Supplemental Tables

Table S1 Dataset characteristics. Average coverage values are at CpG sites only. Average promoter coverage is calculated for CpG sites found within ±500 bp of a gene's TSS as identified by RefSeq gene annotations. Number of bp indicates the number of DNA basepairs (bps) sequenced in the experiment.

Dataset	Sample	Methylation Assay	Avg. Coverage	Avg. Promoter Coverage	Number of Reads	Number of bp
HMEC-HCC1954	HMEC	WGBS	22.6	19.6	6.6×10^8	59.7 Gb
	HCC1954	WGBS	30.7	27.5	8.9×10^8	80.2 Gb
IMR90-H1	IMR90	WGBS	28.5	14.1	11.8×10^8	91.0 Gb
	H1	WGBS	25.6	11.2	11.6×10^8	87.5 Gb
MCF7-T47D	MCF7	Methyl-MAPS	31.3	25.4	34.4×10^6 McrBC 37.9×10^6 RE	na
	T47D	Methyl-MAPS	21.8	22.5	30.3×10^6 McrBC 23.9×10^6 RE	na

Table S2 To examine the rate of false positives due to errant clusters, we randomly permuted the expression values of all genes. We clustered the methylation signatures as before, using default parameters for a 10kb window centered at the TSS. Any clusters that are identified as significant using a scrambled set of expression values are errant clusters. Results for 1000 permutations for each data set are shown. The errant cluster rate is the fraction of simulations with an errant cluster.

	Errant Cluster Rate	Avg Genes per Errant Cluster
HMEC-HCC1954	1.70%	16.3
H1-IMR90	2.30%	16.0
MCF7-T47D	2.10%	14.7

Table S3 Complete set of parameters considered for optimization of DMR-based (upper) and promoter-based (lower) methods. The DMR procedure is described in the Methods.

DMR Approach Parameter Space	
FDR Cutoff	0.01, 0.05, 0.1
Differential CpG Methyl. Threshold	0.3, 0.5, 0.7
First Window Size	100, 200, 300, 400, 500, 1000, 2000, 3000
Min. Number of Differential CpGs in First Window	2, 3, 4, 5, 7, 10
Second Window Size	200, 300, 400, 500, 1000, 2000, 4000
Min. Number of Differential CpGs in Second Window	3, 4, 5, 7, 10, 15, 20

Promoter Approach Parameter Space	
Window Size Upstream of TSS	0, 100, 250, 500, 1000, 2000, 5000
Window Size Downstream of TSS	0, 100, 250, 500, 1000, 2000, 5000
FDR Cutoff	off, 0.01, 0.05, 0.1
Min Number of CpGs	5, 10, 20, 30, 100
Min Average Methylation Change	0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8

Table S4 Sequencing and mapping statistics for RNA-Seq and Methyl-MAPS experiments. Reads sequenced for Methyl-MAPS libraries are for the total number of R3 reads produced by SOLiD sequencing.

	Sample	Library	Reads Sequenced	Read Pairs Uniquely Mapped
Methyl-MAPS	T47D	McrBC	64,666,229	30,290,561
		RE	63,250,316	23,945,047
	MCF7	McrBC	66,068,188	34,427,094
		RE	80,235,306	37,871,733

	Sample	Replicate	Reads Sequenced	Reads Uniquely Mapped
RNA-Seq	T47D	1	6,648,462	4,698,558
		2	8,178,748	5,808,317
	MCF7	1	17,407,241	15,233,215
		2	26,088,462	22,472,874