Supplemental Information

Supplemental figures and tables

Supplemental Figure 1. Additional Massarray methylation data to Figure 2B.

Supplemental Figure 2. Additional data to Figure 3.

Supplemental Figure 3. Additional data to Figure 4E,F.

Supplemental Figure 4. Methylation-RNA expression data.

Supplemental Table 1. Additional data to Figure 1B, lists of m-DMRs.

Supplemental Table 2. Additional data to Figure 5D.

Supplemental Materials and Methods

ES cell culture

E14 ES cells were cultured on feeder cells (mitomycin-inactivated mouse embryonic fibroblasts (MEFs)) for at least two passages after thawing. Cells were split every 2 d with plating densities between 1.5 x 10^6 and 4 x 10^6 cells on 10-cm cell culture plates. ES medium was based on DMEM containing 15% FBS (ES cell qualified), LIF (1,000 U/ml, Millipore, Billerica, MA, USA), 1X non-essential amino acids, 2mM L-glutamine and β-mercaptoethanol (Invitrogen, Carlsbad, CA, USA).

DNA methylation HELP arrays

The HELP (*HpaII* tiny fragment enrichment by ligation-mediated PCR) assay was carried out as previously described[1-3] in the Epigenomics Core Facility of the Weill Cornell Medical College.

Briefly, two samples of one microgram genomic DNA each were digested overnight with *Hpa*II and *Msp*I (New England Biolabs, Ipswich, MA, USA). Adapters were ligated to the DNA ends and the fragments were amplified by ligation mediated PCR optimized for fragment size between 200 and 2,000 bp. The *Hpa*II and *Msp*I representations were then labeled with Cy5 and Cy3, respectively, followed by the co-hybridization of the labeled fragments to Roche 25K custom arrays representing mouse promoters and CpG islands. The arrays were scanned using a GenePix 4000B scanner (Axon Instruments, Sunnyvale, CA, USA).

HELP data normalization

HELP data were preprocessed and normalized using the HELP data analysis package[4] implemented in R. Normalized HELP methylation signal was compared across groups with the Limma R package. Limma implements moderated variance estimates especially useful with small number of biological replicates in each group. The Limma P-values were adjusted for multiple testing with the Benjamini Hochberg method. HELP fragments were considered differentially methylated if their fold-change crossed zero (indicating a qualitative change in average methylation state for the fragment) and the Limma BH adjusted P-values was $q<0.005$ (stringent estimated false discovery rate of less than 0.5%). Differentially methylated regions (DMRs) were classified based on their CpG density and location. CpG islands were defined as >200bp with GC content of 50% or greater. Islands were divided to strong ($CpG^{o/e}$ >0.80) and weak (0.8 >$CpG^{o/e}$ >0.60) and island shores were defined as 2kb regions around islands. Low CpG regions had a $CpG^{o/e}$ <0.60.

DNA methylation sequencing by MassARRAY EpiTYPER

The primers were designed using the Sequenom EpiDesigner beta software (http://www.epidesigner.com/). The primer sequences are displayed below.

Atbf1

1	TTAAGTTTATGTAGTATTTTAGGGGTTTAG	TTCATCTTCAAAACTTACAATCTAAAAAT
2	TTTTAAAAGGATATAATTTAATAGGGTTAG	ACCTCAAATTCATACAACACCTCAA
3	TTTTTTTAAAGGTATTATTGGTTTGG	TTCTCCCCTAAAAATTAACTTCAAC
4	TTGTAATAAGGTGGAGTGTTTTTTT	AAATTATTTTCCCATATACCTATCTATACC
5	GTGGTGAATTTGTAAGAGATGGTGT	AAAACTAAACCCCTAAAATACTACATAAA
6	TTTTTGAGGTGTTGTATGAATTTGA	AAAAACCACCTAAAATCCCTCTACT
7	GTTATTATGGTAATGGTTTTTTAGTTATTT	TAAAAAAACCTCTCCTTTCTCCTTC
8	GTTTTTAAGAAGGAGAAAGGAGAGG	AAAAACAAACCTTCCATACCATACA

Smo

1	GGAATTTATTTTGTAGATTAGGTTGG	AAACTCACAATTCTAAATCATAATCCA
2	AGAAATTTTATGAGGTAGTTGGGTT	AAAACAAACAAAAATTTTCACTCCA
3	TTAAAGATTTAGTTAAGTGTTTTTGGGA	CAACCCCCTAAACTCTCCCTAC
4	TAGTTGGTTTTGTTTTTTGGAATGT	CCAACTAAAAATTCAATCAAATACCTC
5	GAGAGTAGGGTTAGTTAGAGTAATAAAGGA	AAACTATCTTCAACCCTAAAAACC
6	GAAGTTGTTTTTAATTTTGGGAATTT	AAAACTAAAACTCCTCCTCTCCAAC
7	GGAGGGTTTTTAGGGTTGAAGATAG	CCAACAATACCAACAACAACAACTA
8	AGTTGTTGTTGTTGGTATTGTTGG	ACTAACTTCCCTAATCTCTTACCCC
9	TGGTTAAATAGTTAATTTAGTAAAAGTTGA	AAAAATTCCCAAAATTAAAAACAAC
10	AATAGTTTGAGGTTTGAGTTTTTTTT	TCAACTTTTACTAAATTAACTATTTAACCA

Enhanced Reduced Representation Bisulfite Sequencing (ERRBS)

DNA was digested with MspI restriction enzyme. This was followed by end repair and ligation of paired end Illumina sequencing adaptors fully methylated at all cytosines. Size selection for library sizes of 150-400 bp was performed followed by a single round of bisulfite treatment using the EZ DNA Methylation Kit (Zymo Research). PCR amplification using Illumina PCR PE1.0 and 2.0 was followed by product isolation using AMPure XP beads per manufacturer's

recommended protocol (Agencort). Quality control was performed using quantitation on a Qubit 2.0 fluorometer (Invitrogen) and library visualization using a Quant-iT dsDNA HS Assay Kit for (Agilent 2100 Bioanalyzer). The amplified libraries were sequenced using a 50bp single end read run on a HiSeq2000 per manufacturer's recommended protocol. Image capture, analysis and base calling were performed using Illumina's CASAVA 1.7.

ERRBS read mapping

The last bisulfite plugin parallelizes alignments with the last aligner, and otherwise follows the recommended protocol for aligning bisulfite reads with this aligner (i.e., see http://last.cbrc.jp/doc/bisulfite.txt).

ERRBS methylation rate estimation

Methylation rates were estimated with GobyWeb and the SEQ_VAR_GOBY_METHYLATION plugin. This plugin determine when methylation events occurs at a given genomic location. Events are defined as observing a C in the read when the reference has a C (methylation event on the forward strand) or observing a T in the read when the reference has a C (non-methylation event on the forward strand). Similarly, G/G and G>A observations define methylation and non-methylation events for the reverse strand, respectively. Methylation rates were estimated for sites were more than 35 events were observed. Methylation rates were estimated as the number of methylation events divided by the sum of non-methylation and methylation events. A methylation rate of 100% indicates that all events support methylation at this site. To identify differentially methylated sites, we calculate a Fisher Exact test comparing the number of methylation events and non-methylation events at a site between two groups of samples as reported earlier[5]. The Fisher p-values are adjusted for multiple testing across all sites observed with more than 35 events across the genome. This is done with the Goby fdr implementation of

the Benjamini-Hochberg method (see http://goby.campagnelab.org). Sites are considered significantly differentially methylated when the adjusted q-value is less than 0.01 and the difference in methylation rate is larger than 20% between the groups.

RNA extraction and RNA sequencing (RNA-Seq)

Total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, CA, USA). RNA concentration was measured using NanoDrop® ND-1000 (Thermo Scientific, Wilmington, DE, USA) and RNA quality control was performed by using Bioanalyzer 2100 (Agilent Technologies, Palo Alto, California, USA). RNA was fragmented with divalent cations at high temperature and converted to cDNA libraries following the Illumina recommended sample preparation guide (Document 1004898 Rev. D) using Illumina kits (Illumina, San Diego, CA, USA)[6]. The libraries were sequenced on the Illumina GAIIx instrument (one sample per lane), with the single end protocol and 42 cycles of sequencing.

RNA-Seq Data Analysis

RNA-Seq data were received as FASTQ files from the core facility and uploaded to a local instance of GobyWeb (http://gobyweb.campagnelab.org). Alignments were performed with and the bwa aligner[7] against the MM9 mouse reference genome. Alignments were filtered to keep only reads that matched with less than 5% sequence differences (accepting 2 mismatches at most over a 42 bp read) and to exclude those generated from reads that mapped in more than one location in the reference genome. Differential expression analysis was conducted with GobyWeb. Briefly, alignments were used to estimate the number of reads that match gene annotations with the Goby alignment-to-counts mode. Annotations were obtained the Ensembl release corresponding to NCBI37.55/MM9. Gene counts were estimated as the sum of the number of reads that partially overlap with any of the exons of a gene, but do not lie completely

within the introns of the gene. Counts were compared between groups with a Fisher exact test (R implementation) adjusted for multiple testing with the method of Benjamini Hochberg (adjusted Fisher exact test P-value<0.01 and fold-change>1.3 in either direction of change).

Functional Enrichment Analysis

The list of differentially methylated genes was analyzed through the use of Ingenuity Pathways Analysis (http://www.ingenuity.com). The Functional Analysis identified the biological functions that were most significant to the data set. Right-tailed Fisher's exact test was used to calculate a p-value determining the probability that each biological function assigned to the data set is due to chance alone. Analysis of functional enrichment was carried out using MetaCore from MetaCore from Thomson Reuters (http://www.genego.com/metacore.php, version 6.10) searching for enrichment in the manually curated GeneGO Process Networks, representing a pre-set network of protein interactions. Statistical significance of a process was determined using the Hypergeometric distribution and adjusted for multiple testing by MetaCore.

Supplemental references

1. Figueroa ME, Melnick A, Greally JM. Genome-wide determination of DNA methylation by Hpa II tiny fragment enrichment by ligation-mediated PCR (HELP) for the study of acute leukemias. *Methods Mol Biol* 2009; **538**: 395-407.
2. Figueroa ME, Reimers M, Thompson RF, Ye K, Li Y, Selzer RR *et al.* An integrative genomic and epigenomic approach for the study of transcriptional regulation. *PLoS ONE* 2008; **3**: e1882.
3. Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E *et al.* Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res* 2006; **16**: 1046-1055.
4. Thompson RF, Reimers M, Khulan B, Gissot M, Richmond TA, Chen Q *et al.* An analytical pipeline for genomic representations used for cytosine methylation studies. *Bioinformatics* 2008; **24**: 1161-1167.
5. Gu H, Bock C, Mikkelsen TS, Jäger N, Smith ZD, Tomazou E *et al.* Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. *Nature methods* 2010; **7**: 133-136.
6. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific

cleavage and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 2005; **102**: 15785-15790.
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-1760.