. Feinberg, Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat. Genet.* **49**, 719–729 (2017). doi:10.1038/ng.3811 Medline

54. H. H. McAdams, A. Arkin, It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69 (1999). doi:10.1016/S0168-9525(98)01659-X Medline

55. A. Alaghi, J. P. Hayes, Survey of stochastic computing. *ACM Trans. Embed. Comput. Syst.* **12**, 1 (2013). doi:10.1145/2465787.2465794

56. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324 Medline

57. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, M. A. DePristo, From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 1–33 (2013). Medline

58. Z. Huang, N. Rustagi, N. Veeraraghavan, A. Carroll, R. Gibbs, E. Boerwinkle, M. G. Venkata, F. Yu, A hybrid computational strategy to address WGS variant analysis in >5000 samples. *BMC Bioinformatics* **17**, 361 (2016). doi:10.1186/s12859-016-1211-6 Medline

59. D. Challis, J. Yu, U. S. Evani, A. R. Jackson, S. Paithankar, C. Coarfa, A. Milosavljevic, R. A. Gibbs, F. Yu, An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* **13**, 8 (2012). doi:10.1186/1471-2105-13-8 Medline

60. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011). doi:10.1038/ng.806 Medline

61. A. Rimmer, H. Phan, I. Mathieson, Z. Iqbal, S. R. F. Twigg, A. O. M. Wilkie, G. McVean, G. Lunter; WGS500 Consortium, Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014). doi:10.1038/ng.3036 Medline

62. Y. Wang, J. Lu, J. Yu, R. A. Gibbs, F. Yu, An integrative variant analysis pipeline for accurate genotype/haplotype inference in population NGS data. *Genome Res.* **23**, 833–842 (2013). doi:10.1101/gr.146084.112 Medline

63. A. Abyzov, A. E. Urban, M. Snyder, M. Gerstein, CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011). doi:10.1101/gr.114876.110 Medline

64. C. Coarfa, F. Yu, C. A. Miller, Z. Chen, R. A. Harris, A. Milosavljevic, Pash 3.0: A versatile software package for read mapping and integrative analysis of genomic and epigenomic variation using massively parallel DNA sequencing. *BMC Bioinformatics* **11**, 572 (2010). doi:10.1186/1471-2105-11-572 Medline

65. Y. Liu, K. D. Siegmund, P. W. Laird, B. P. Berman, Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biol.* **13**, R61 (2012). doi:10.1186/gb-2012-13-7-r61 Medline

66. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi:10.1038/nature11247 Medline

67. F. Fang, E. Hodges, A. Molaro, M. Dean, G. J. Hannon, A. D. Smith, Genomic landscape of human allele-specific DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7332–7337 (2012). doi:10.1073/pnas.1201310109 Medline

68. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi:10.1093/bioinformatics/btp352 Medline

69. R. M. Kuhn, D. Haussler, W. J. Kent, The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013). doi:10.1093/bib/bbs038 Medline

70. W. Guo, P. Zhu, M. Pellegrini, M. Q. Zhang, X. Wang, Z. Ni, CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34**, 381–387 (2018). doi:10.1093/bioinformatics/btx595 Medline

71. S. G. Coetzee, G. A. Coetzee, D. J. Hazelett, motifbreakR: An R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015). Medline

72. K. G. Ardlie, D. S. Deluca, A. V. Segre, T. J. Sullivan, T. R. Young, E. T. Gelfand, C. A. Trowbridge, J. B. Maller, T. Tukiainen, M. Lek, L. D. Ward, P. Kheradpour, B. Iriarte, Y. Meng, C. D. Palmer, T. Esko, W. Winckler, J. N. Hirschhorn, M. Kellis, D. G. MacArthur, G. Getz, A. A. Shabalin, G. Li, Y.-H. Zhou, A. B. Nobel, I. Rusyn, F. A. Wright, T. Lappalainen, P. G. Ferreira, H. Ongen, M. A. Rivas, A. Battle, S. Mostafavi, J. Monlong, M. Sammeth, M. Mele, F. Reverter, J. M. Goldmann, D. Koller, R. Guigo, M.

I. McCarthy, E. T. Dermitzakis, E. R. Gamazon, H. K. Im, A. Konkashbaev, D. L. Nicolae, N. J. Cox, T. Flutre, X. Wen, M. Stephens, J. K. Pritchard, Z. Tu, B. Zhang, T. Huang, Q. Long, L. Lin, J. Yang, J. Zhu, J. Liu, A. Brown, B. Mestichelli, D. Tidwell, E. Lo, M. Salvatore, S. Shad, J. A. Thomas, J. T. Lonsdale, M. T. Moser, B. M. Gillard, E. Karasik, K. Ramsey, C. Choi, B. A. Foster, J. Syron, J. Fleming, H. Magazine, R. Hasz, G. D. Walters, J. P. Bridge, M. Miklos, S. Sullivan, L. K. Barker, H. M. Traino, M. Mosavel, L. A. Siminoff, D. R. Valley, D. C. Rohrer, S. D. Jewell, P. A. Branton, L. H. Sobin, M. Barcus, L. Qi, J. McLean, P. Hariharan, K. S. Um, S. Wu, D. Tabor, C. Shive, A. M. Smith, S. A. Buia, A. H. Undale, K. L. Robinson, N. Roche, K. M. Valentino, A. Britton, R. Burges, D. Bradbury, K. W. Hambright, J. Seleski, G. E. Korzeniewski, K. Erickson, Y. Marcus, J. Tejada, M. Taherian, C. Lu, M. Basile, D. C. Mash, S. Volpi, J. P. Struewing, G. F. Temple, J. Boyer, D. Colantuoni, R. Little, S. Koester, L. J. Carithers, H. M. Moore, P. Guan, C. Compton, S. J. Sawyer, J. P. Demchok, J. B. Vaught, C. A. Rabiner, N. C. Lockhart, K. G. Ardlie, G. Getz, F. A. Wright, M. Kellis, S. Volpi, E. T. Dermitzakis; GTEx Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi:10.1126/science.1262110 Medline

73. A. Mathelier, O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, W. W. Wasserman, JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44** (D1), D110–D115 (2016). doi:10.1093/nar/gkv1176 Medline

74. W. D. Fakhouri, F. Rahimov, C. Attanasio, E. N. Kouwenhoven, R. L. Ferreira De Lima, T. M. Felix, L. Nitschke, D. Huver, J. Barrons, Y. A. Kousa, E. Leslie, L. A. Pennacchio, H. Van Bokhoven, A. Visel, H. Zhou, J. C. Murray, B. C. Schutte, An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. *Hum. Mol. Genet.* **23**, 2711–2720 (2014). doi:10.1093/hmg/ddt664 Medline

75. J. Malone, E. Holloway, T. Adamusiak, M. Kapushesky, J. Zheng, N. Kolesnikov, A. Zhukova, A. Brazma, H. Parkinson, Modeling sample variables with an experimental factor ontology. *Bioinformatics* **26**, 1112–1118 (2010). doi:10.1093/bioinformatics/btq099 Medline

76. K. Eilbeck, S. E. Lewis, Sequence ontology annotation guide. *Comp. Funct. Genomics* **5**, 642–647 (2004). doi:10.1002/cfg.446 Medline

77. C. J. Mungall, C. Torniai, G. V. Gkoutos, S. E. Lewis, M. A. Haendel, Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* **13**, R5 (2012). doi:10.1186/gb-2012-13-1-r5 Medline

78. R. Tewhey, D. Kotliar, D. S. Park, B. Liu, S. Winnicki, S. K. Reilly, K. G. Andersen, T. S. Mikkelsen, E. S. Lander, S. F. Schaffner, P. C. Sabeti, Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**, 1519–1529 (2016). doi:10.1016/j.cell.2016.04.027 Medline