Supplementary Methods

Read Mapping

Single end 41 bp Illumina reads, which covered 11 cell lines, were merged with newer paired end 36 bp Illumina reads generated from 8 cell lines. Merged together, the combined data set represented 16 cell lines. The single end reads did not contain quality scores while the paired end data set did contain them (+64 offset). Single end reads were uniformly assigned 20 as a Phred-like quality score for each base. We assigned this score so that the single end reads would pass quality filtering but were not given the same confidence as reads assigned higher quality scores by Illumina's software. For each cell line, the reads were sorted into two categories. The first category contained reads that represented bisulfite converted sequences and the second category included the reverse complementary reads. The ratio of T/C nucleotides was compared to the ratio of A/G nucleotides to determine to which category a read belonged. Reads that represented bisulfite converted sequences had a higher T/C ratio than A/G, since unmethylated cytosines were converted to thymines. Reads classified as reverse complementary were converted to bisulfite converted sequences (i.e. the reverse complement was taken). The reads were merged, demthylated in silico, and mapped to the unmethylated bisulftie converted hg18 genome. Soap 2.20 mapped sequences to the reference via an end-to-end policy that allowed up to two mismatches. Reads that mapped to multiple locations were discarded. For the paired end datasets, the paired ends were mapped independently of each other. Mates of paired end sequences that mapped to separate chromosomes were treated as single end reads. If both mates of a paired end sequence mapped uniquely, the mapped locations of these reads were checked. If the mates' sequences overlapped (i.e. distance less than 36 bp) or were greater than 250 bp from each other, the mates were treated as single end reads. For the remaining paired end reads, the distribution of distances between mates was calculated per cell line. The distributions were assumed to be normal and the distribution's parameters were calculated for each cell line. Paired end reads whose mate distances fell outside of 2.5 standard deviations from the average mate distance were treated as single end reads. The remaining reads were considered valid paired end reads. The generation of methylation frequency values for each targeted CpG was calculated analogously to the previously published protocol.

Sanger reads were mapped to a reference template using blat. Due to the much larger size of the sequence length and its known location on the genome, gaps and mismatches were allowed during the alignment. Analogous to Illumina reads, T/C and A/G ratios for Sanger sequences were calculated and if needed, Sanger sequences were reverse complemented so that they would match the reference template in the forward direction. The sequences were then demethylated in silico and mapped to an unmethylated template using blat. The mapped read coordinates for each in silico demethylated read were then transferred to the original sequence. Sanger Sequence diagrams were created by aligning reads according to CpG sites and grouped together based on the nucleotide present at the SNP position.

SNP Calling

Heterozygous SNPs were detected using an algorithm that assigned each genotype a probability. For example, a diploid cell was tested for ten possible genotypes at a given nucleotide position: AA, AT, AG, AC, GG, GC, GT, CC, CT, TT. The possible genotypes were bisulfite converted. The inputs for the SNP calling algorithm are the following: (1) reads that covered the nucleotide position, (2) the quality scores of those reads, and (3) the SNP129 dbSNP candidate positions and SNP identities. Base calls at the examined SNP site needed to have a minimum Phred-like quality score of 15 and the three flanking base calls on needed to have an minimum quality score of 15 on either side. Bisulfite conversion eliminates the complementarity between the Watson and Crick strands and the Watson and Crick strands were thusly treated independently. If reads mapped sufficiently to both strands, the bases at this position were examined to ensure that the bases on Watson and Crick strands were indeed

reverse complementary. If a certain base was present in more than 20% of the Watson strand reads, its reverse complement needed to be present on at least 20% of the Crick strand reads. If these criteria were not met, this location was not analyzed. For sites that passed this filter, a nucleotide frequency matrix was constructed for each nucleotide position based on the read data. The Phred-like scores were used to weight the nucleotide count contributions to the nucleotide frequency matrix. For example, a nucleotide call with a Phred-like score of 20 led to a 0.9 contribution while a nucleotide call of 30 led to a contribution of 0.999. This weighting scheme attenuated the effects from lower quality base calls. The weighted matrix was normalized so that the sum of all entries equaled 1. Each entry in the matrix was then multiplied by the read count.

A Fisher's Test was used to calculate the probability that the nucleotide frequency matrix represented a specific genotype. To test genotype AG, the Fisher's Test compared the frequency that A appeared in the read data to the expected frequency A would appear if this position were an A/G heterozygous SNP. A second Fisher's Test was performed that compared the frequency G from the read data to the expected frequency at an A/G SNP site. The two p-values were multiplied together to generate a stranded p-value product for a specific genotype. The product of the strand specific p-value products represented the likelihood of a specific genotype at a nucleotide position. The likelihoods of all genotypes were then normalized so that the sum of these likelihoods was 1. To filter out false positives, a SNP candidate site needed to have an odds ratio greater than 10 relative to the next most likely genotype. In the case of Hybrid1, the most likely genotype at a SNP candidate site needed to have an odds ratio greater than 100,000 relative to the second most likely genotype. This stricter threshold was implemented due to SNP calling on a tetraploid cell line. SAMtools was not designed for tetraploid cells and we were unable to verify our SNP calls with SAMtools for Hybrid1. Only SNP sites reported by the SNP129 database were examined. Each examined SNP candidate site needed to have at least 10x read depth. With double stranded information, the SNP calling algorithm was able to discern the original SNP identity. Some single stranded SNP identities, however, could not be clearly identified with bisulfite converted reads. Since reads were demethyalted in silico during the SNP analysis, C/T SNPs were not called and A/C and A/T SNPs were not resolved for single stranded SNP calls. However, since SNP calls were made at SNP129 sites, the single stranded SNP identity could be clearly discerned based on the allelic information in the SNP 129. SNP calls that were not consistent with the SNP129 database were excluded. If a called SNP created a CpG dinucleotide or destroyed a CpG dinculeotide relative to the reference genome, it was recorded. The methylation frequency of found CpG dinucleotides not present in the reference were investigated analogous to the method state above. These new CpG dinucleotide sites were used in the SNP ASM and LD analyses.

We also called SNPs using SAMtools v 0.1.7 in order to improve the confidence in SNP calls made with bisulfite converted read data. Mapped reads were demethylated and SAMtools created a consensus sequence. Examining sites with at least 10x read depth, we searched for base calls in the consensus sequence with a minimum SNP score of 20 that did not match the reference. Analogous to our own SNP calling, SNP sites with double stranded coverage allowed for unambiguous SNP calling. We used the intersection of SNP calls between SAMtools to form a confident SNP candidate list, which we used in the other analyses.

ASM SNP Identification

A SNP site was labeled as an allele specifically methylated (ASM) region if the region met one of the following criteria: (1) there was a single CpG site that showed significant ASM, (2) all CpGs, when summed together, showed significant ASM, and (3) all non-SNP overlapping CpGs, when summed together, showed significant ASM. To calculate ASM, a contingency table was created where the columns represented alleles and the rows represented the cytosine (methylated) and thymine (unmethylated) counts at CpG sites found on those alleles. A Fisher's Exact Test was used to produce a two sided p-value, which served as the metric for ASM. For

(1), each CpG site was treated independently and a contingency table was thusly made for each CpG site. For (2) and (3), one contingency table was made and the cytosine and thymine counts were summed across multiple CpGs per allele. If the p-value for any of these analyses was less then 0.001 and the methylation frequency difference between the alleles was greater than 0.1, the SNP region was labeled as ASM. The methylation frequency difference serves to filter out regions with very high read coverage (1,000+) that the Fisher Test would identify as significant ASM even though the allelic methylation frequency difference would be quite small. There were three ASM categories. Category I ASM is not solely dependent on a SNP present in a CpG site while category II ASM is solely dependent on the presence of a SNP at a CpG site. Category III regions showed no ASM. For example, if a SNP region had a significant p-value in analysis (1) but the ASM CpG overlapped with a SNP, then this region was usually labeled as category II ASM. If a SNP site had a significant p-value in analysis (2) but not in (3) or (1), it was labeled as category II ASM. SNPs that had a significant p-value in (1) for CpGs that did not overlap with SNPs or had a significant p-value in (3) were labeled as category I ASM. Using the fdrtool in R, we calculated the FDR for our per CpG ASM and average ASM calls across all cell lines. A p-value cutoff of 0.001 yields a 0.62% FDR for our per CpG ASM calls and a 0.25% FDR for our average SNP ASM calls. Sanger sequences were labeled using the same criteria as the Illumina reads. Regarding the Illumina data, CpG sites with less than 5x read depth on either allele were not considered in the ASM SNP analysis. There was no minimum read depth requirement for the Sanger data.

LD analysis

The r^2 metric was adopted from linkage disequilibrium analysis to measure the organization of methylation at CpG sites on the same read sequence. Reads that contained more than one CpG were used in this analysis and the methylation status of all present CpGs was recorded. For the Illumina data, only CpG pairs that were covered by at least 10 reads were considered in this analysis. For Sanger data, only CpG pairs covered by at least 3 reads were considered. For each recorded CpG pair, a contingency table was constructed, which counted the combinatorial methylation states of the CpG pair (i.e. both methylated, both unmethylated, or mixed methylation states):

CpG 1 / CpG 2	**Methylated**	**Unmethylated**
Methylated	$F_{Methyl\|Methyl}$	$F_{Methyl\|Unmethyl}$
Unmethylated	$F_{Unmethyl\|Methyl}$	$F_{Unmethyl\|Unmethyl}$

The r^2 values were calculated based on data in the contingency table via the following equation:

$$r^2 = \frac{\left(F_{Methyl|Methyl}F_{Unmethyl|Unmethyl} - F_{Methyl|Unmethyl}F_{Unmethyl|Methyl}\right)^2}{F_{Methyl|*}F_{Unmethyl|*}F_{*|Unmethyl}F_{*|Methyl}}$$

CpG sites that displayed uniform methylation patterns were not considered in this analysis since they did not produce meaningful r2 values. LD blocks were created based on the presence of a CpG pair that contained a significant r^2 value (i.e. $r^2 > 0.3$). These blocks were extended if there was either an overlapping CpG pair with a significant r^2 value or there was a CpG pair with a significant r^2 value within 100 bp. Since we were interested in looking at regions of extended organized methylation, LD blocks less than 100 bp or containing less than 10 CpG pairs with $r^2 > 0.3$ were filtered out.