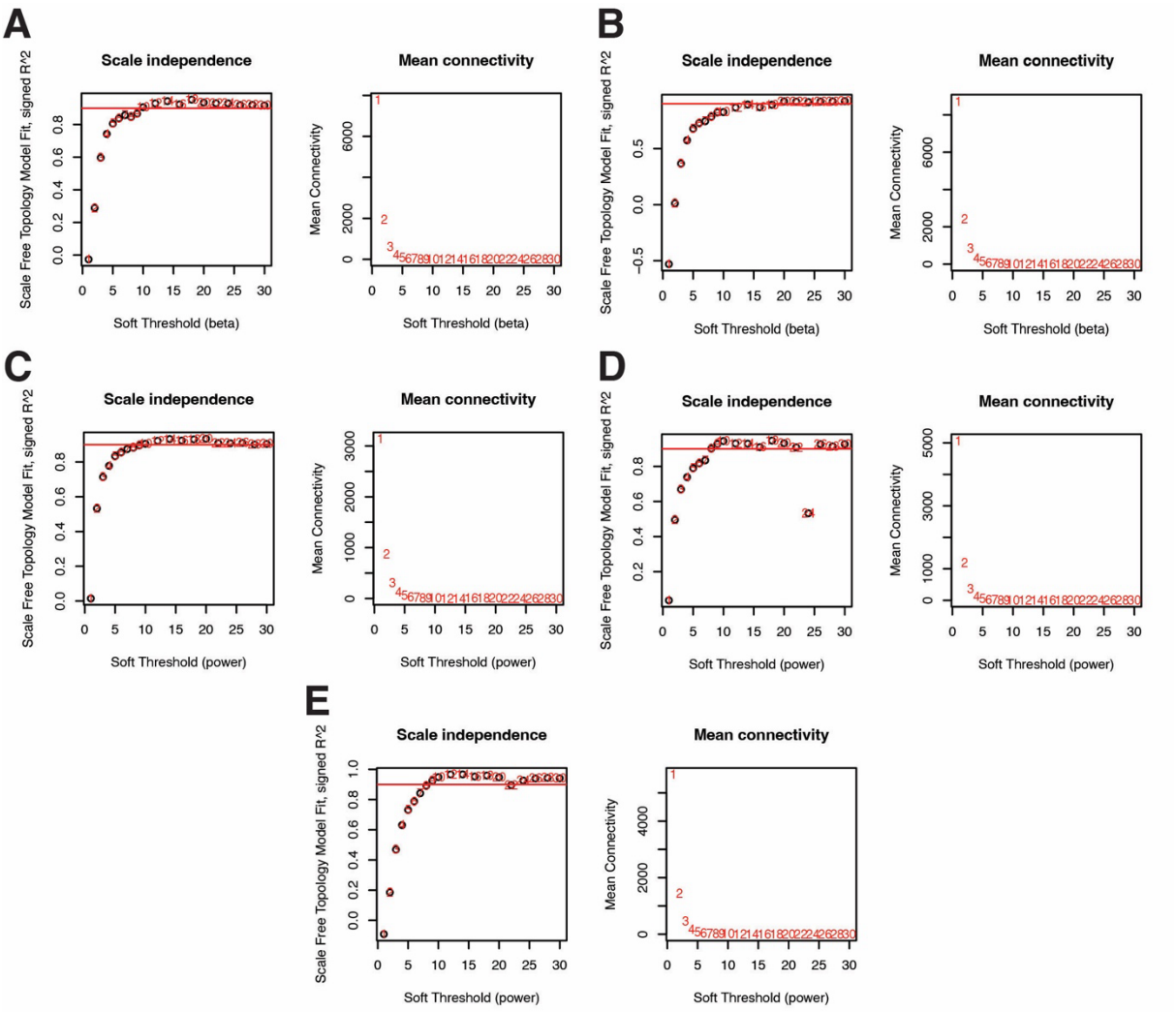
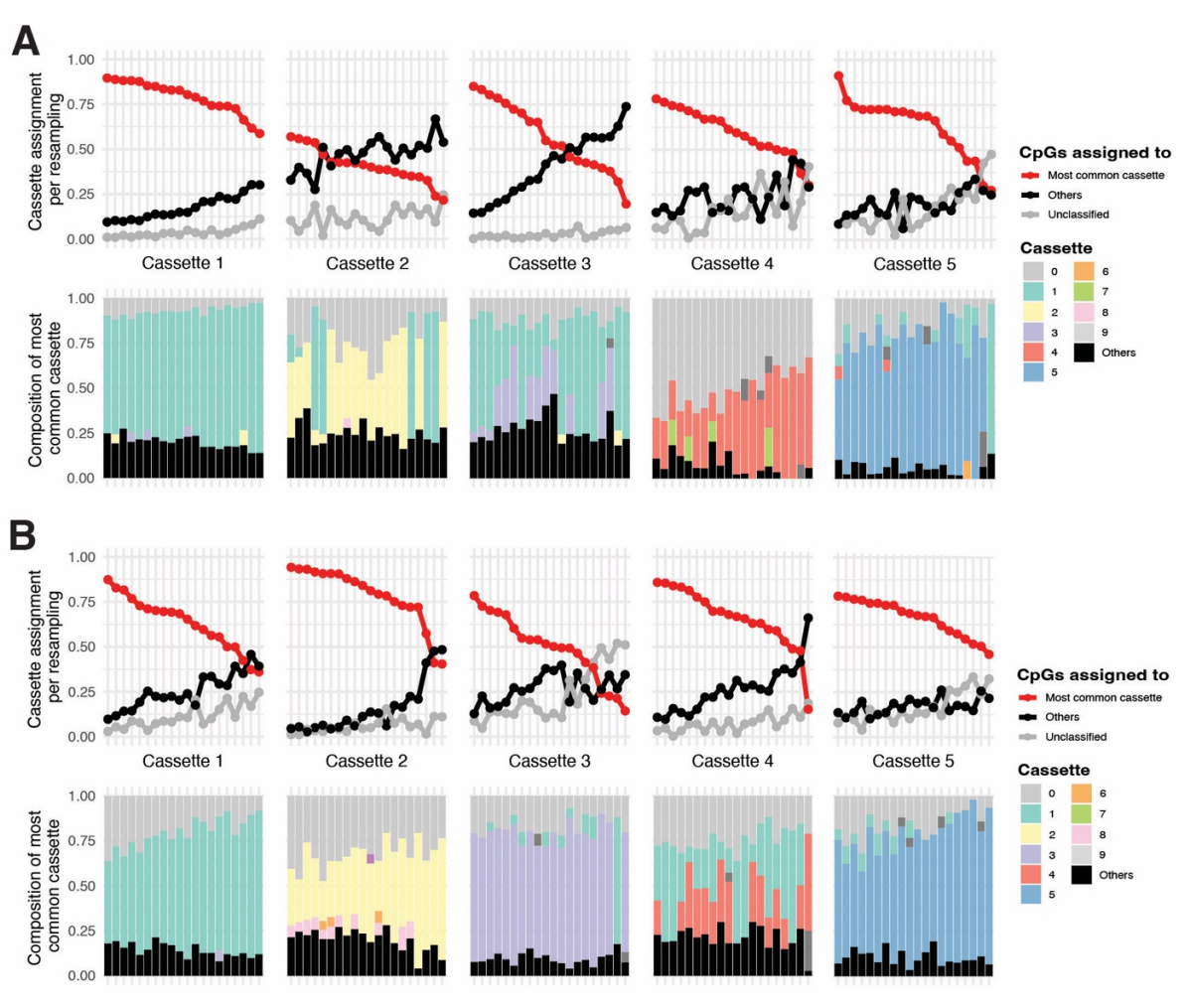
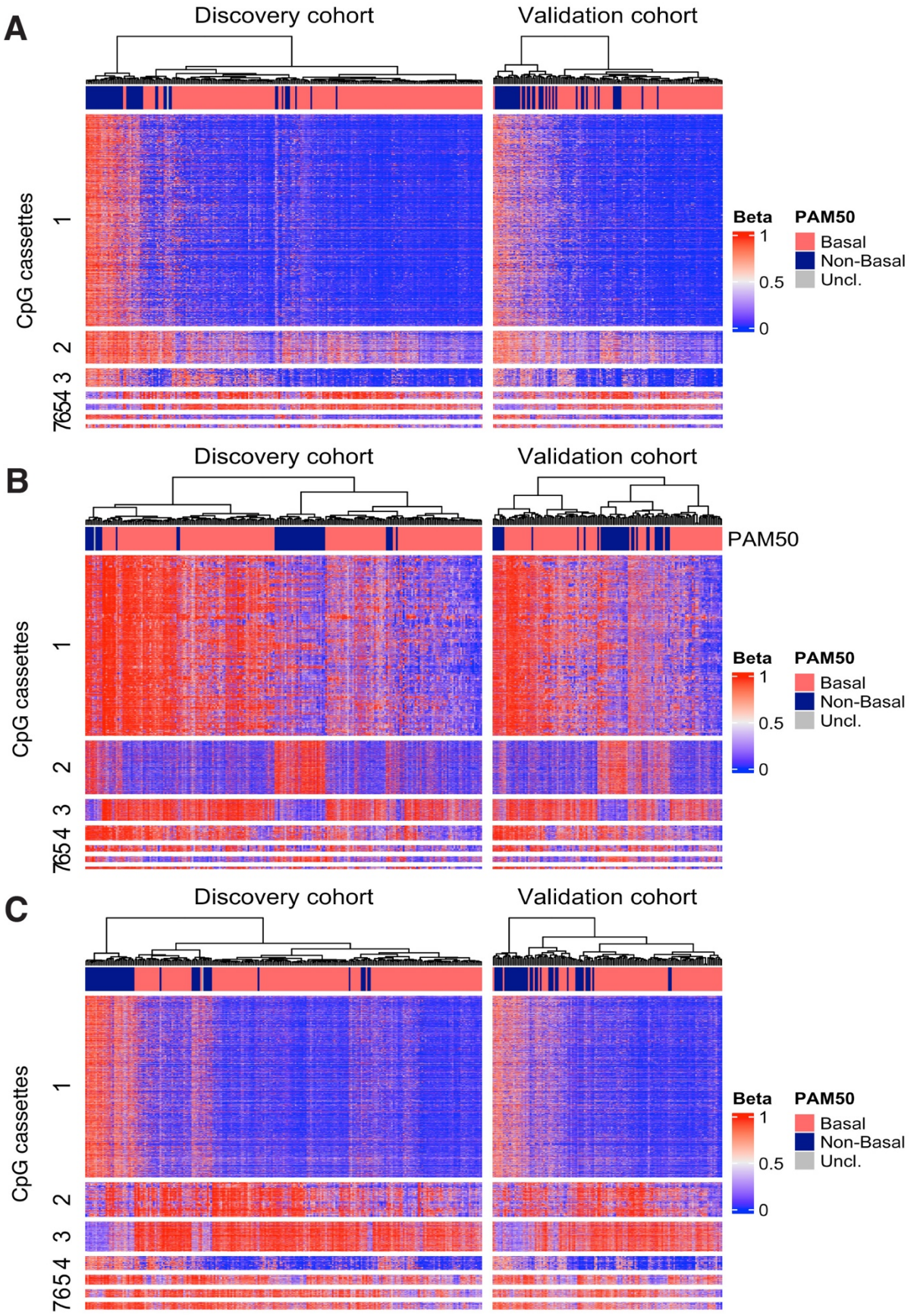


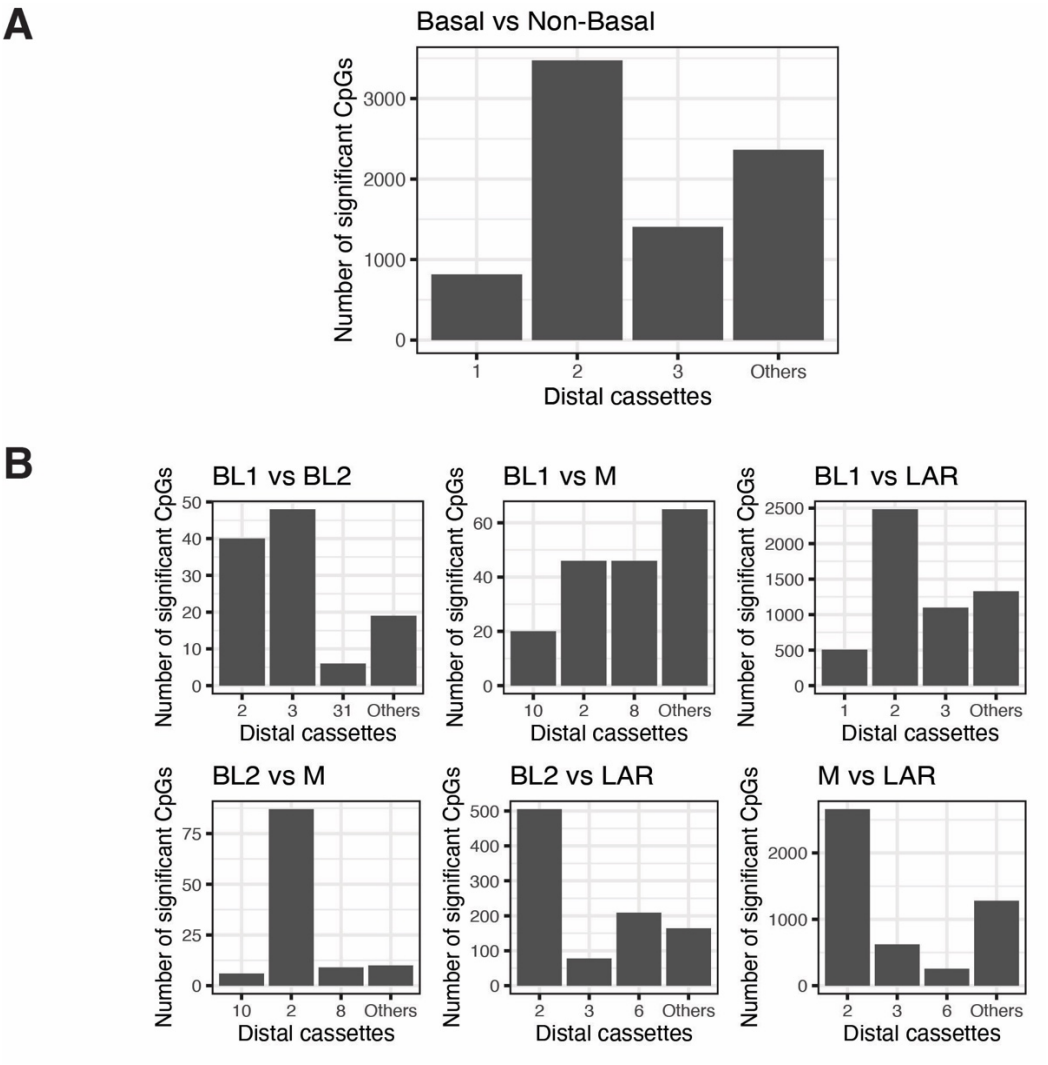
Supplementary Figure 1. Selection of soft thresholding power in WGCNA using the discovery cohort. Soft threshold (β) selection for **A**) all purity adjusted CpG beta values, **B**) all unadjusted CpG beta values, **C**) promoter purity-adjusted CpG beta values, **D**) proximal purity-adjusted CpG beta values, and **E**) distal purity-adjusted CpG beta values. Low-variance CpGs were excluded to reduce computational load: CpGs with variance<0.1 were removed for panels A and E, and those with variance<0.05 for panels C and D. For panel B, the same number of high-variance CpGs as in A was used. In each case, the left panel shows the scale-free topology model fit versus soft threshold, identifying the saturation point to be between 9 and 12. The right panel shows connectivity versus soft threshold, indicating a sharp decrease when beta increases, reflecting increasing network sparsity. Based on these plots, β=10 was selected for all subsequent WGCNA analyses.



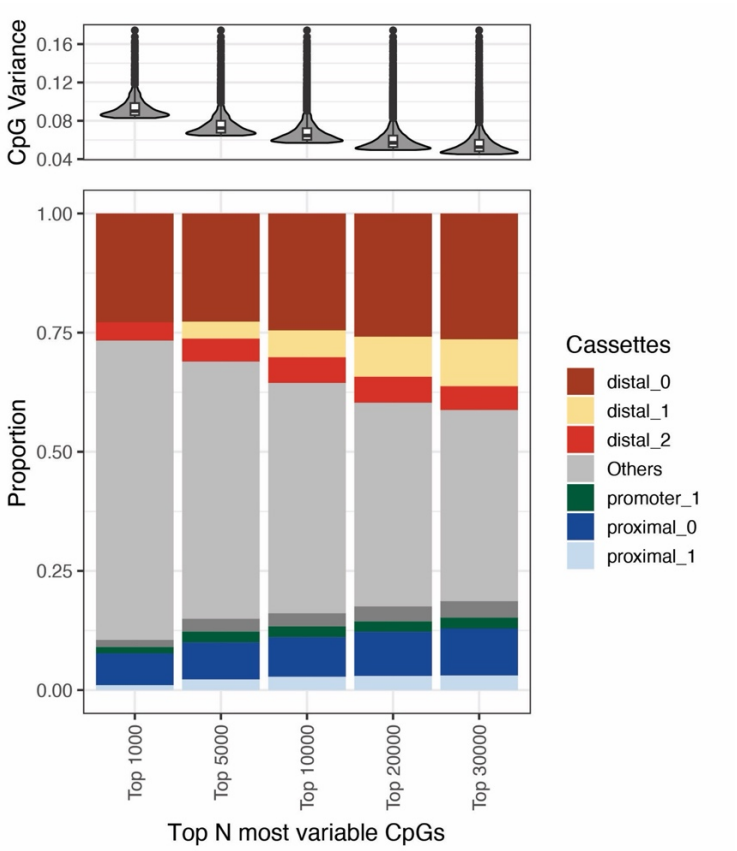
Supplementary Figure 2. Stability analysis of major promoter (A) and proximal (B) cassettes. To assess the stability of the detected cassettes, 100 samples were randomly resampled 20 times while preserving proposed DNA methylation epitype proportions from Aine et al. Top row: for each CpG site originally assigned to a given cassette, the proportion of resamplings in which it was assigned to the most common cassette (red), to any cassette (black), or to a cassette other than the most common (gray) is shown. A high red line indicates consistent reassignment to the same cassette across resamplings. Bottom row: proportional cassette assignments for each CpG site across resamplings, with colors representing the cassettes to which CpGs were assigned. Uniform colors within a cassette indicate strong preservation of the original group structure. CpGs from the original proximal and promoter cassettes generally clustered together across resamplings, forming cassettes largely composed of CpGs from the same original cassette. Nevertheless, some cassettes, such as promoter cassette 3 or proximal cassette 4, often appeared merged with promoter cassette 1 and proximal cassette 1, respectively, even if its CpGs were clustered together due to the strong anticorrelation between both cassettes and the unsigned configuration of the WGCNA implementation. The cassettes annotated as “0” represent the unclassified CpGs.



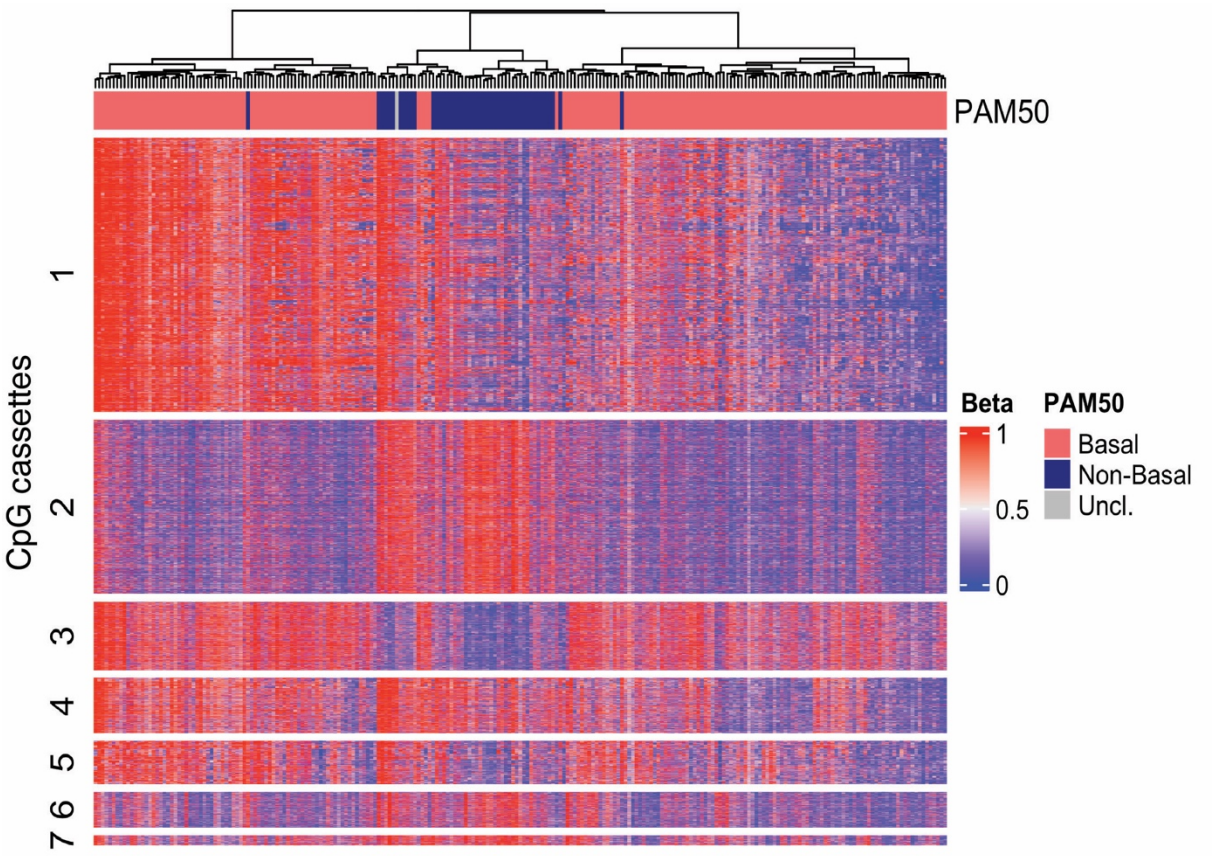
Supplementary Figure 3. CpG cassettes are detected consistently in the SCAN-B validation cohort. Heatmaps show the seven largest CpG context derived cassettes determined from the SCAN-B discovery and validation cohorts side by side: **A**) promoter CpGs, **B**) distal CpGs, and **C**) proximal CpGs. For each heatmap, the information included in the top annotations is detailed in respective heatmap's legends. Rows in the heatmaps correspond to CpGs, columns to samples. For each CpG context the patterns obtained from the discovery cohort are consistently preserved in the validation cohort.



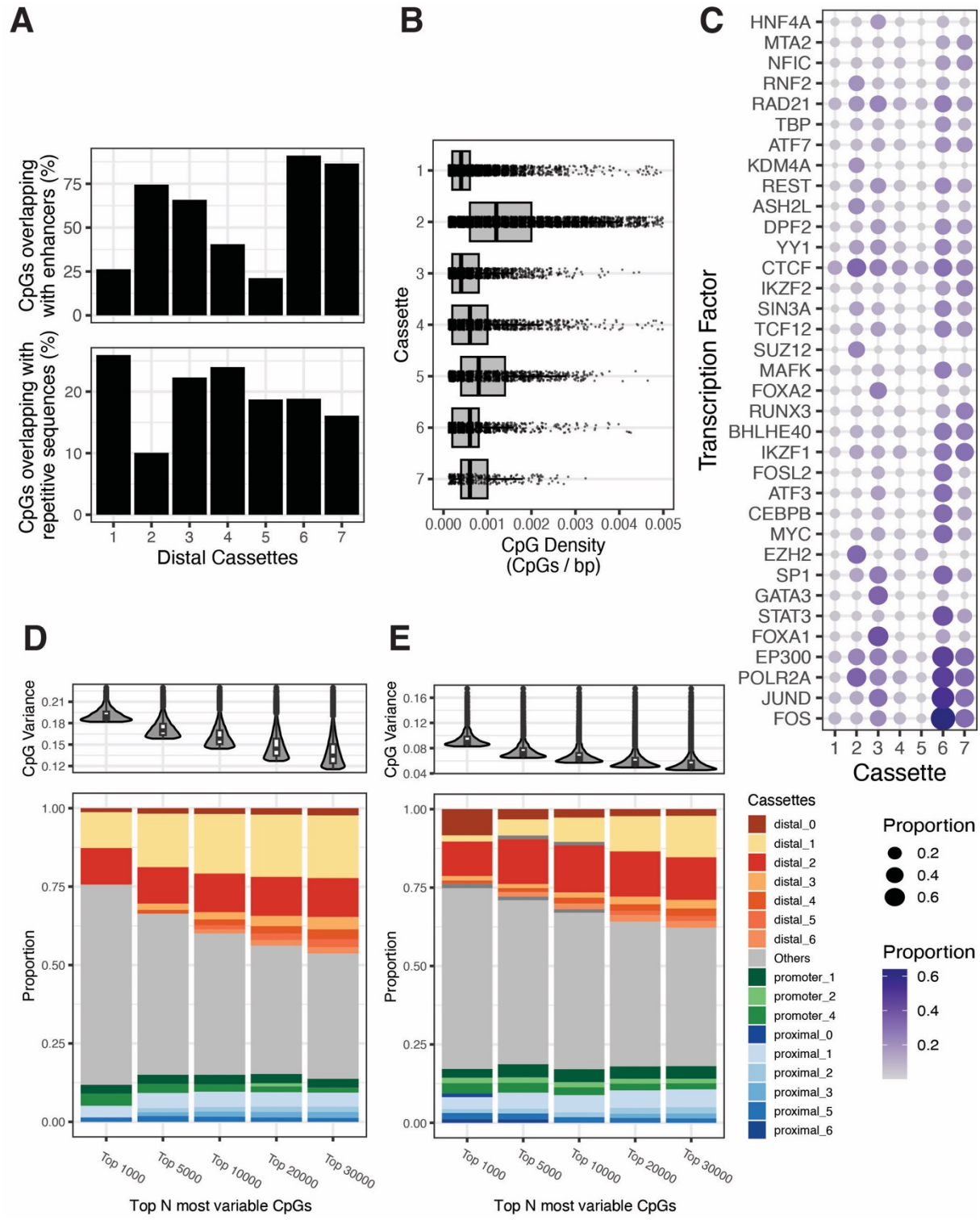
Supplementary Figure 4. Distal CpG cassette distribution of CpG sites selected by differential methylation analyses. The purity adjusted beta values of all CpGs from the methylation cohort were analysed pairwise to detect differentially methylated CpG sites in Basal and non-Basal TNBC samples (**A**) and in every combination of the Lehmann et al. TNBC mRNA subtypes (TNBCtype) (**B**). The bar plots show the counts of CpGs selected (bonferroni adjusted $p \leq 0.05$) belonging to each of the distal cassettes. Cassettes with proportions of total CpGs below 0.05 were grouped in the ”Others” category.



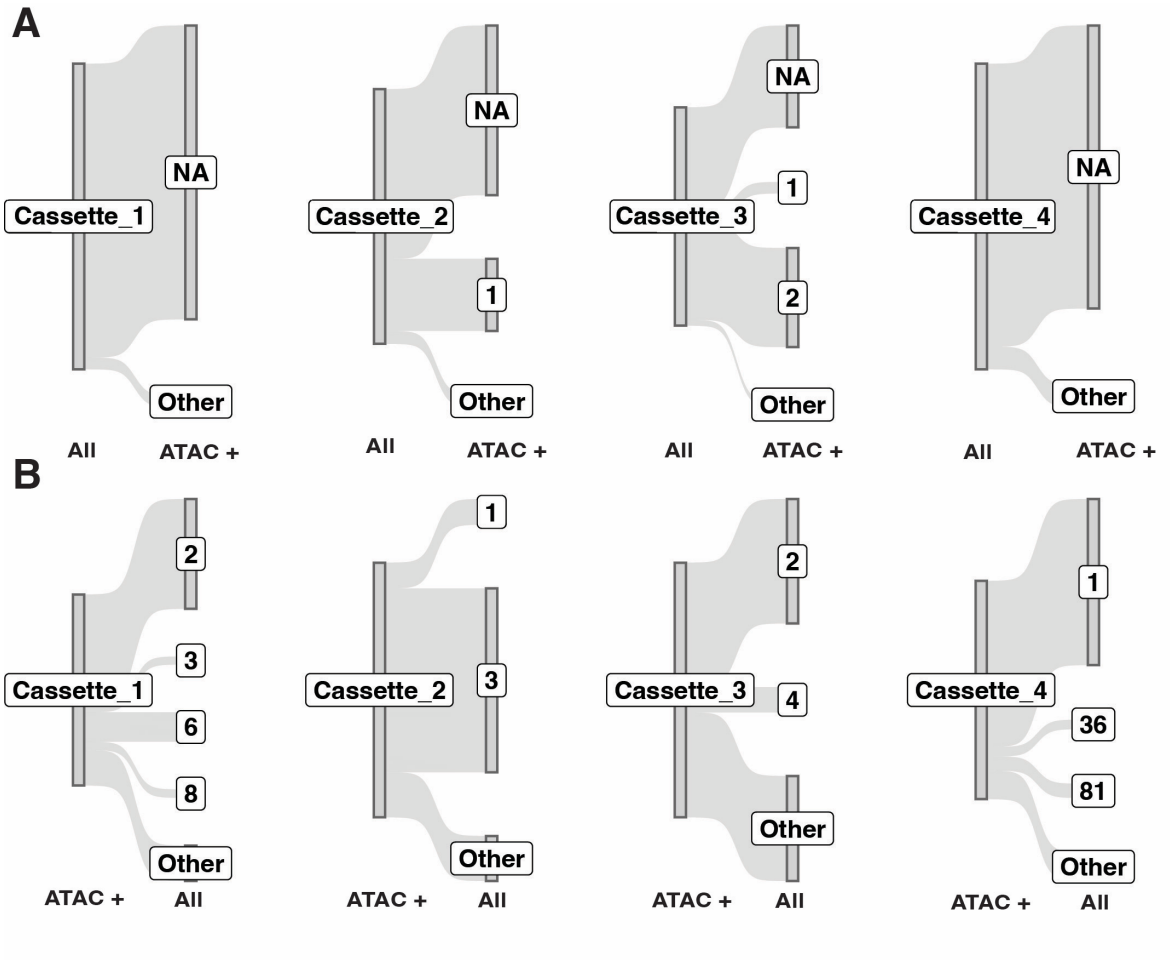
Supplementary Figure 5. Proportion of CpG cassettes per N top variable CpGs in the discovery cohort. The variance used to select the CpGs was obtained from unadjusted beta values. The top violin plot shows the measured variance from unadjusted betas per group, and the bar plot the proportion of cassettes for respective N. CpGs below 1% of the total CpGs of each variance-based group were grouped and identified as ”Others”. Cassettes annotated as “0” represent the unclassified CpGs.



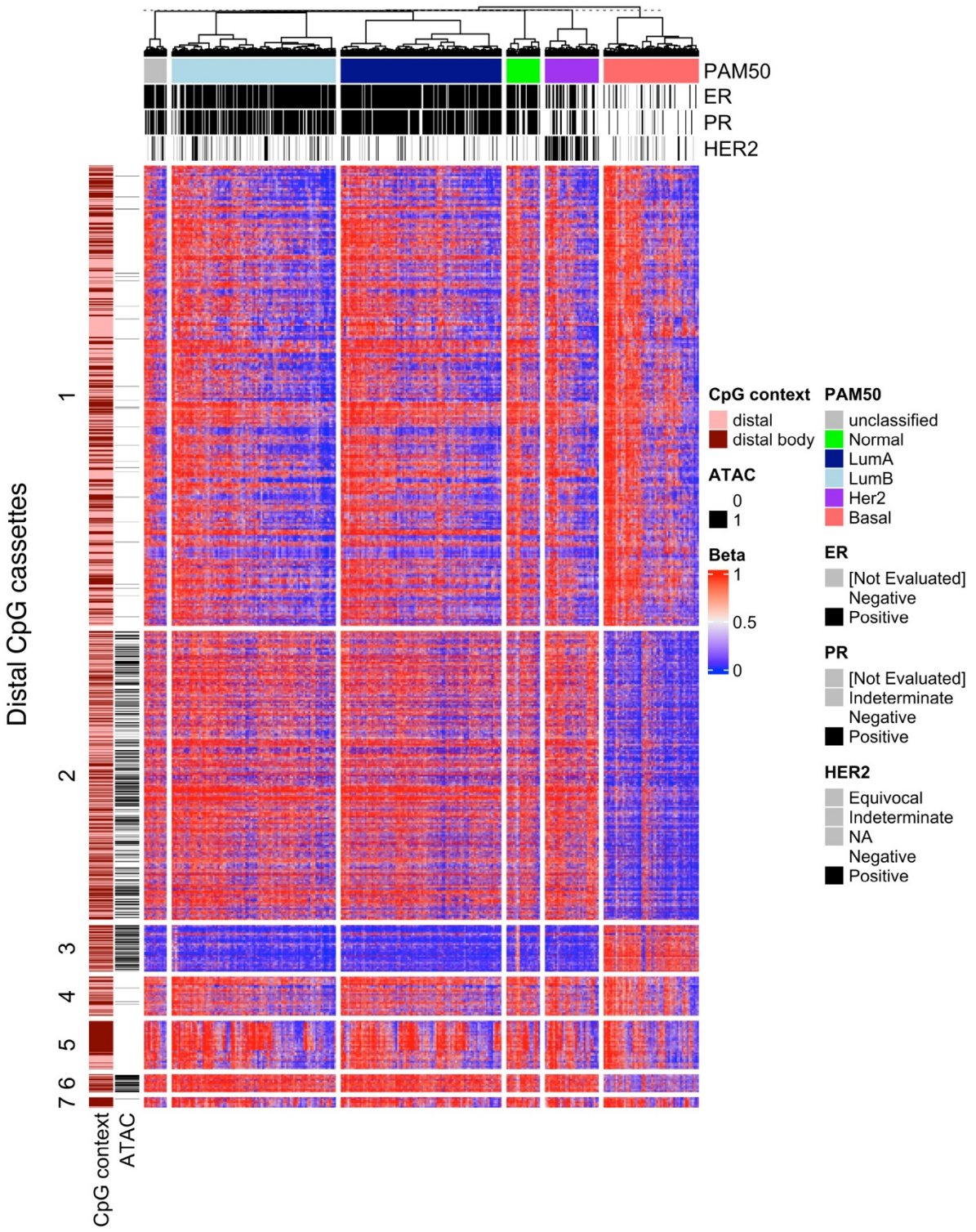
Supplementary Figure 6. Distal CpG cassettes detected in the discovery cohort using β=5 as the soft-thresholding value to minimize the proportion of unclustered CpGs. The remaining parameters and variance-threshold were kept unchanged from the ones described in the Methods section. Rows in the heatmap correspond to CpGs, columns to samples.



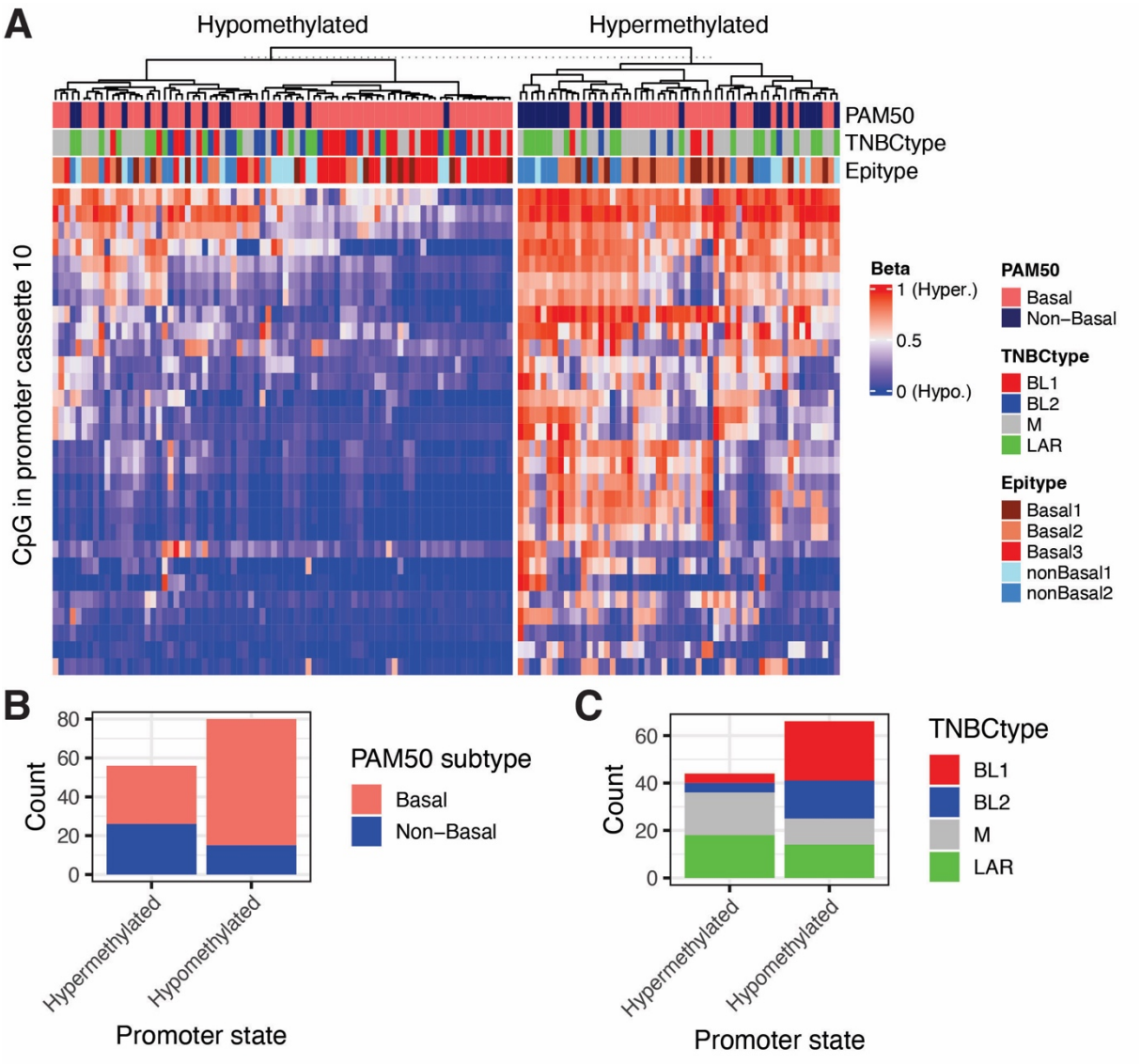
Supplementary Figure 7. Analysis of distal CpG cassettes and proportion of cassettes after variance-based filtering using β=5 as soft-thresholding power. A) Fraction of CpGs for the seven largest distal CpG cassettes overlapping with enhancers and repetitive regions. **B**) CpG density for the seven largest distal CpG cassettes. The density was determined using windows of 5000 base pairs. **C**) Fraction of CpGs in the seven largest distal cassettes overlapping with transcription factor binding sites. The transcription factors shown are the 35 ones with the highest pairwise differences between any two analyzed cassettes. **D**) Proportion of CpG cassettes per N top variable CpGs after variance-based filtering on purity adjusted methylation data. The top violin plot shows the measured variance from adjusted beta values per group, and the bar plot the proportion of different CpG cassettes. CpGs below 1% of the total CpGs were grouped and identified as ”Others”. **E**) Proportion of CpG cassettes per N top variable CpGs after variance-based filtering on non-purity adjusted methylation data. The top violin plot shows the measured variance from unadjusted betas per group, and the bar plot the proportion of cassettes. CpGs below 1% of the total CpGs were grouped and identified as ”Others”. Cassettes annotated as “0” represent the unclassified CpGs.



Supplementary Figure 8. Effect of ATAC-seq peak filtering on detected CpG cassettes when β=5. A) Sankey plot between cassettes determined from all CpGs (left, first four) and from ATAC-seq peak filtered CpGs (right). The NA category represents filtered CpGs and ”Other” includes cassettes whose proportion is below 4%. **B**) Sankey plot between CpG cassettes determined from ATAC-seq peak filtered CpGs (left, first four cassettes) and from all CpGs (right). The ”Other” class includes cassettes whose proportion is below 4%.



Supplementary Figure 9. Methylation patterns of distal CpG cassettes 1 to 7 in the TCGA breast cancer validation cohort. Heatmap of CpG beta values for CpGs in the seven largest distal cassettes across 645 breast cancers from the TCGA cohort. Only the CpGs overlapping between the Illumina EPIC and Infinium 450K methylation array platforms could be analyzed. The CpG cassette annotations displayed is from the annotations from the EPIC dataset. The samples are ordered based on PAM50 subtypes. The information included in the top annotations is detailed in the heatmap's legends. Rows in the heatmap correspond to CpGs, columns to samples.



Supplementary Figure 10. Promoter CpG cassette 10 in the SCAN-B validation cohort. A) Methylation pattern of promoter cassette 10 in the SCAN-B validation cohort. The samples are grouped using k-means clustering with k=2 on the detected cassette's beta values in the validation cohort. The information included in the top annotations is detailed in the heatmap's legends. Rows in the heatmap correspond to CpGs, columns to samples. **B)** Distribution of PAM50 Basal/non-Basal subtypes per methylation state. **C)** Distribution of Lehmann et al. TNBC mRNA subtypes (TNBCtype) per methylation state.

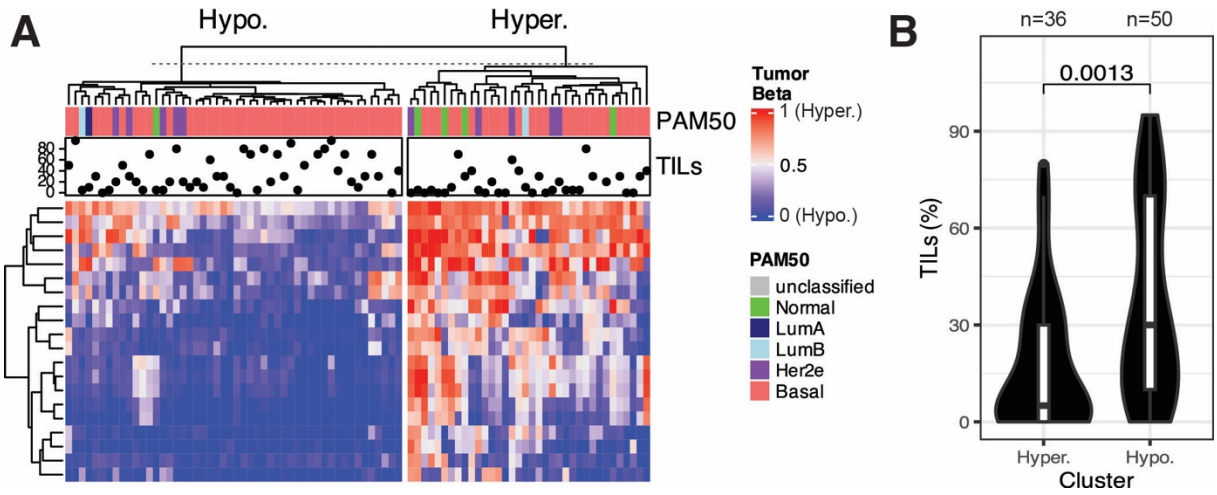
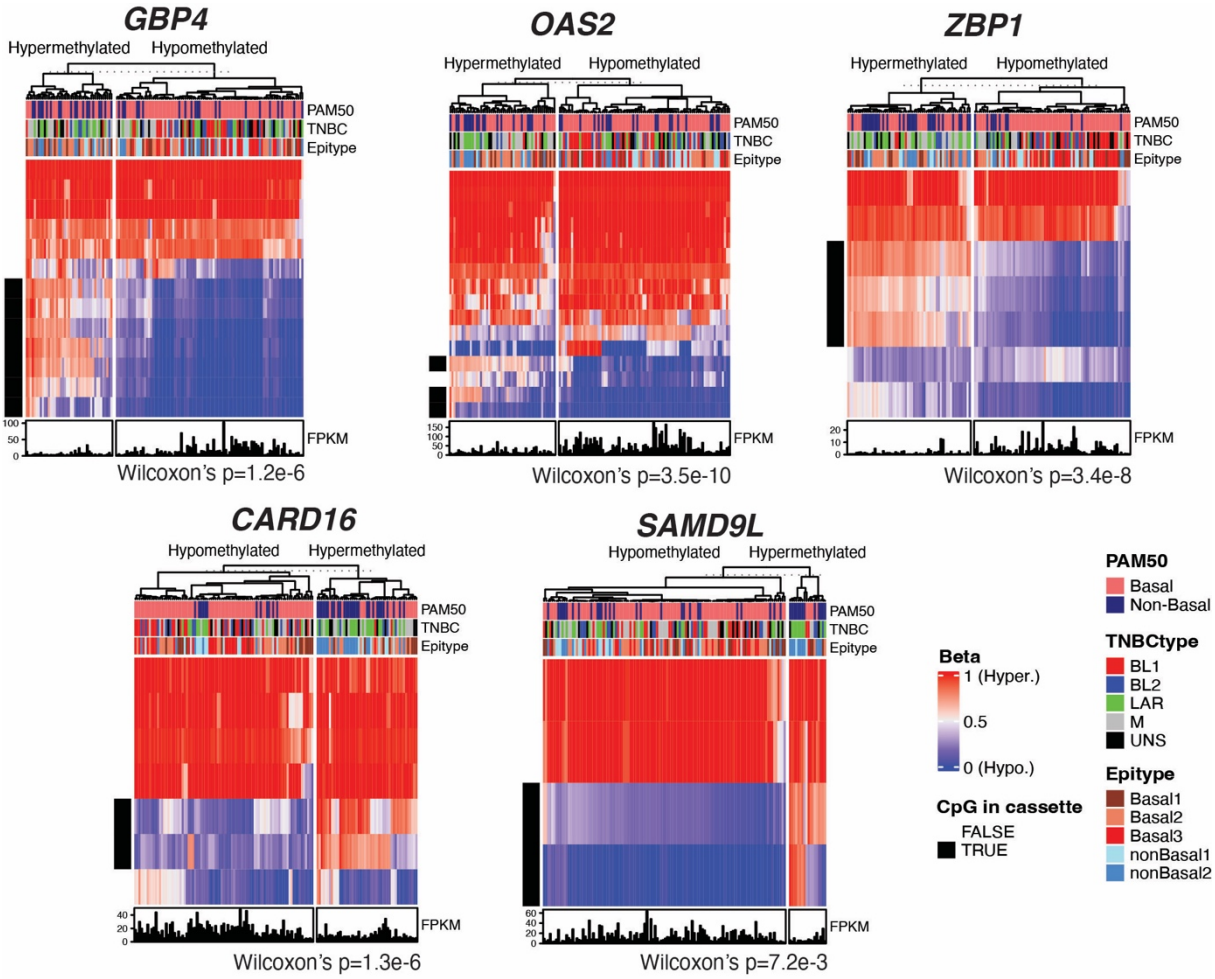
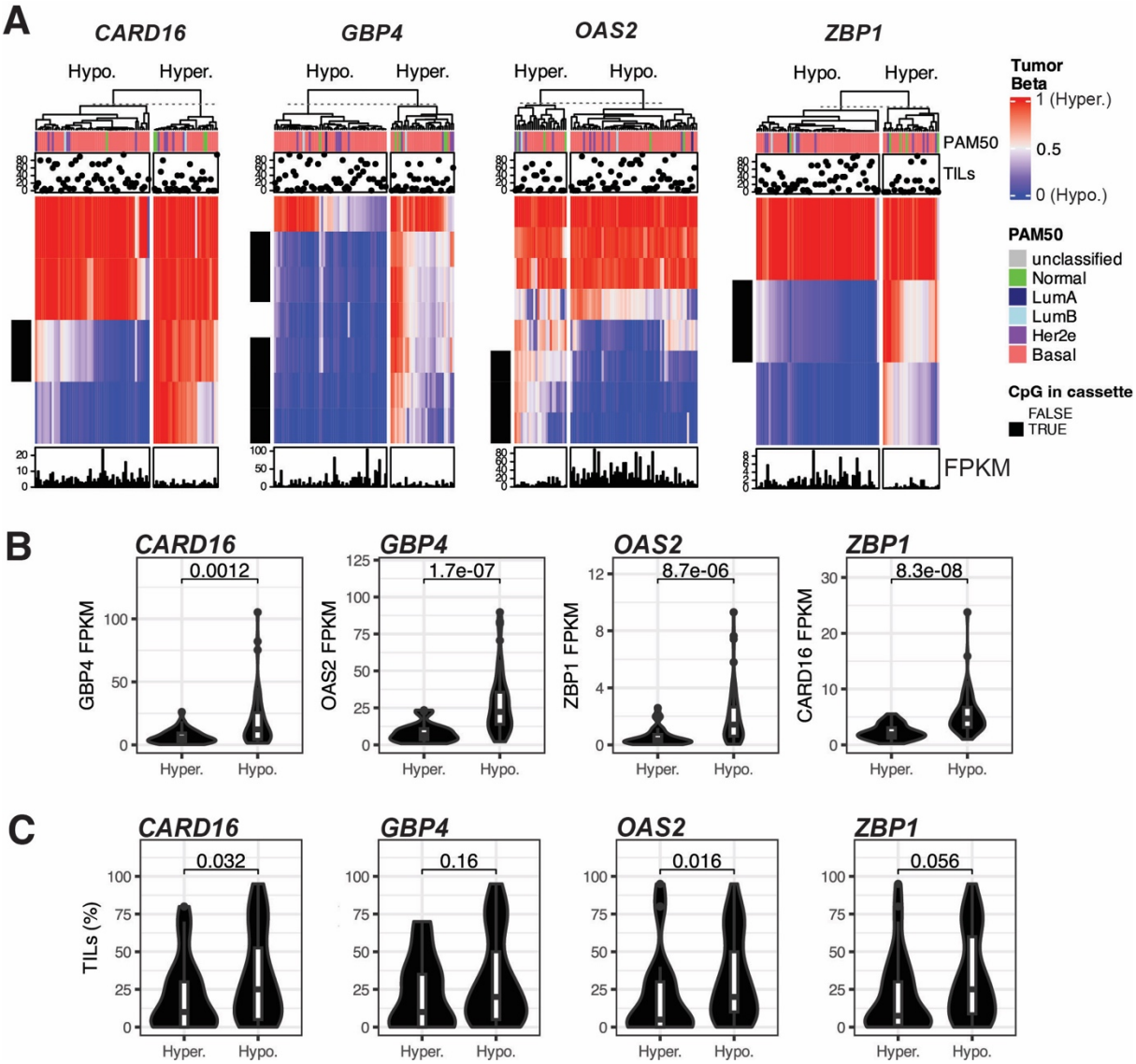


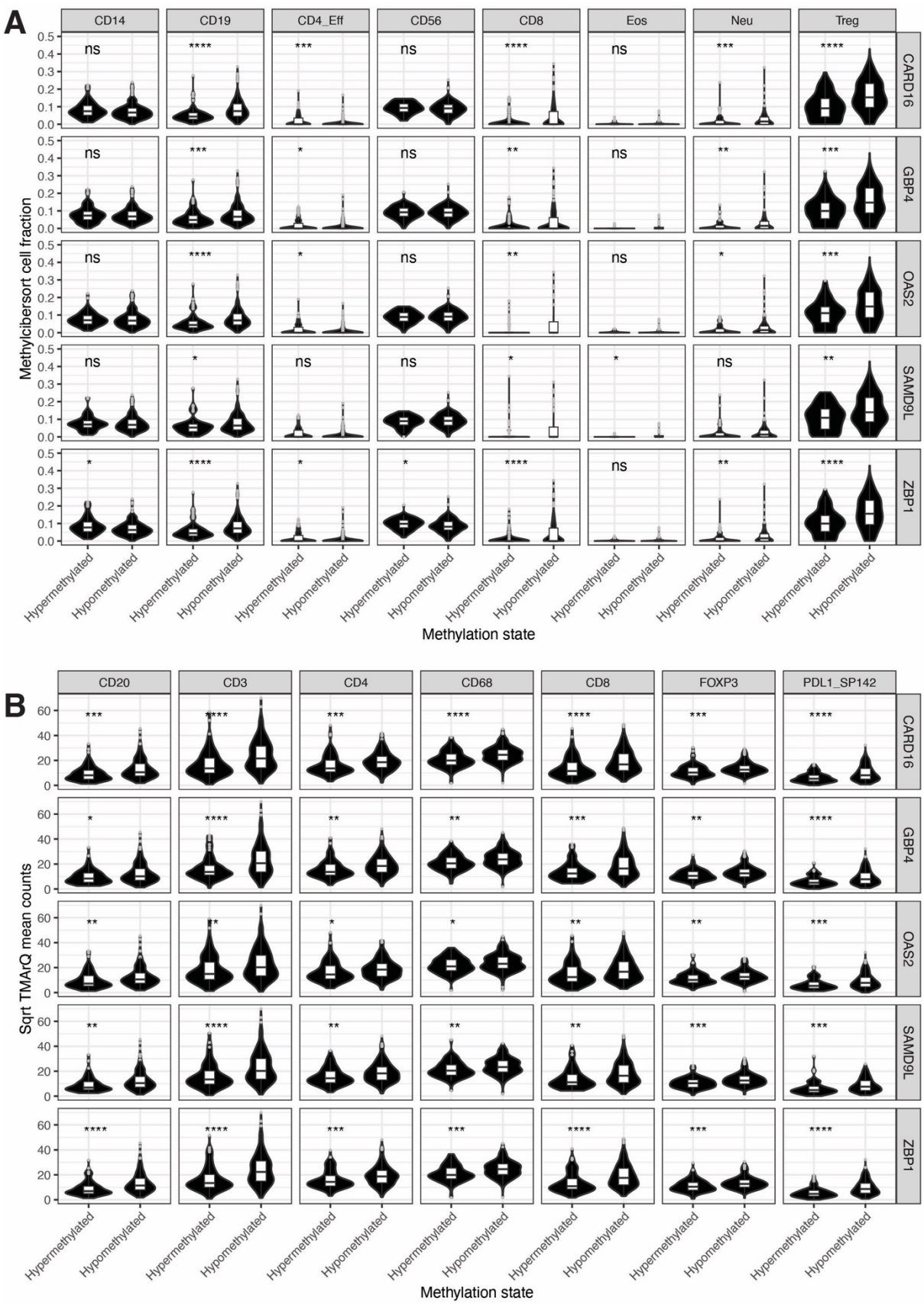
Figure 11. Promoter Cassette 10 in TNBC tumors from the TCGA breast cancer validation cohort. A) Methylation pattern of promoter cassette 10 in the external validation cohort. Only CpGs included in the available Illumina Infinium 450K methylation arrays in this cohort were used. The samples are grouped using k-means clustering with k=2 on the detected cassette's beta values in the external validation cohort. The information included in the top annotations is detailed in the heatmap's legends. Rows in the heatmap correspond to CpGs, columns to samples. **B)** Association between promoter cassette 10 and immune infiltration estimated by TILs in TCGA TNBCs. Samples were clustered as performed in A. Statistical significance was assessed using Wilcoxon's test.



Supplementary Figure 12. Methylation state and gene expression of the five genes in promoter CpG cassette 10 in the SCAN-B validation cohort. Methylation state of CpGs overlapping the five genes targeted in promoter cassette 10 in the SCAN-B validation cohort. The samples are grouped using k-means clustering and k=2 on CpGs overlapping with the gene. The rows marked with black labels represent the CpGs included in promoter CpG cassette 10. FPKM gene expression of each gene per sample is shown in the bottom panel, which was significantly different per methylation cluster for the five genes analyzed using Wilcoxon's test. The information included in the top annotations is detailed in the heatmap's legends. Rows in the heatmap correspond to CpGs, columns to samples.

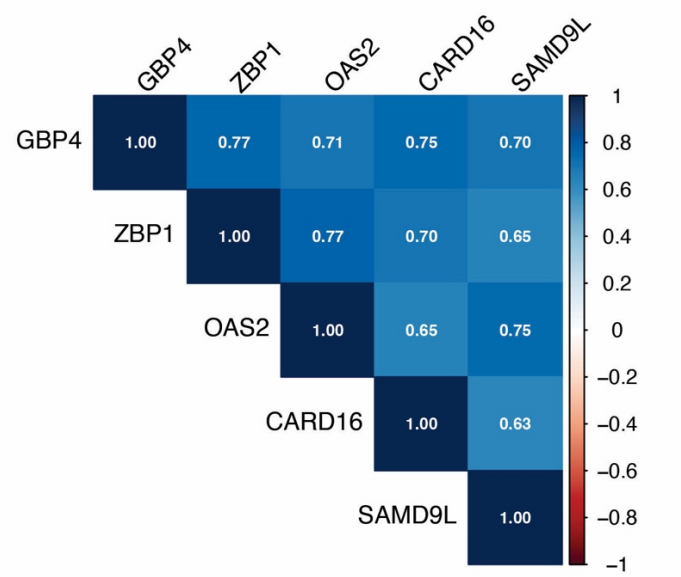


Supplementary Figure 13. Methylation state, gene expression and association to TILs of the genes in promoter CpG cassette 10 in the TNBCs from the TCGA breast cancer validation cohort. *SAMD9L* was excluded due to lack of CpG coverage in the available data. **A)** Methylation state of CpGs overlapping the four available genes targeted in promoter cassette 10 in the TCGA validation cohort. The samples are grouped using k-means clustering and k=2 on CpGs overlapping with the gene. The rows marked with black labels represent the CpGs included in promoter CpG cassette 10. The information included in the top annotations is detailed in the heatmap's legends. Rows in the heatmap correspond to CpGs, columns to samples. **B)** Association between methylation status defined in A with gene expression. Statistical significance was evaluated using Wilcoxon's test. **C)** Association between methylation status defined in A with TILs. Statistical significance was evaluated using Wilcoxon's test.

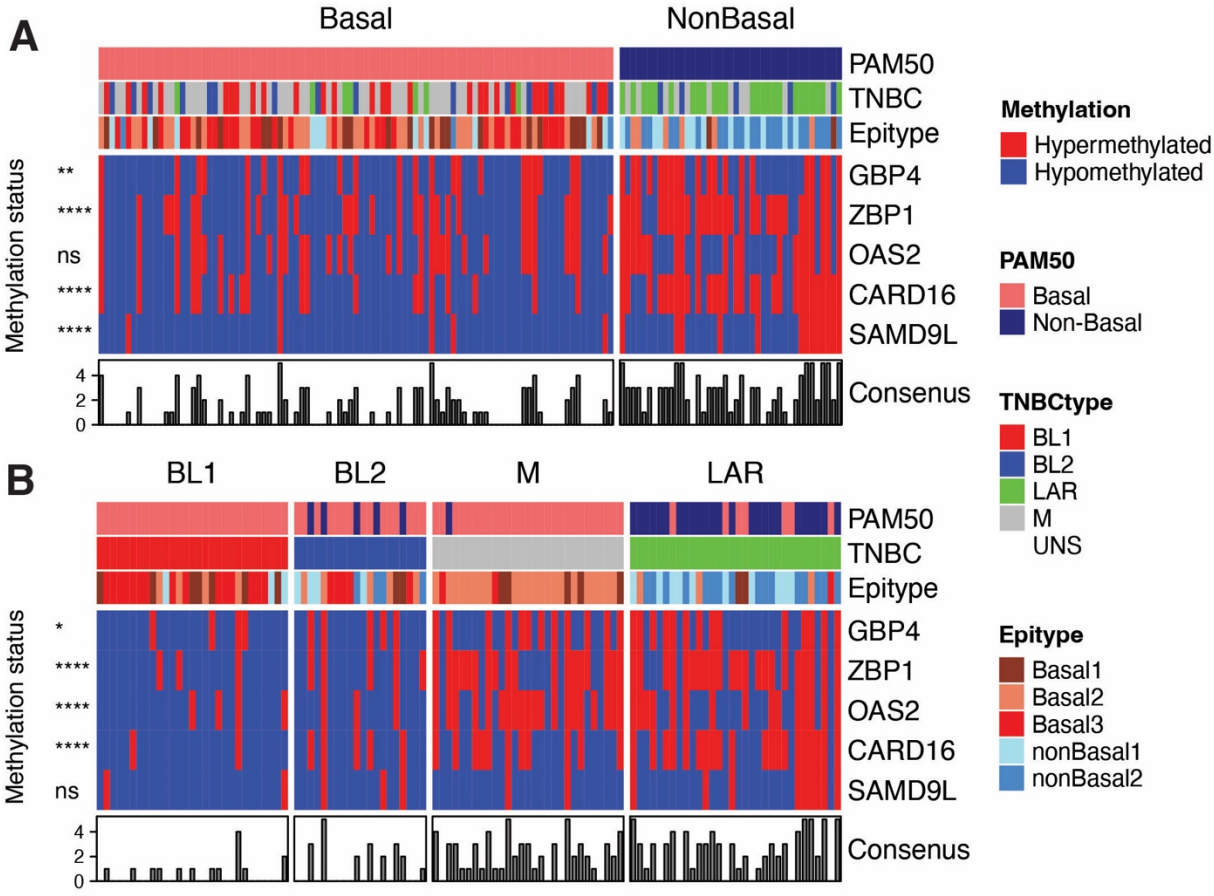


Supplementary Figure 14. Infiltration of different immune cell types based on methylation state of detected immune genes. A) Immune cell proportions determined by MethylCIBERSORT versus methylation status of each of the analyzed genes. CD14 stands for macrophages and monocytes, CD19 for B-cells, CD4_Eff stands for Helper T lymphocytes, CD56 for NK cells, CD8 for cytotoxic T lymphocytes, Neu for neutrophile and Treg for Regulatory T-cells. The methylation status was defined in Figure 5A. **B)** Square rooted TMArQ digital cell counts for immune cell type markers per methylation status of each of the analyzed genes. Here, PDL1 counts are based on IHC analysis using the

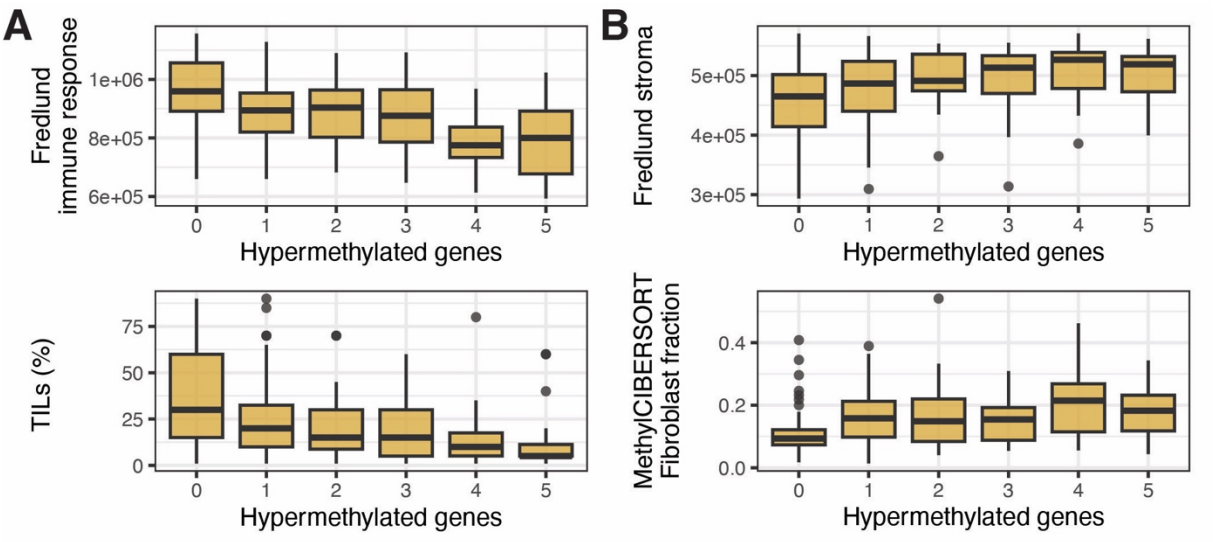
SP142 antibody (see Roostee et al. for details). The statistical significance values displayed were obtained using Wilcoxon's test.



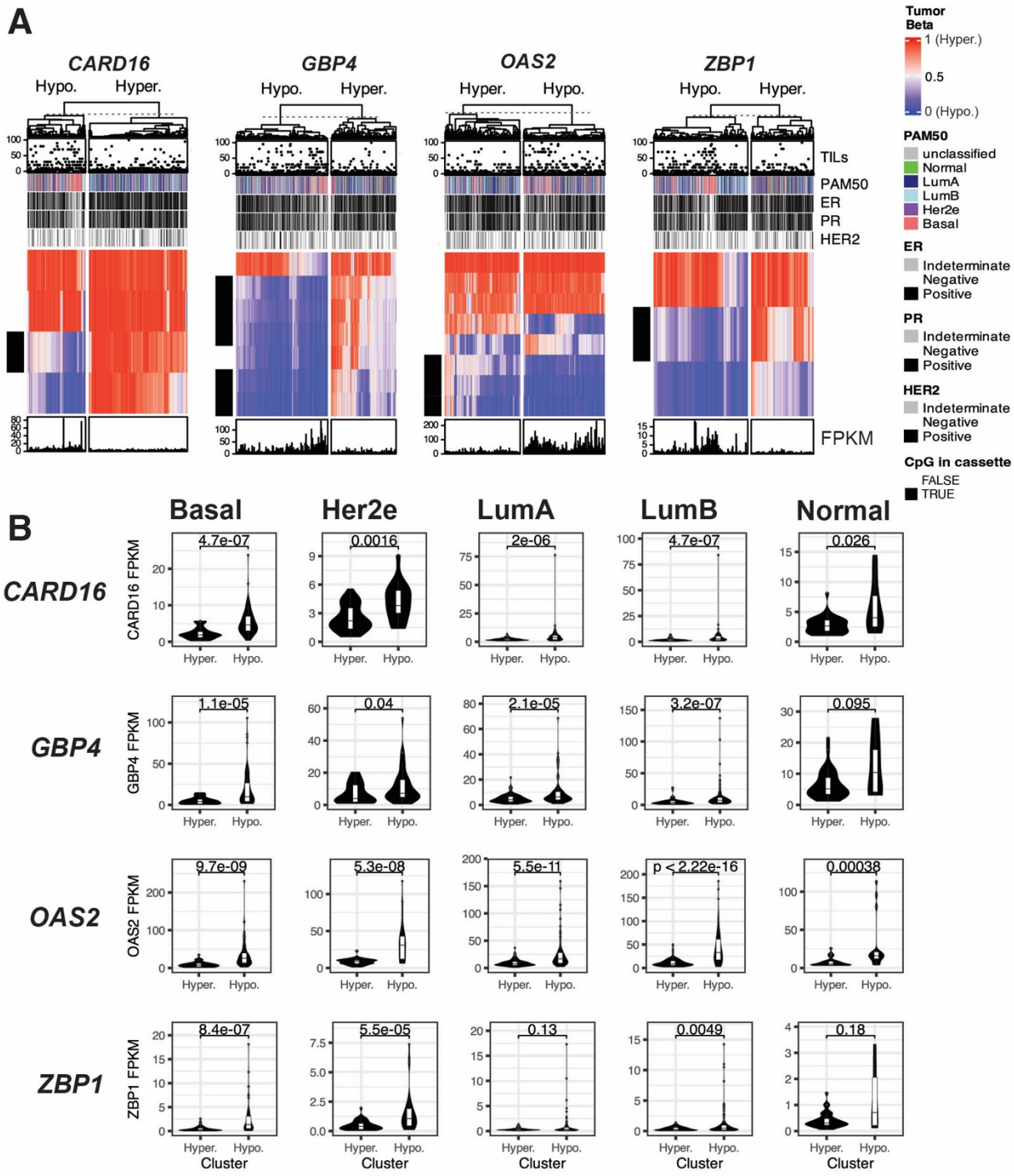
Supplementary Figure 15. Correlation plot between expression of the five genes from promoter CpG cassette 10 in the discovery cohort. The correlation values were calculated using the Spearman correlation coefficient from log1p-transformed FPKM gene expression values.



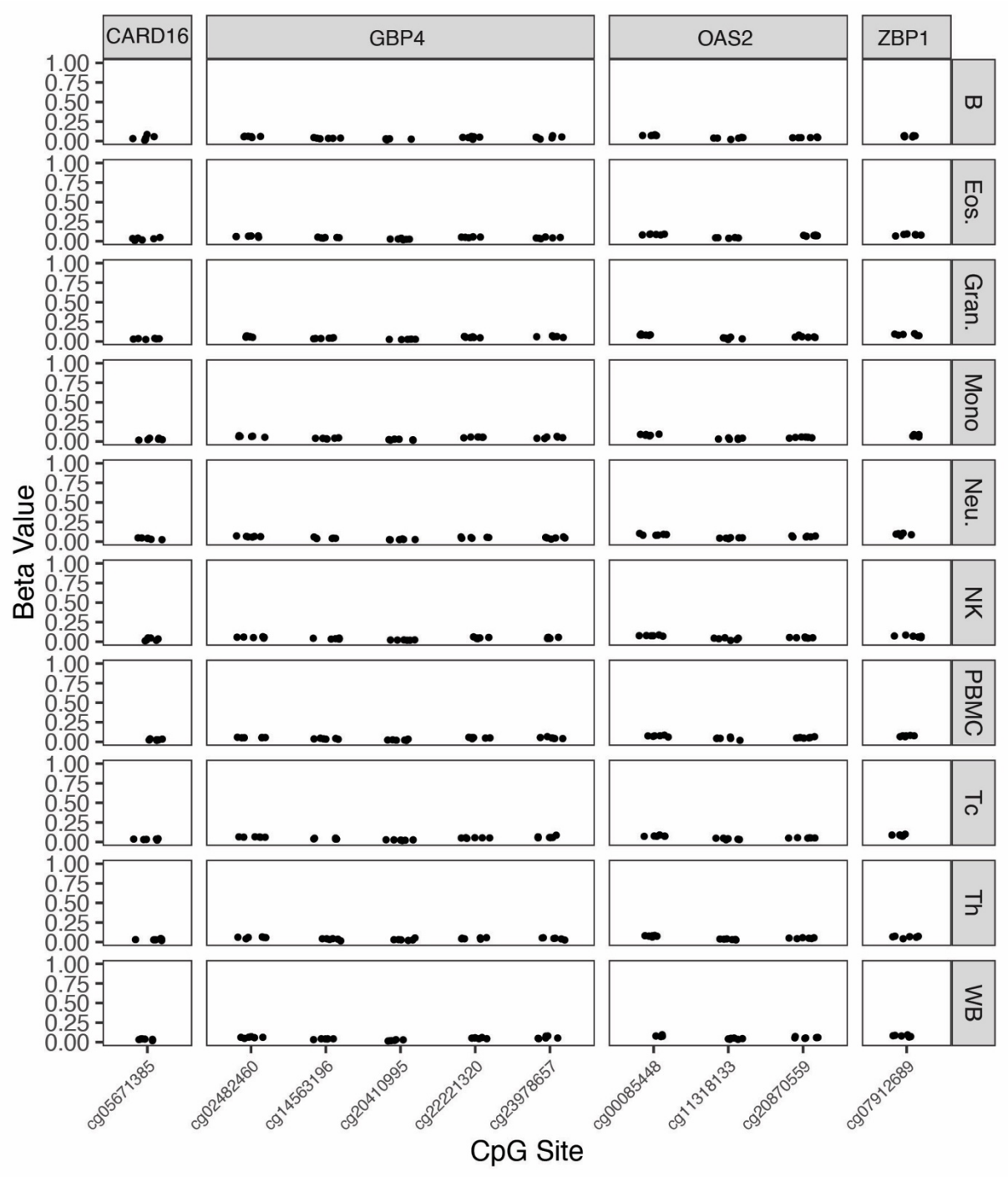
Supplementary Figure 16. Co-occurrence of the methylation patterns detected per gene in TNBC subtypes in the SCAN-B validation cohort. The samples are split based on Basal/non-Basal PAM50 groups (**A**) or Lehmann TNBCtype4 subtypes (**B**). The methylation status was defined in Supplementary Figure 12. Statistical significance values displayed on the left were obtained using the Chi-square test. *: p<0.05, **: p<0.01, ****: p<0.0001



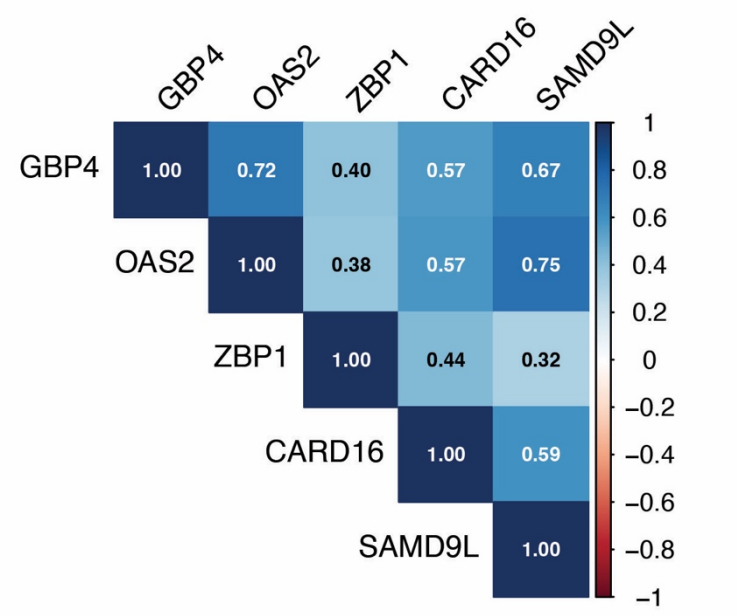
Supplementary Figure 17. Cumulative gene hypermethylation versus tissue composition in the discovery cohort. A) Effect of cumulative hypermethylation of the five genes in promoter CpG cassette 10 versus immune infiltration measured by rank-based scores for a gene expression immune response metagene (top) and TILs (bottom). **B**) Effect of cumulative promoter hypermethylation in the five genes versus stromal estimates by a rank-based gene expression stroma metagene (top) and estimated fibroblast fraction by MethylCIBERSORT (bottom). The methylation status was defined in Figure 5A.



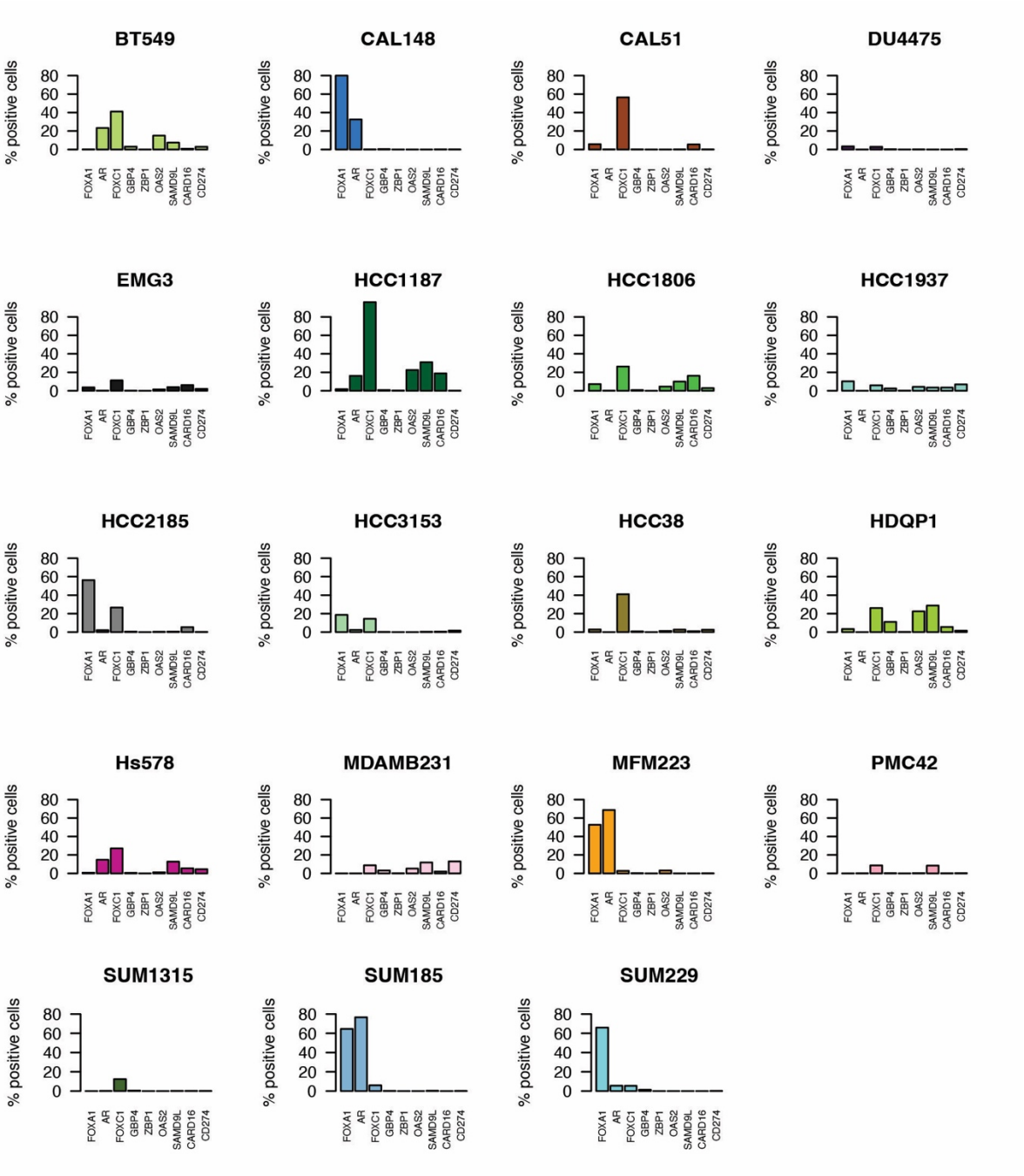
Supplementary Figure 18. Methylation state and gene expression of the five genes in promoter CpG cassette 10 in the TCGA breast cancer validation cohort. *SAMD9L* was excluded due to lack of CpG coverage in the available data. **A)** Methylation state of CpGs overlapping the four available genes targeted in promoter cassette 10 in the external validation cohort. The samples are grouped using k-means clustering and k=2 on CpGs overlapping with the gene. The rows marked with black labels represent the CpGs included in promoter CpG cassette 10. The information included in the top annotations is detailed in the heatmap's legends. Rows in the heatmap correspond to CpGs, columns to samples. Her2e: HER2-enriched. **B)** Association between methylation status defined in A with gene expression across PAM50 subtypes. Statistical significance was evaluated using Wilcoxon's test.



Supplementary Figure 19. Methylation state of CpGs in promoter CpG cassette 10 in immune cells from flow cytometry-sorted blood samples. CpGs not included were not available in the Reinius et al. dataset due to usage of the Illumina Infinium 450K platform. Cell types included are B-cells (B), eosinophiles (Eos.), granulocytes (Gran.), monocytes (Mono.), neutrophiles (Neu.), natural killer cells (NK cells), peripheral blood mononuclear cells (PBMC), cytotoxic T-cells (Tc), helper T-cells (Th), and whole blood (WB). Each cell type was profiled in six independent donors.



Supplementary Figure 20. Correlation plot between expression of the five genes from promoter CpG cassette 10 in bulk RNA-sequencing data from 34 TNBC cell lines. The correlation values were calculated using the Spearman correlation coefficient from log1p-transformed FPKM gene expression values.



Supplementary Figure 21. scRNA-seq expression of the five genes in promoter CpG cassette 10 in 19 TNBC cell lines. The expression is summarized as proportion of cells with detected expression (>0 counts). Four additional genes were added to the plots as controls: for a Basal/non-Basal phenotype, *FOXA1*, *AR* (androgen receptor), and *FOXC1*; and for immune response, *CD274* (PD-L1).