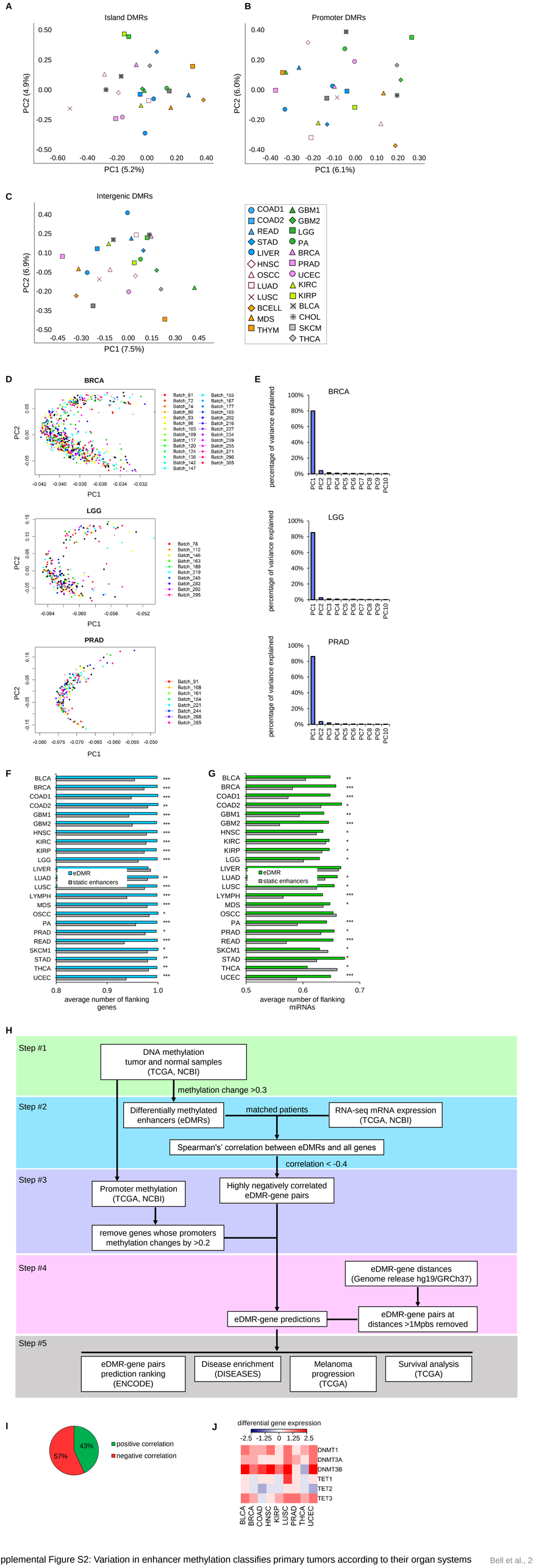

Supplemental Figure S2: Variation in enhancer methylation classifies primary tumors according to their organ systems

Supplemental Figure S2: Variation in enhancer methylation classifies primary tumors according to their organ systems

(A-C) Principal components analysis (PCA) of DMRs within (A) islands, (B) promoters, and (C) intergenic regions show no meaningful clustering of cancers, is based on the differential methylation levels of these genomic features (compare to Fig. 2B). First and second principal components (x-axis: PC1 and y-axis: PC2, respectively) account for 10.1%, 12.1%, and 14.4% of variation, respectively. (D) PCA to assess the batch effect of samples. Enhancer methylation changes were used to create each plot. PCA plots of (top) BRCA, (middle) LGG and (bottom) PRAD (TCGA samples, https://tcga-data.nci.nih.gov/tcga) show no separation between library preparation batches; each batch is colored similarly, each circle represents a patient sample. We used the top quartile of most variable eDMRs in this analysis. (E) Bar plots show the percent of variance explained by the top 10 principle components (PC1 to PC10) corresponding to the PCA plots shown in panel D. (F-G) Compared to enhancers that are not differentially methylated between normal and primary tumors ('static enhancers'), eDMRs are flanked by significantly more genes (F) and miRNAs (G). X-axes represent average number of genes or miRNAs within 1 Mbp of enhancers in each cancer dataset (binomial distribution, P<0.05, *; P<0.001, **, P<1e-6, ***; FDR corrected). (H) Flowchart summarizing the steps in our model for predicting genes that are potentially regulated by eDMR (eDMR-associated genes). Step #1, Green background: First, we calculated the methylation changes in enhancers and promoters between normal and cancerous samples, we retained only the eDMRs, and used the level of promoter differential methylation in the third step (see below Step #3, Purple background). Step #2, Cyan background: Second, matched patients were used to compute Spearman's correlations between methylation patterns of the eDMRs and expression patterns of all genes and miRNAs. Highly correlated eDMR-genes or eDMR-miRNAs pairs were retained (negative correlation below -0.4). Step #3, Purple background: Third, we filtered out genes for which promoter differential methylation was greater 0.2, enriching our set for genes with variation due to enhancer, rather than by promoter, methylation changes; this step was performed only for predicted eDMR-associated genes and not for predicted eDMR-associated miRNAs, since the annotation for miRNA promoters is incomplete. Step #4, Pink background: Fourth, we selected eDMR-gene pairs (or eDMR-miRNA pairs) located on the same chromosome, and with a maximal distance of 1 Mbp between the center of the enhancer and the transcription start site (for eDMR-associated genes) or the start position of the pre-miRNA sequences (for eDMR-associated miRNAs). Step #5, Grey background: Fifth, eDMR-gene predictions were subjected to post-prediction analyses: (leftmost box) ranking of the eDMR-gene pair predictions (see Supplemental Table S4), (middle left box) enrichment analysis of eDMR-associated genes for diseases (see Fig. 2E); (middle right box) study of the contribution of eDMR-gene pairs to metastasis using melanoma as a model (see Fig. 3-5); (rightmost box) survival analyses correlating eDMRs and expression of their paired genes with survival of patients with metastatic melanoma (see Fig. 4-5). (I) Correlation between predicted pairs of eDMRs and associated genes is more frequently negative (Spearman's correlations). (J) Heatmap shows that DMNT3B expression is upregulated in most cancers. Colors depict differential expression of DMNT family genes encoding methylases and TET family genes encoding demethylases.