

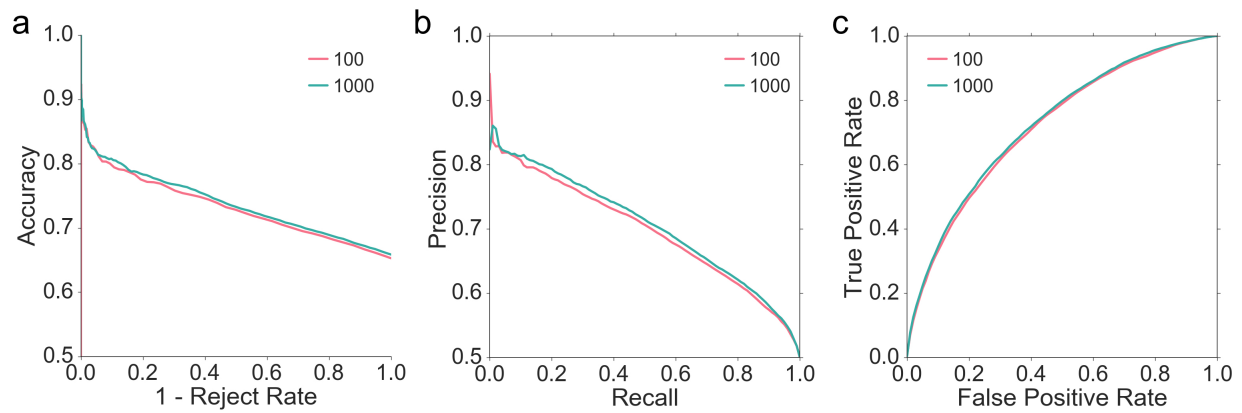
## Supplementary Tables and Figures

**Supplementary Table 1.** Roadmap Epigenomics Project (REP) biological and technical replicate counts.

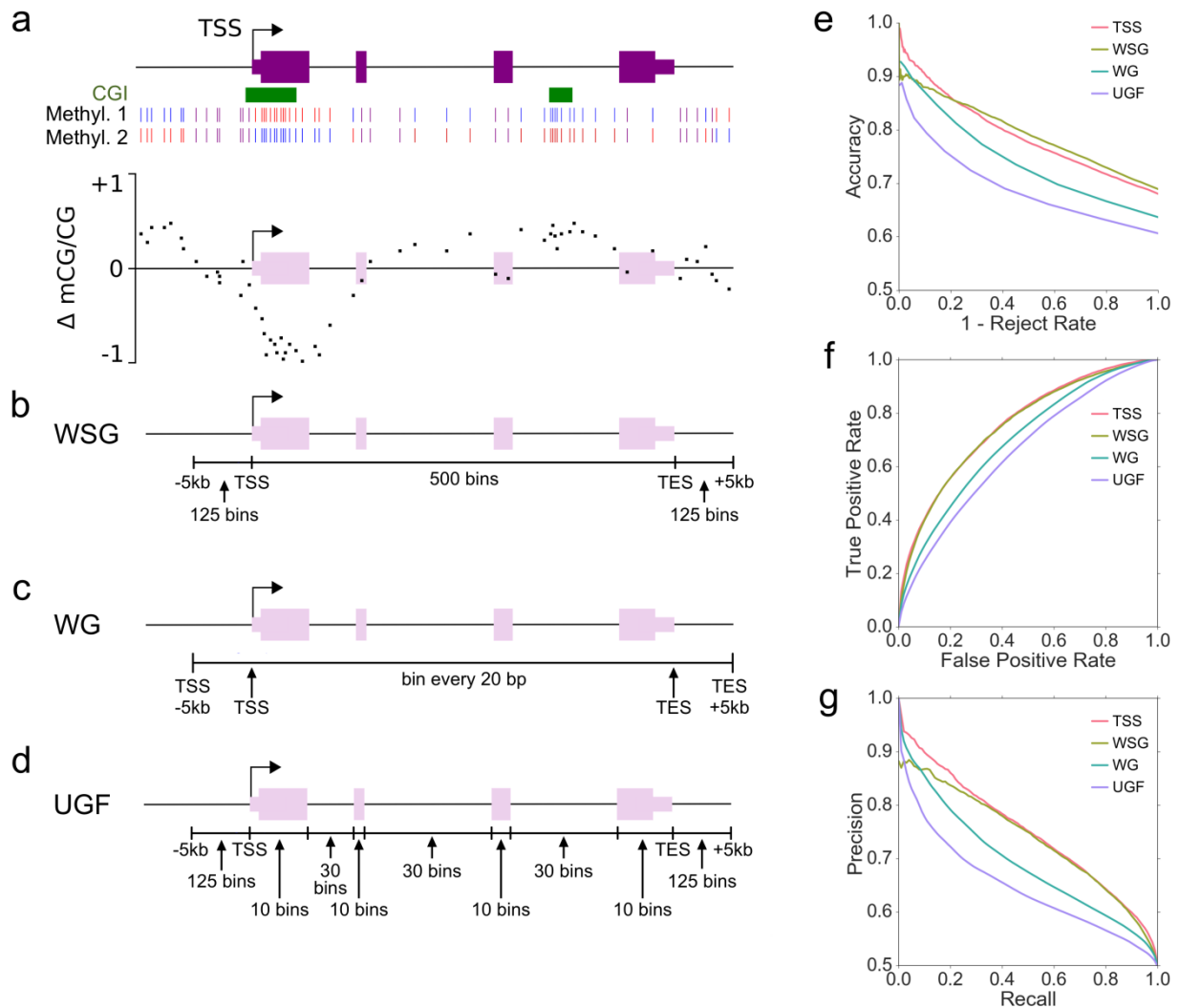
ID	Tissue	Tissue Short Name	Sources	Technical Replicates	
				WGBS	RNA-seq
E058	Penis Foreskin Keratinocyte	Keratin.	1	2	3
E065	Aorta	Aorta	1	6	2
E066	Adult Liver	Liver	3	2	2
E071	Brain Hippocampus Middle	Hippo.	2	3	2
E079	Esophagus	Esoph.	1	2	2
E094	Gastric	Gastric	1	5	3
E095	Left Ventricle	L. Ventr.	1	4	2
E096	Lung	Lung	1	2	2
E097	Ovary	Ovary	1	2	1
E098	Pancreas	Pancreas	1	2	2
E100	Psoas Muscle	Psoas	2	3	3
E104	Right Atrium	R. Atrium	1	3	1
E105	Right Ventricle	R. Ventr.	2	5	2
E106	Sigmoid Colon	Colon	2	2	3
E109	Small Intestine	Intest.	2	4	3
E112	Thymus	Thymus	1	2	1
E113	Spleen	Spleen	1	3	3

**Supplementary Table 2.** Differentially expressed and ME-Class interpolated gene counts from Roadmap Epigenomics Project (REP).

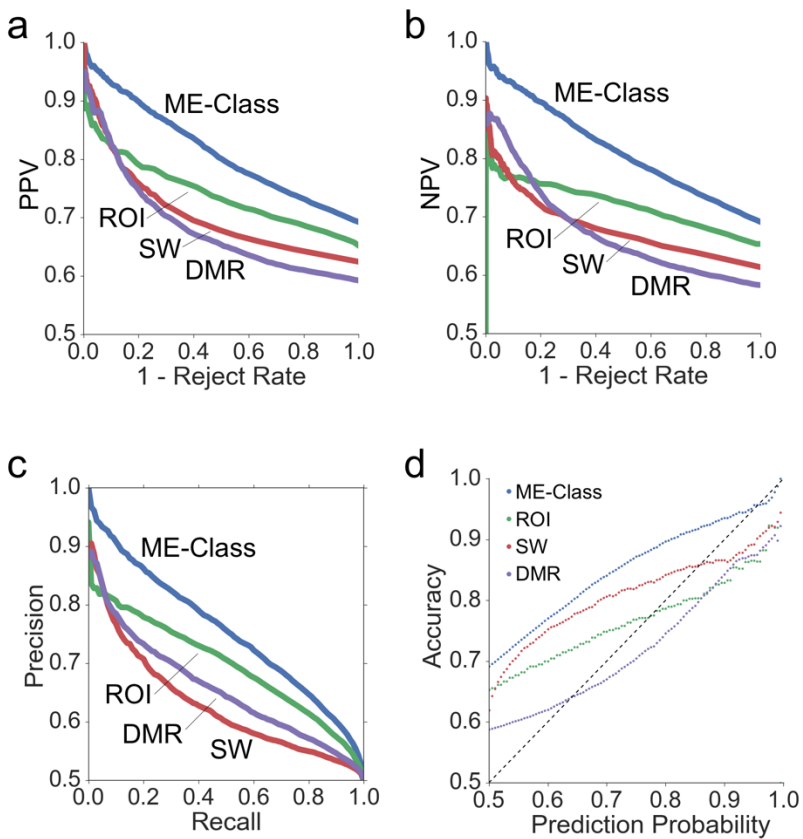
Tissue 1	Tissue 2	Diff. Expr. Genes	Interpolated Genes
Penis_Foreskin_Keratinocyte (E058)	Aorta (E065)	4596	3804
Aorta (E065)	Adult_Liver (E066)	3891	3186
Adult_Liver (E066)	Brain_Hippocampus_Middle (E071)	5020	4136
Brain_Hippocampus_Middle (E071)	Esophagus (E079)	4270	3484
Esophagus (E079)	Gastric (E094)	2801	2220
Gastric (E094)	Left_Ventricle (E095)	3438	2738
Left_Ventricle (E095)	Lung (E096)	3711	2942
Lung (E096)	Ovary (E097)	3227	2521
Ovary (E097)	Pancreas (E098)	3931	3148
Pancreas (E098)	Psoas_Muscle (E100)	5240	4285
Psoas_Muscle (E100)	Right_Atrium (E104)	3437	2854
Right_Atrium (E104)	Right_Ventricle (E105)	1128	860
Right_Ventricle (E105)	Sigmoid_Colon (E106)	3573	2906
Sigmoid_Colon (E106)	Small_Intestine (E109)	1121	874
Small_Intestine (E109)	Thymus (E112)	3899	3163
Thymus (E112)	Spleen (E113)	2068	1628
Spleen (E113)	Penis_Foreskin_Keratinocyte (E058)	4588	3711



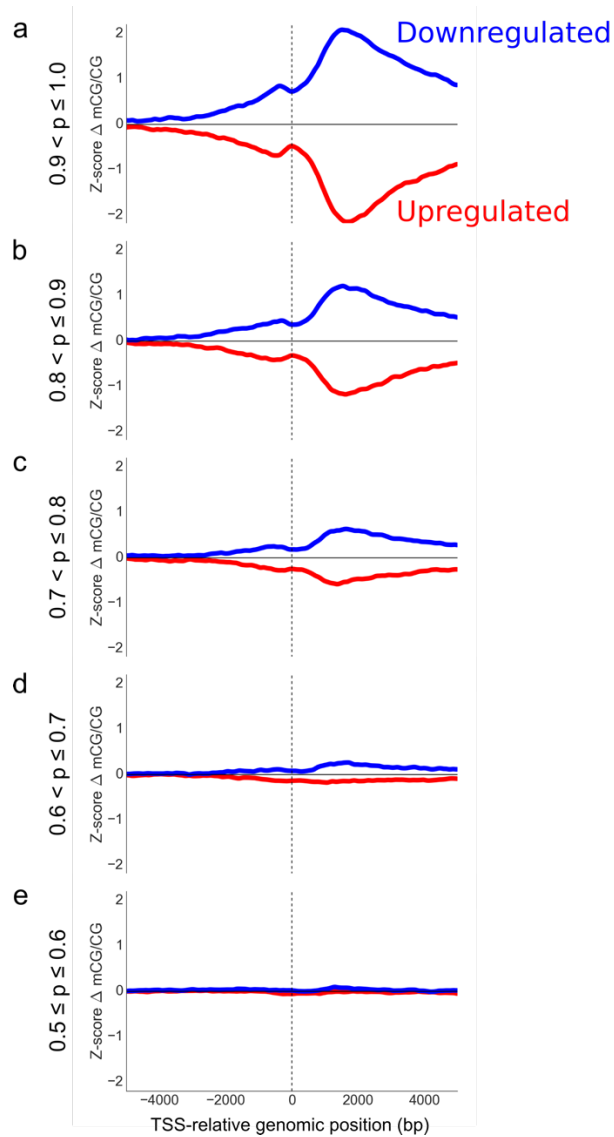
**Supplementary Figure 1.** Increasing the number of RF estimators from 100 to 1000 for the ROI classifier does not substantially increase performance as evaluated by: a) accuracy versus 1-reject rate, b) precision versus recall (PR AUC; 100 estimators: 0.70, 1000 estimators: 0.71), and c) ROC curve (ROC AUC; 100 estimators: 0.72, 1000 estimators: 0.73).



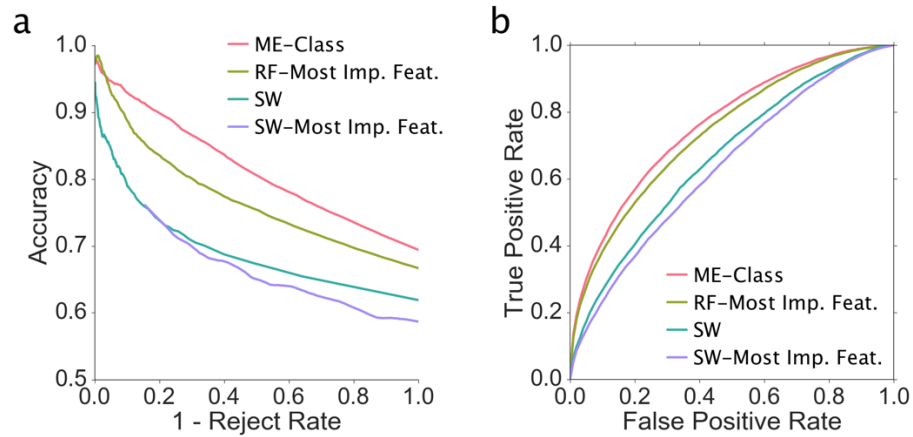
**Supplementary Figure 2.** Alternative full-gene methylation representations do not outperform TSS-centric representations. a) Heat map indicates methylation status at individual CpG sites – red is fully methylated, blue is fully unmethylated – for an example gene in two samples (Methyl. 1 and Methyl. 2). Individual points below indicate differential DNA methylation (Methyl. 2 – Methyl. 1) across the example gene at individual CpG sites. b) Whole Scaled Gene (WSG), c) Whole Gene (WG) and d) Uniform Gene Features (UGF) representation of the gene in (a). See additional description of each method in the Materials and Methods. Performance plots of TSS, WSG, WG, and UGF as reported by: e) accuracy versus 1-reject, f) ROC curve (ROC AUC; TSS: 0.76, WSG: 0.75, WG: 0.70, UGF: 0.65), and g) precision versus recall (PR AUC; TSS: 0.75, WSG: 0.74, WG: 0.69, UGF: 0.65). CGI = CpG island.



**Supplementary Figure 3.** Additional evaluation metrics for each method using 17 REP tissue differential samples: a) positive predictive value (PPV) versus 1- reject rate, b) negative predictive value (NPV) versus 1- reject rate, c) precision versus recall (PR AUC; ME-Class: 0.75, ROI: 0.70, DMR: 0.63, SW: 0.66) and d) accuracy versus the classifier's probability of prediction.

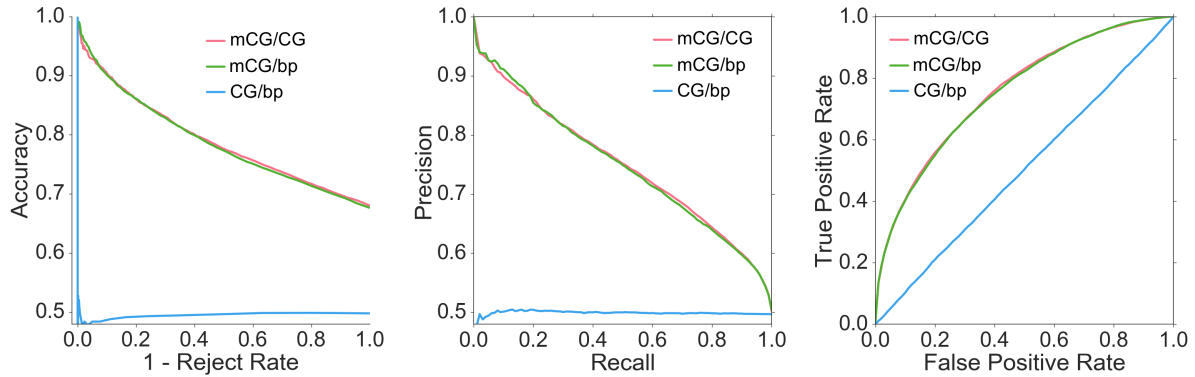


**Supplementary Figure 4:** Metagene plots of genes identified by ME-Class in REP data at different probabilities of prediction  $p$ . Blue curves represent the average Z-score normalized methylation difference between each sample for downregulated genes while red curves represent the average for upregulated genes.

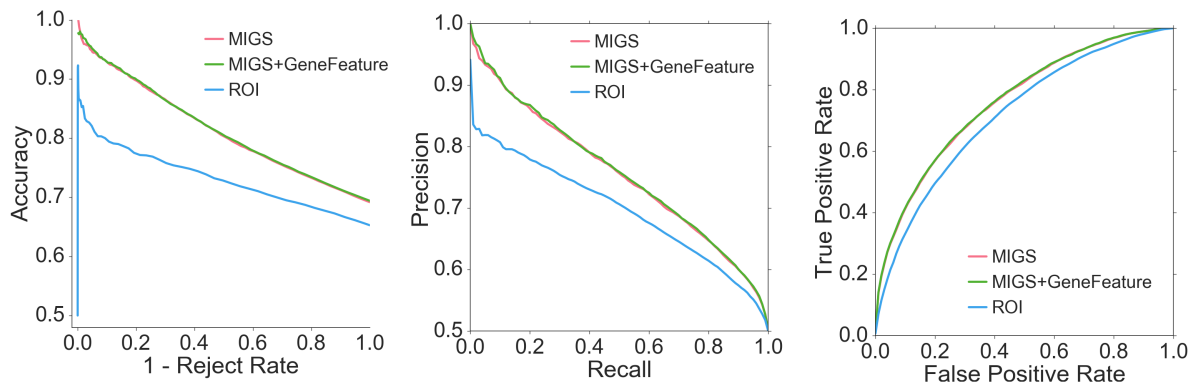


**Supplementary Figure 5.** ME-Class outperforms classifiers using REP data based on only the most important methylation features, [+0.5kb, +2.5kb] around the TSS, as evaluated by: a) accuracy versus 1-reject rate, and b) ROC curve (ROC AUC; ME-Class: 0.76, RF-Most Imp. Feat.: 0.74, SW: 0.67, SW-Most Imp. Feat.: 0.64). RF-Most Imp. Feature is an ME-Class like classifier built using features from only the region [+0.5kb, +2.5kb] around the TSS. SW-Most Imp. Feat. is similar to the SW approach, but only using methylation from [+0.5kb, +2.5kb] around the TSS.

### a Methylation density

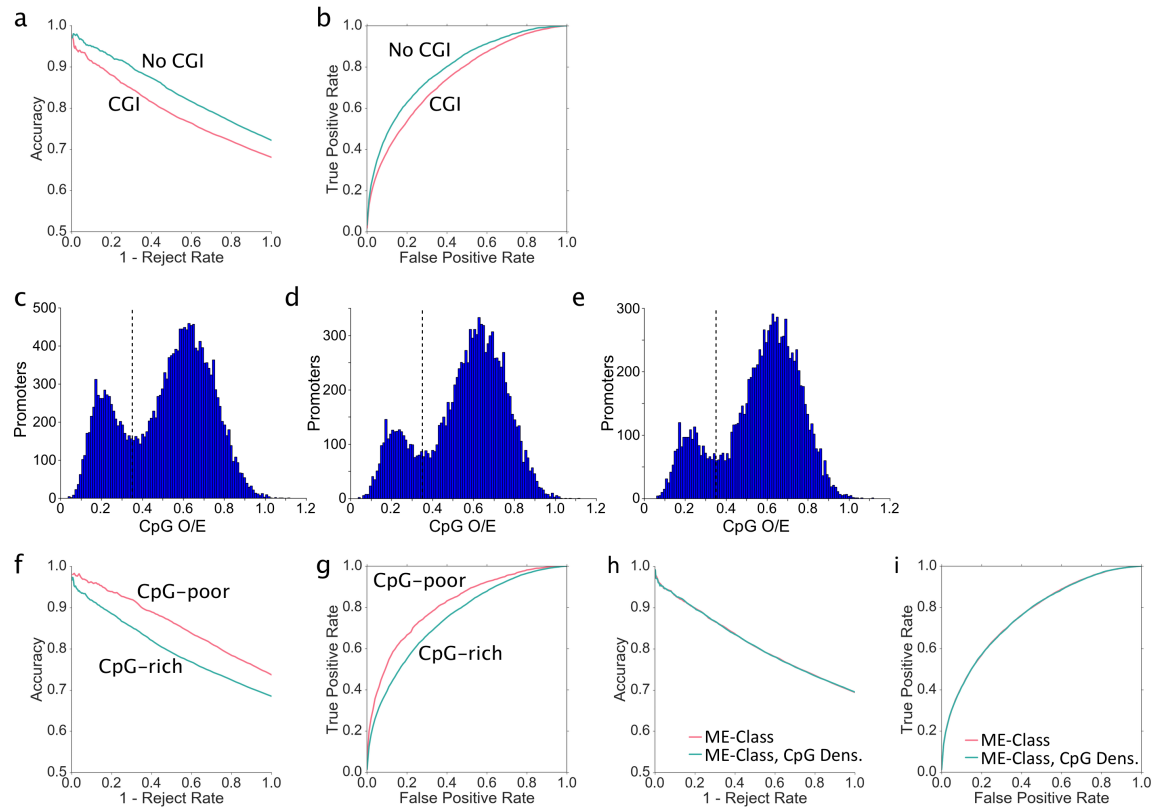


### b Gene features

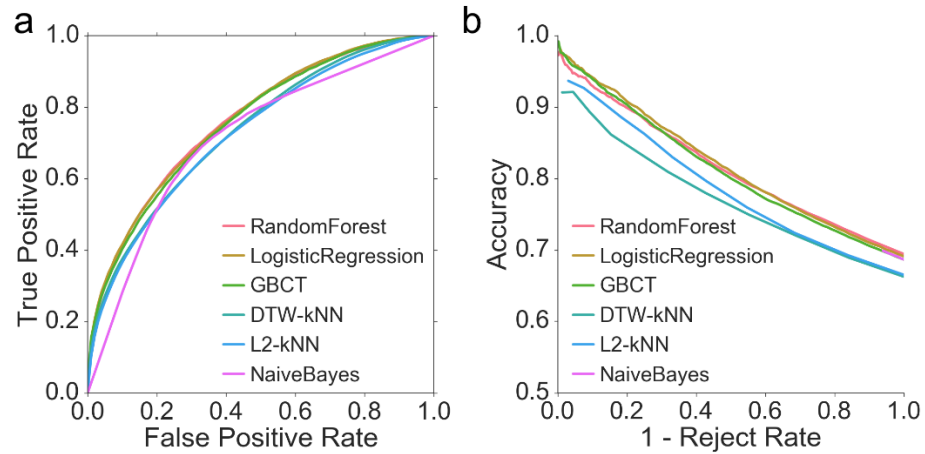


**Supplementary Figure 6.** The addition of methylated CpG density and gene body features (GF) does not increase ME-Class performance. a) Performance plots of ME-Class altered to use either mCG/CG, mCG/bp, or CpG density (200bp resolution, CG/bp) as input. (PR AUC; mCG/CG: 0.75, mCG/bp: 0.75, CG/bp: 0.50; ROC AUC; mCG/CG: 0.75, mCG/bp: 0.75, CG/bp: 0.50) b) Performance plots of ME-Class with and without adding gene body features (GF) from the ROI classifier including average internal exons, introns, and downstream features. ROI features are in Fig. 1d. (PR AUC; ME-Class: 0.75, ME-Class+GF: 0.76, ROI: 0.70; ROC AUC; ME-Class: 0.76, ME-Class+GF: 0.76, ROI: 0.72).



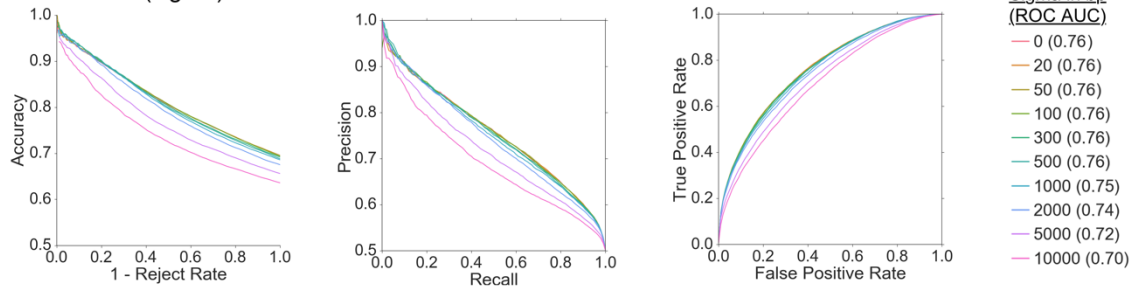


**Supplementary Figure 7.** CpG-poor genes are more predictive of expression classification. ME-Class performance for genes overlapping a CpG Island (CGI) by  $\geq 1$ bp is reported as: a) accuracy versus 1-rejection rate and b) ROC curve analysis (ROC AUC: No-CGI: 0.79; CGI: 0.75) c) Histogram of all genes with complete start and stop annotation according to RefSeq (n=19,175). Low CpG density genes comprise 26.0% (4,977 genes) while high CpG density genes comprise 74.0% (14,198 genes). d) Histogram of differentially expressed RefSeq genes (n=12,064 genes), where low CpG density genes comprise 18.8% (2,265 genes) while high CpG density genes comprise 81.2% (9,799 genes). e) Histogram of differentially expressed, interpolated RefSeq genes (n=10,524 genes) after applying our filtering parameters (see Materials and Methods). Low CpG density genes comprise 17.5% (1,842 genes) while high CpG density genes comprise 82.5% (8,681 genes). ME-Class performance is reported as: f) accuracy versus 1-rejection rate and g) ROC curve analysis (ROC AUC: CpG-poor: 0.8; CpG-rich: 0.75) Cutoff between low and high CpG density genes at 0.35 observed/expected normalized CpGs  $\pm 1500$ bp of TSS. ME-Class performance with or without added feature of observed/expected normalized CpG density  $\pm 1500$ bp of TSS is reported as h) accuracy versus 1-rejection rate and i) ROC curve analysis (ROC AUC: ME-Class: 0.76; ME-Class, CpG Density: 0.76).

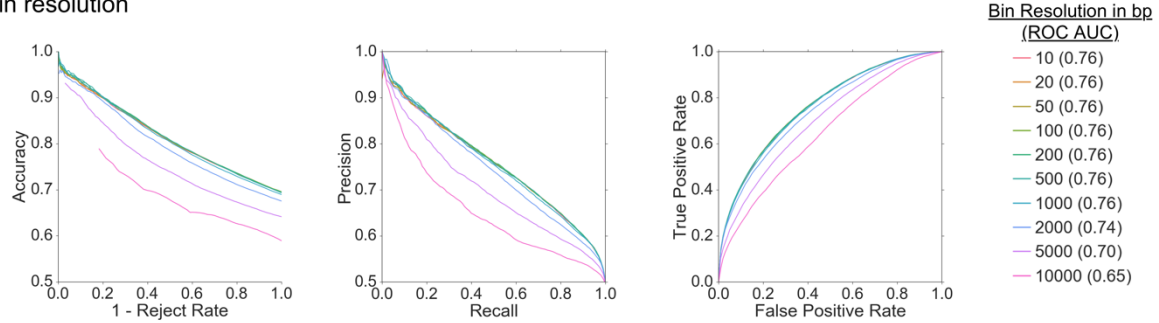


**Supplementary Figure 8.** Random Forest classifier performs similarly or outperforms alternatives based on classification performance as measured by a) ROC curve (ROC AUC; RF: 0.76, LR: 0.76, GBCT: 0.76, DTW-kNN: 0.73, L2-kNN: 0.73, Naïve Bayes: 0.71) and b) accuracy versus 1-reject rate. LR = Logistic Regression, GBCT= Gradient Boosted Classification Trees, DTW-kNN = Dynamic Time Warping based k-Nearest Neighbor, L2-kNN = Euclidean distance (L2) based k-Nearest Neighbor.

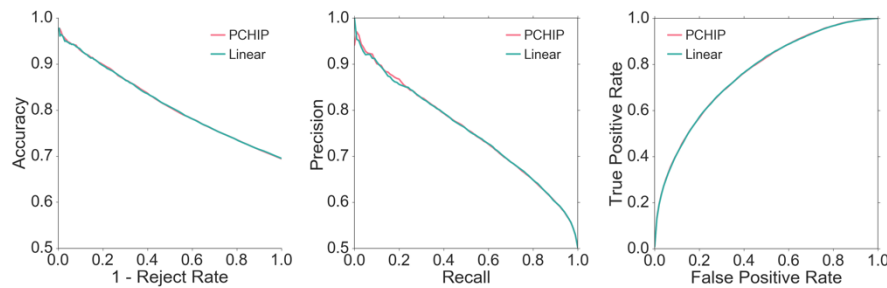
### a Gaussian filter (sigma)



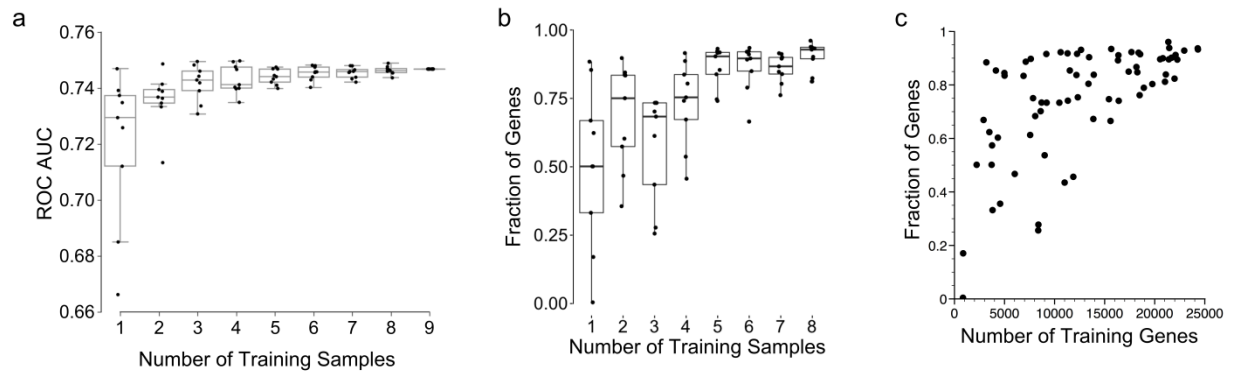
### b Bin resolution



### c Interpolation method

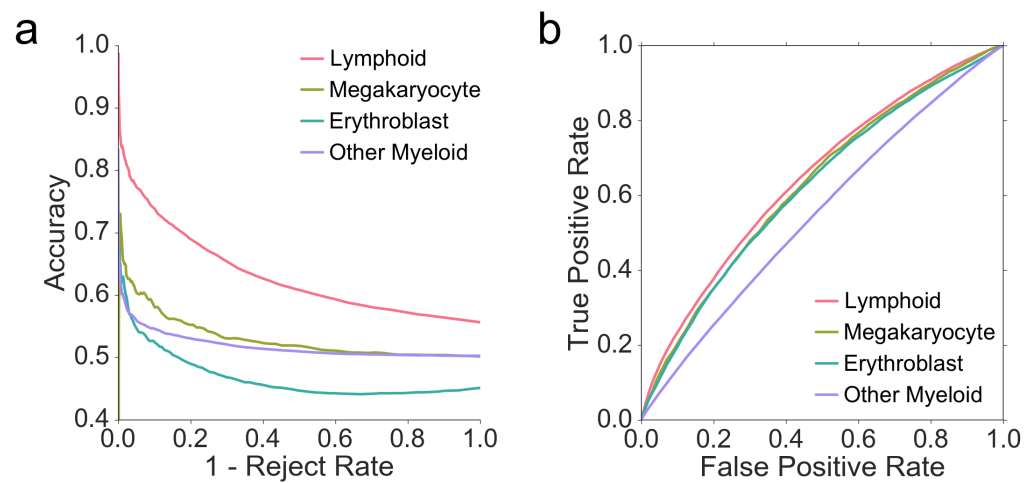


**Supplementary Figure 9.** Effect on ME-Class performance of tuning parameters for smoothing, bin resolution, and interpolation. Performance is reported as: accuracy versus 1-reject, precision versus recall, and ROC curve. a) Relationship between ME-Class performance and sigma for Gaussian smoothing with a constant bin resolution of 20bp. b) Relationship between ME-Class performance and the size of the bin resolution at a constant sigma of 50bp. c) Relationship between ME-Class performance and alternative interpolation method (PR AUC; PCHIP: 0.76, Linear: 0.76; ROC AUC; PCHIP: 0.76, Linear: 0.76).

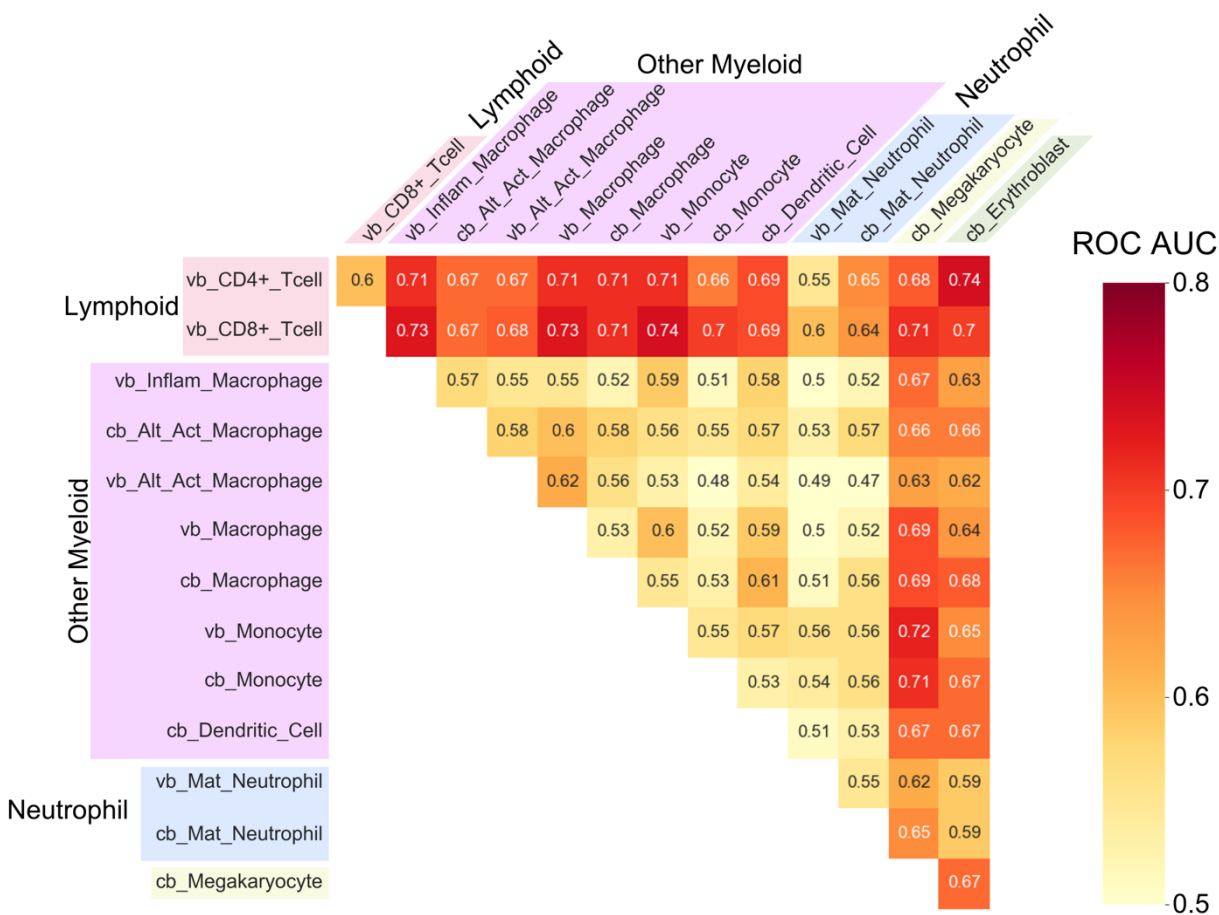


**Supplementary Figure 10.** Number of training samples and genes determine ME-Class performance.

The testing ROC AUC as a function of a) the number of training samples. The fraction of correctly identified genes using ME-Class with 9 evaluation samples as a function of b) the number of training samples and c) the total number of training genes from the training samples. In (a) and (b), each point indicates the performance across all genes in an individual sample comparison. In (c), each point indicates the number of training genes and fraction of genes returned for an individual sample comparison and set of training samples. Training genes in (c) means the number of genes summed across all training samples. A gene can be counted multiple times if it shows up in multiple samples. Although it will likely have different methylation profiles and expression values in each comparison. Permuted sets of all differential training samples ( $n=8$ ) and a fixed set of differential evaluation samples ( $n=9$ ) are randomly chosen from the REP dataset.



**Supplementary Figure 11.** Performance of Blueprint neutrophil samples in comparison to other hematopoietic cell types. ME-Class is trained from the full REP dataset and performance is reported as: a) accuracy versus 1- reject rate and b) ROC curve analysis (ROC AUC; Lymphoid: 0.65, Megakaryocyte: 0.63, Erythroblast: 0.62, Other Myeloid: 0.55).



**Supplementary Figure 12.** Performance of Blueprint Epigenome samples using a similar leave-one-out differential sample evaluation cross-validation framework as used for the REP data (see Fig. 1e). The performance of ME-Class trained and evaluated solely using Blueprint samples is similar to that of a ME-class model trained from the REP dataset. Shown are ROC AUC of differential comparisons of randomly chosen single samples of each of the 14 cell types from Blueprint dataset.