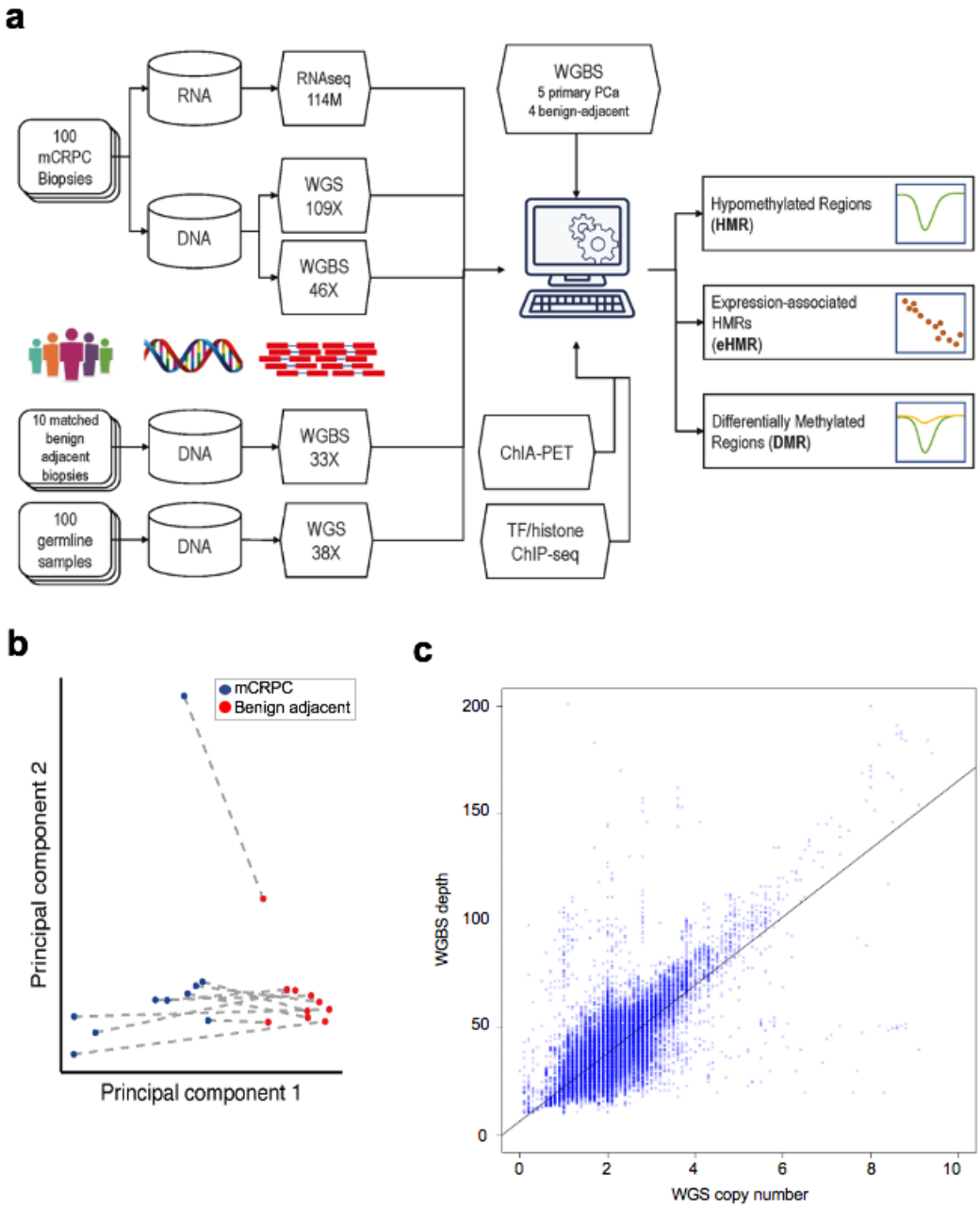
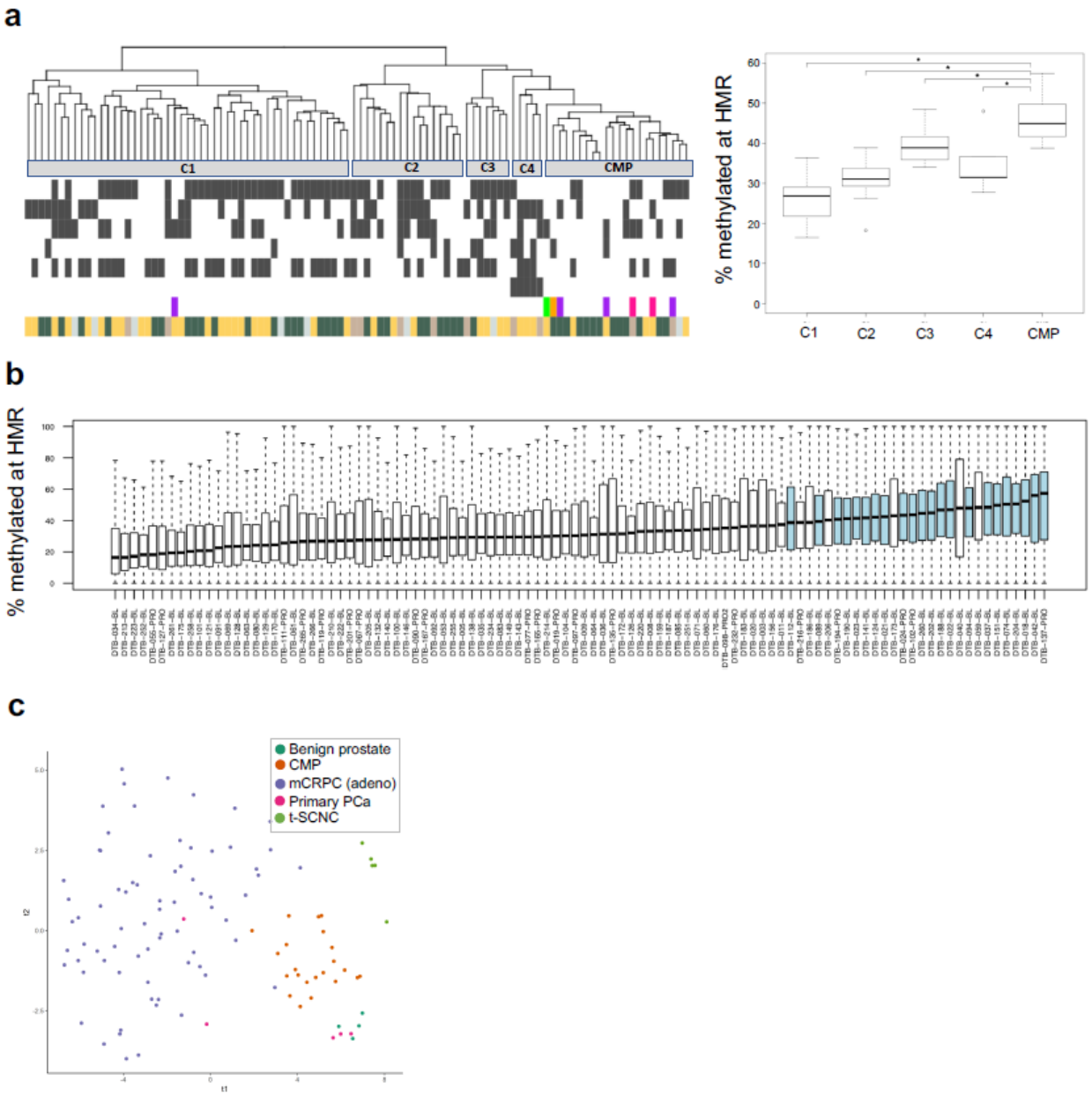
SUPPLEMENTAL FIGURES (EXTENDED DATA)

Supplemental Figure 1



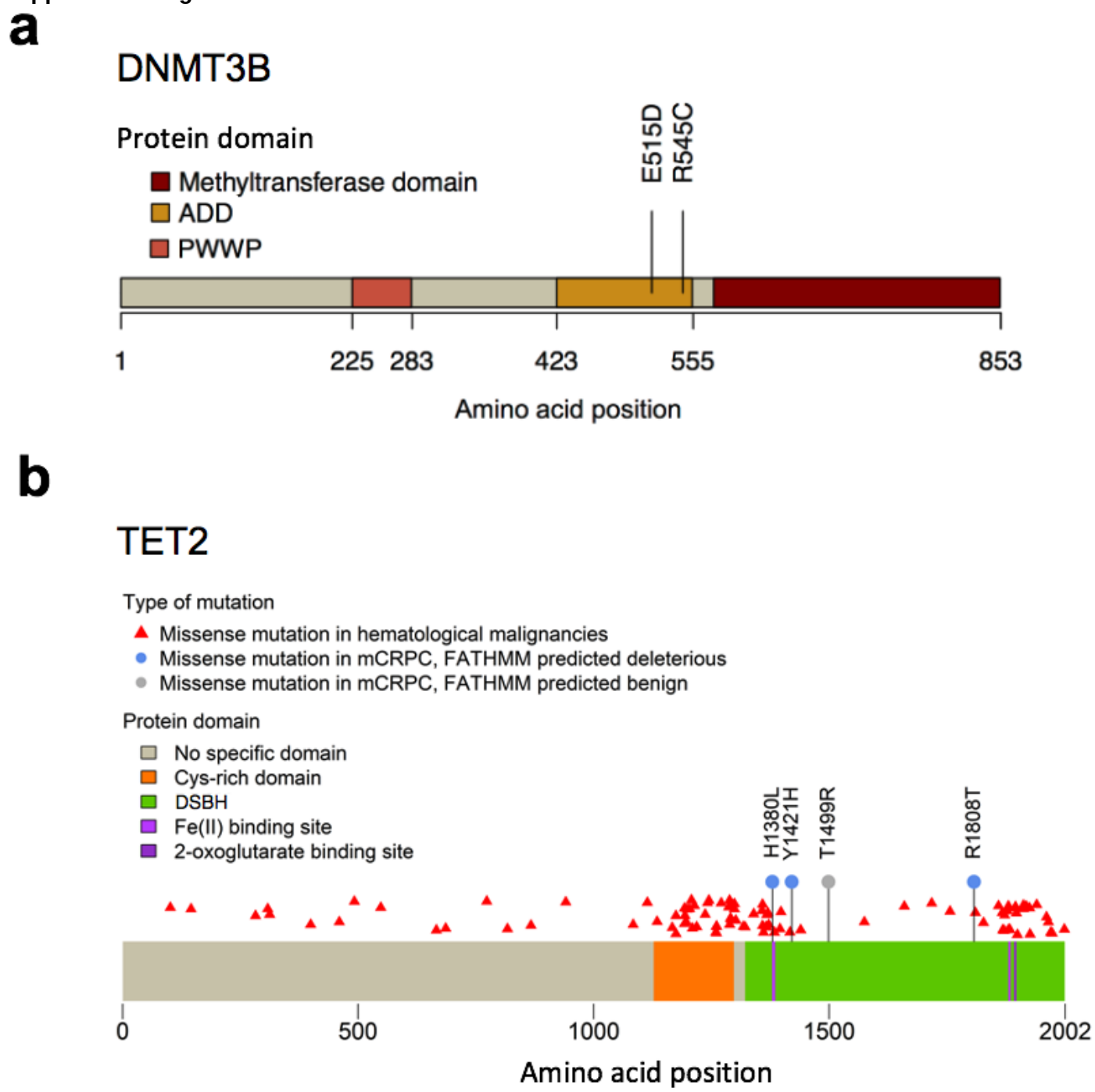
Supplemental Figure 1: Overview of study design and quality control
a) Flowchart of the study design. **b)** Principal component analysis of CpG methylation at all recurrent HMRs in the 10 paired mCRPC tumor and benign-adjacent tissue. Dotted grey lines represent distance between each pair of samples in the principal component space. The percent of variance explained by principal components 1 and 2 are 46.1% and 9.5% respectively. **c)** To check consistency of DNA samples between WGS and WGBS experiments, we plotted the copy number estimates (a measure of depth) derived from the WGS and the depth derived from the WGBS sequencing of the same DNA samples across the entire genomes of all 100 samples. Data were highly concordant in copy number estimates (Spearman's Correlation = 0.66). Median copy number in 3 Mb windows tiled along the genome was calculated from WGS data as described in (Quigley Cell 2018). Median WGBS sequencing depth at the same 3 Mb windows was calculated from Bismark calls at CpG islands.

Supplemental Figure 2



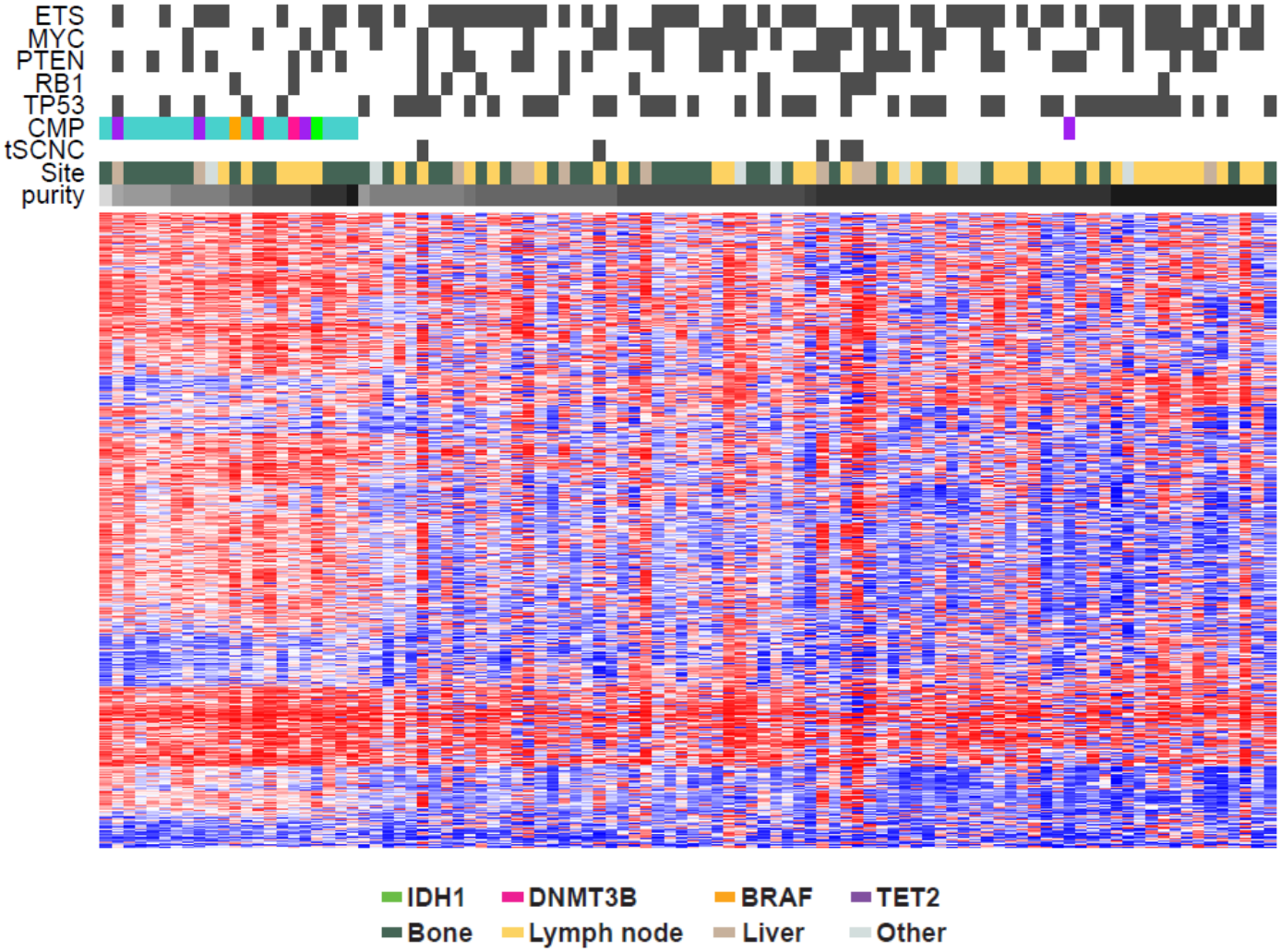
Supplemental Figure 2: CMP tumors harbor elevated methylation. a) Boxplots showing the average methylation levels at HMRs by cluster. Tumors in the CMP cluster have significantly higher methylation levels when compared to all other clusters. N's: C1=49 (vs. CMP, $P < 2.2x10^{-16}$), C2=17 (vs. CMP, $P = 8.048x10^{-14}$), C3=7 (vs. CMP, $P = 0.008$), C4=5 (vs. CMP, $P = 0.008$), CMP=22 (asterisk indicates $P < 0.05$, two-sided Wilcoxon test). **b)** Boxplots show the median, first, and third quartiles, and outliers are shown if outside 1.5x the inter-quartile range for methylation levels at HMR for each tumor. Non-CMP or CMP samples with white and blue background, respectively. **c)** t-SNE dimensionality reduction plot of the mCRPC samples as well as the publicly available primary prostate cancer and benign prostate samples based on methylation at all recurrent HMRs.

Supplemental Figure 3



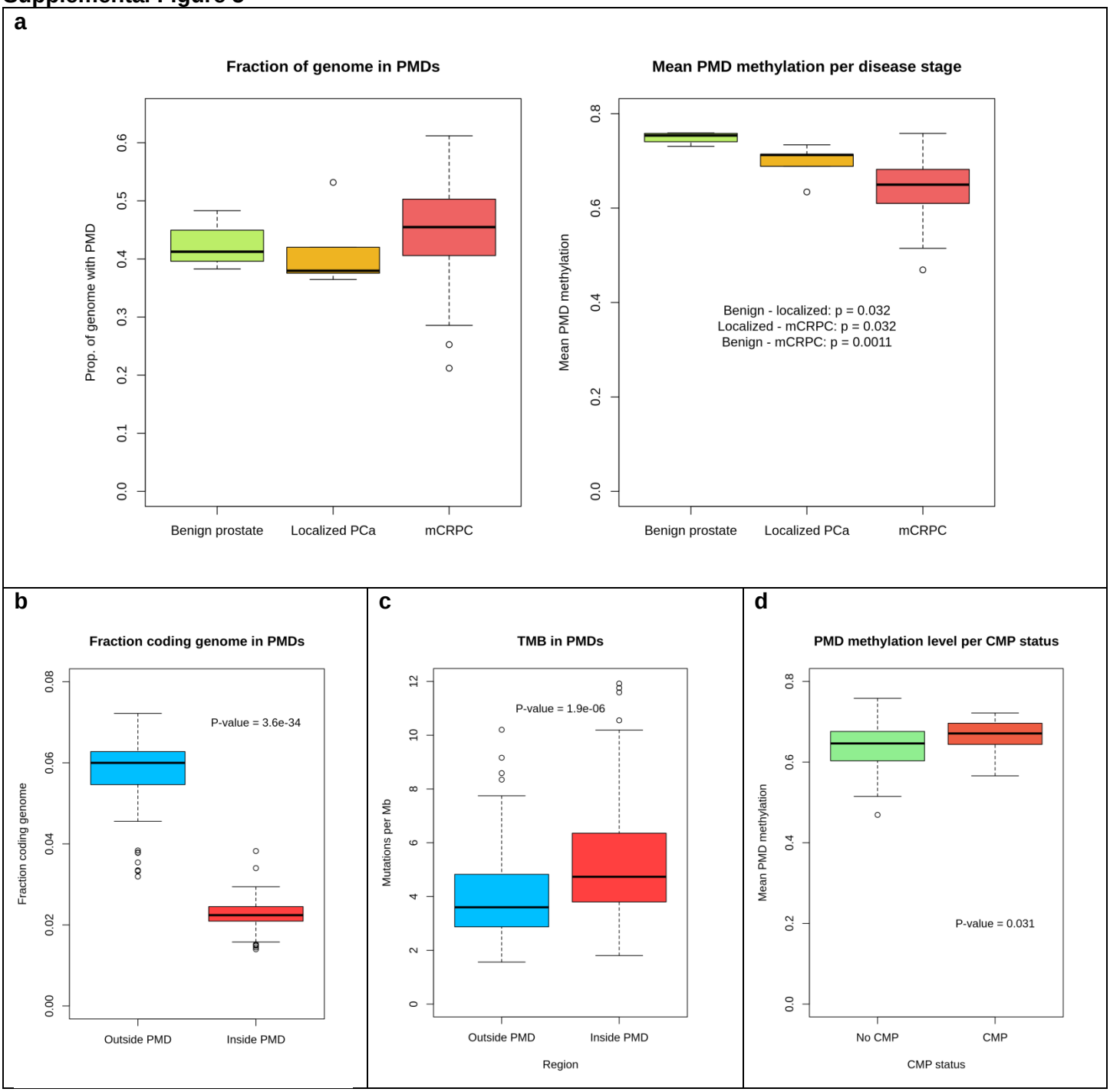
Supplemental Figure 3: Mutations observed in DNMT3B and TET2
a) Lollipop plot depicting the protein domains of DNMT3B and the two somatic missense mutations in our cohort. **b)** Lollipop plot depicting the protein domains of TET2 and the four somatic missense mutations in our cohort. Three of the four are associated with the CMP subtype, which are the 3 that fall near hotspots of missense mutations in leukemias, and the three that are computationally predicted by FATHMM to be deleterious.

Supplemental Figure 4



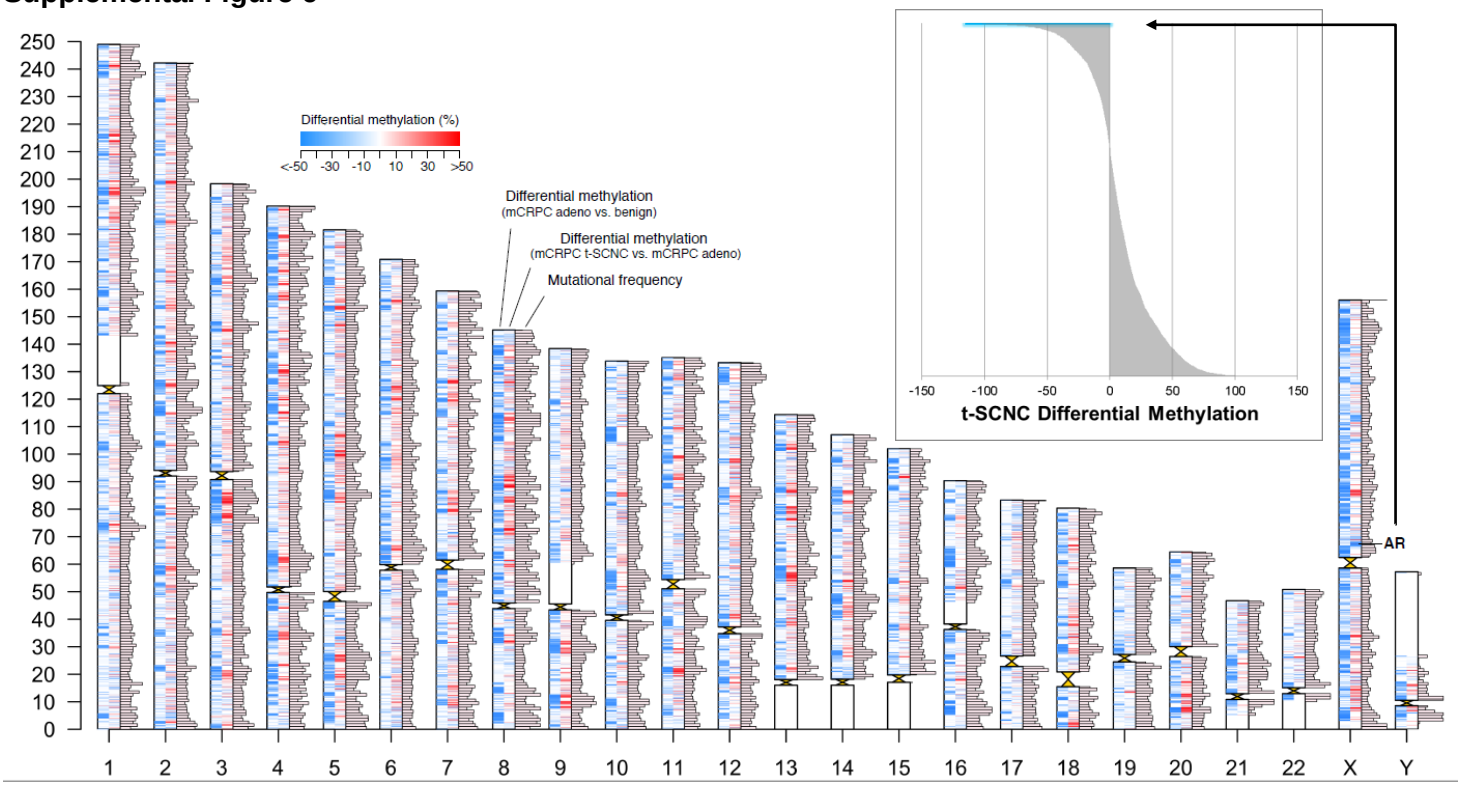
Supplemental Figure 4: Heatmap of the 10% most variable (based on methylation value standard deviation) recurrent HMRs in the mCRPC samples (the same loci as in Figure 1B). The samples are first grouped into CMP and non-CMP, and then within these groups, sorted by the histologic tumor purity. White=0% and black=100% tumor purity. Both the CMP and non-CMP tumors have a wide range of tumor purities, and the pattern of methylation within each group does not correlate with the degree of tumor purity.

Supplemental Figure 5



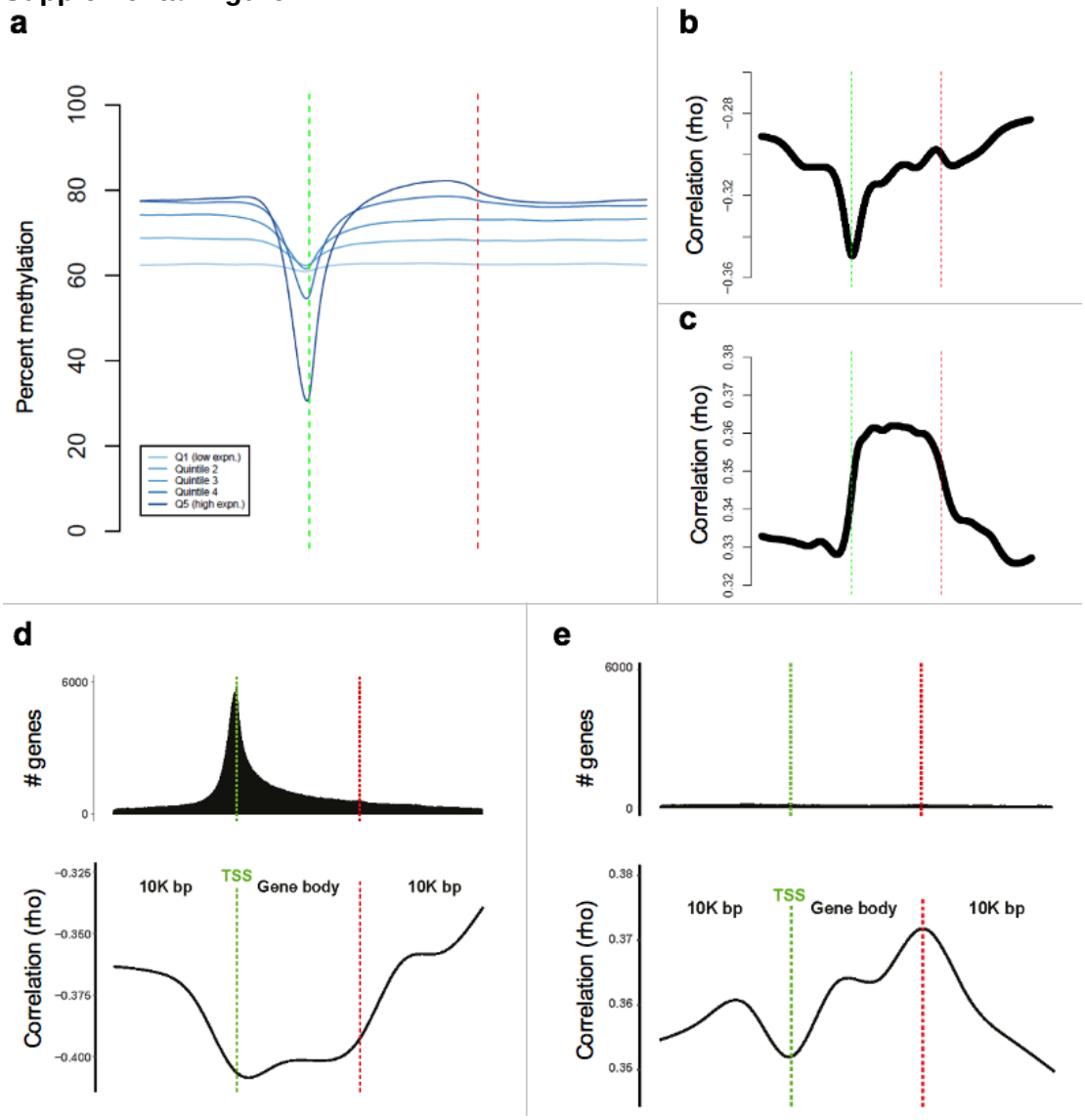
Supplemental Figure 5: Partially Methylated Domains (PMDs) identified using MethylseekR. a) Left: the fraction of the genome covered by PMDs did not differ between benign prostate (N=4), localized prostate cancer (N=5), and mCRPC (N=100). Right: the degree of methylation within PMDs was significantly lower in localized cancer (N=5) vs. benign prostate (N=4) and mCRPC (N=100) vs. localized cancer (N=5), two-sided Wilcoxon test. **b)** The fraction of bases inside PMDs (N=100) encoding exons is significantly lower than outside PMD regions (N=100). **c)** The mutational density (tumor mutational burden (TMB), calculated as mutations per Mb) is significantly lower outside of PMD regions. The two hyper-mutated samples are excluded from this analysis so the N=98 in both groups. **d)** The level of methylation in PMDs is higher in the CMP subtype (N=22) than the non-CMP mCRPC samples (N=78). A two-sided Wilcoxon test was used for all three comparisons. Boxplots show the median, first, and third quartiles, and outliers are shown if outside 1.5x the inter-quartile range.

Supplemental Figure 6



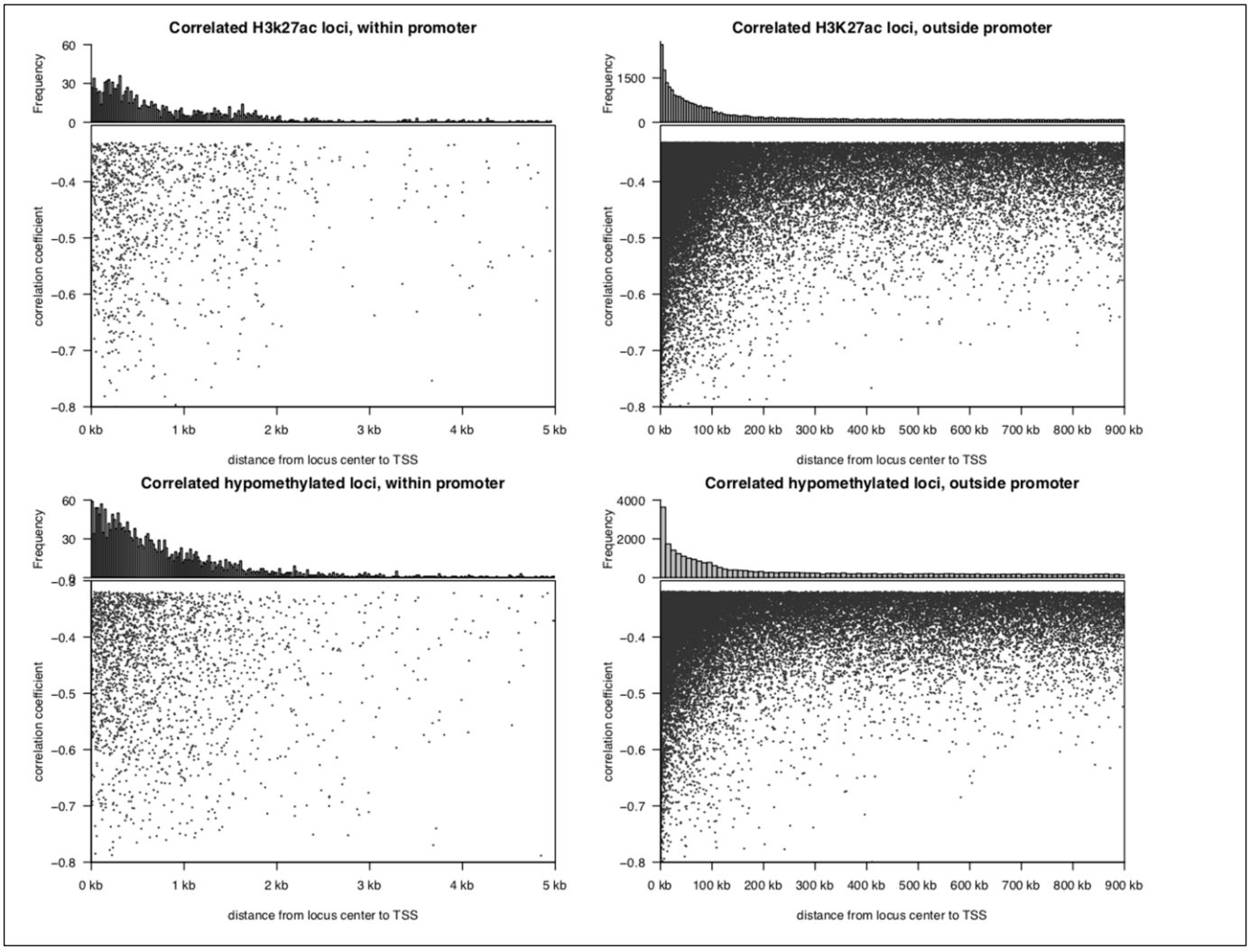
Supplemental Figure 6: Differential methylation comparing t-SCNC with mCRPC was performed using DSS. Differential methylation comparing benign prostate (N=4) and mCRPC adenocarcinoma (N=95) is shown as a control, and demonstrates the expected global hypo-methylation (blue). The comparison of t-SCNC (N=5) samples to adenocarcinoma (N=95) demonstrates that t-SCNC tumors have hyper-methylation at many regions (red). However, the *AR* locus on chromosome X harbors a broad region of hypo-methylation (blue). To formally assess this phenomenon, we binned the genome into 1 Mbp regions, calculated the sum of the DMRs within them, and found that the region containing *AR* was the top differentially hypo-methylated region genome-wide (inset).

Supplemental Figure 7



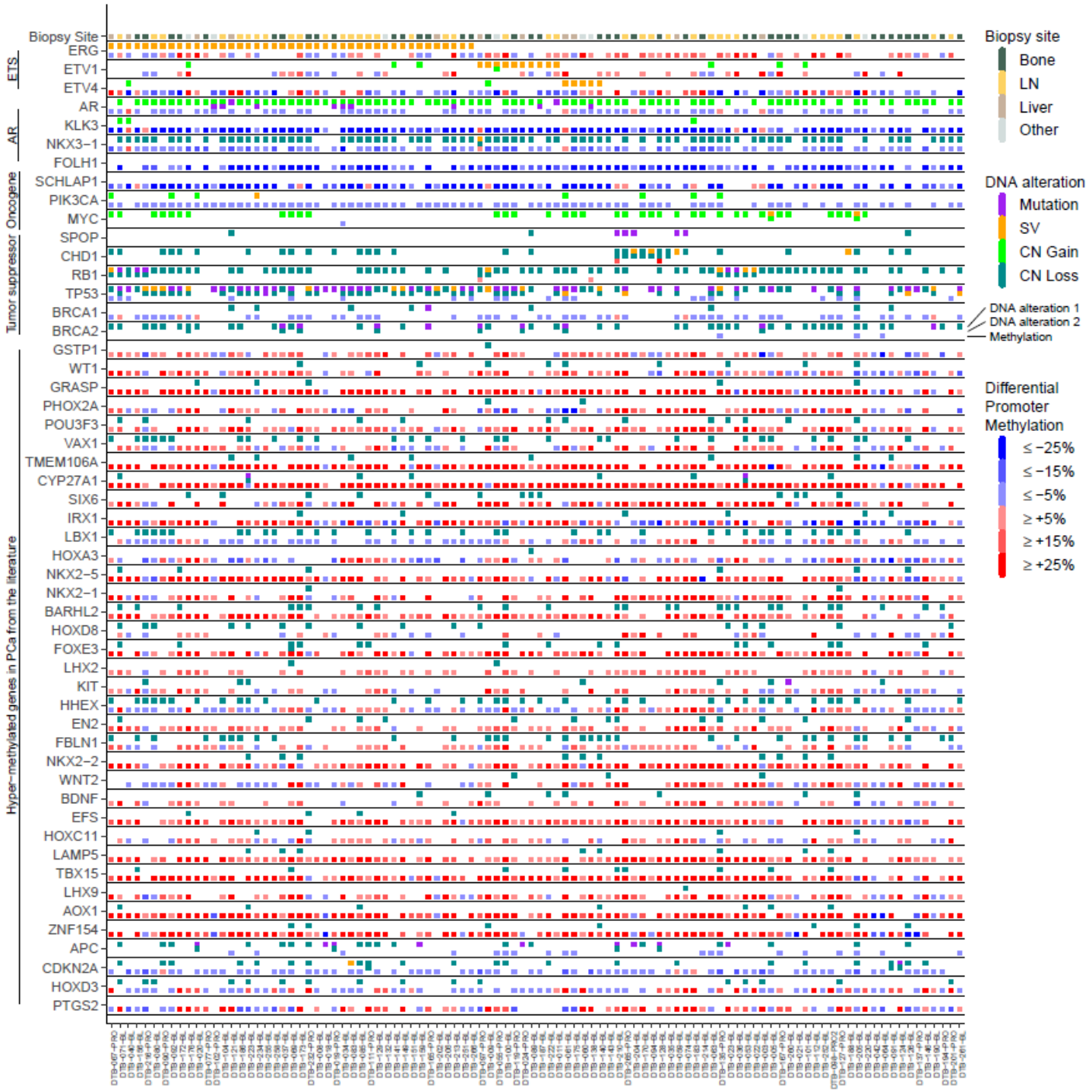
Supplemental Figure 7: eHMR strength across promoter and gene body regions. a) For each gene, CpG methylation within the gene body and 10kb upstream and downstream flanking region were considered. Methylation values were smoothed on a per-gene basis using the R smooth.spline function and then averaged across all genes, with genes stratified into quintiles based on mean gene expression in mCRPC. The x-axis represents normalized genomic position (relative to the transcription start site in green and the gene terminator in red).Spearman's correlation between CpG methylation within the gene body and 10kb upstream and downstream flanking region and gene expression was calculated for all genes. **b,c)** CpG's with a significant correlation are plotted, with negative correlation (b) and positive correlation (c). A spline was fit across all genes in aggregate. **d,e)** Top: density of eHMRs in which methylation was significantly (d) negatively and (e) positively correlated with gene expression. The width of eHMR segments overlapping gene bodies was scaled such that the width of the portion of the eHMR overlapping the gene body was proportional to the width of the gene of interest. Bottom: mean Spearman's correlation values of eHMRs in which methylation was significantly (d) negatively and (e) positively correlated with gene expression. The width of eHMR segments overlapping gene bodies was scaled as described above, and a spline was fit to the Spearman correlation values of all eHMR segments within 10Kbp of their respective genes. The green line represents the transcription start site (TSS) and the red line represents the transcription terminator. For all panels, N=100 independent mCRPC samples, and a two-sided Spearman's correlation was used with a cutoff of FDR ≤0.05.

Supplemental Figure 8



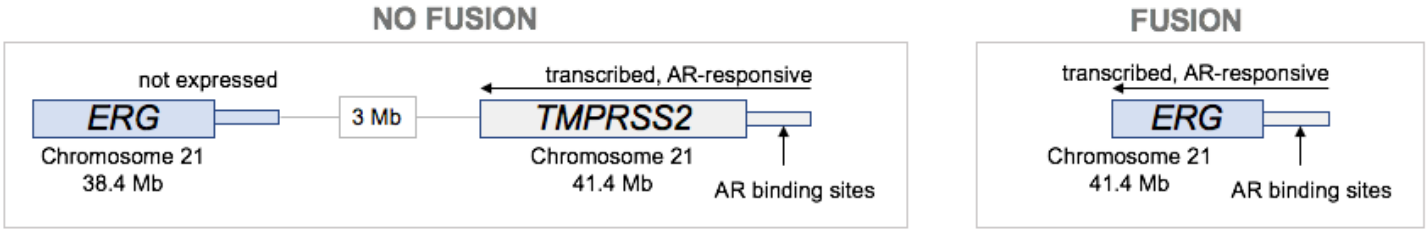
Supplemental Figure 8: Methylation associated with gene expression across local genomic regions. Plots show strength of correlation between % methylation and gene expression in candidate enhancer (left) or rHMR (right) loci. Correlation coefficient on Y axis, limited to negative correlations, vs. distance between gene TSS and the mid-point of the methylated region (X axis). Candidate enhancer regions defined by H3K27ac peaks and rHMR within 900 kb upstream or 100 kb downstream of a given gene were defined as described in the methods. Plots shown in both the full 1 megabase window around the TSS (upper) and a 10 kb window around the TSS (lower) for each of 58,381 transcripts. For all panels, N=100 independent mCRPC samples, and a two-sided Spearman's correlation was used with a cutoff of FDR ≤0.05.

Supplemental Figure 9



Supplemental Figure 9: Sample-level summary of mutations, copy number alterations, structural variants, and promoter methylation (in recurrent HMRs) for key PCa genes and for genes reported in the literature to be hyper-methylated in PCa. Promoter methylation values represent the difference between the methylation levels in our mCRPC samples versus the mean in benign prostate tissue. Negative and positive values represent hypo-methylation and hyper-methylation in mCRPC, respectively.

Supplemental Figure 10



Supplemental Figure 10: Schematic showing how methylation at AR binding sites in the TMPRSS2 promoter can modulate ERG expression in the TMPRSS2-ERG fusion tumors.