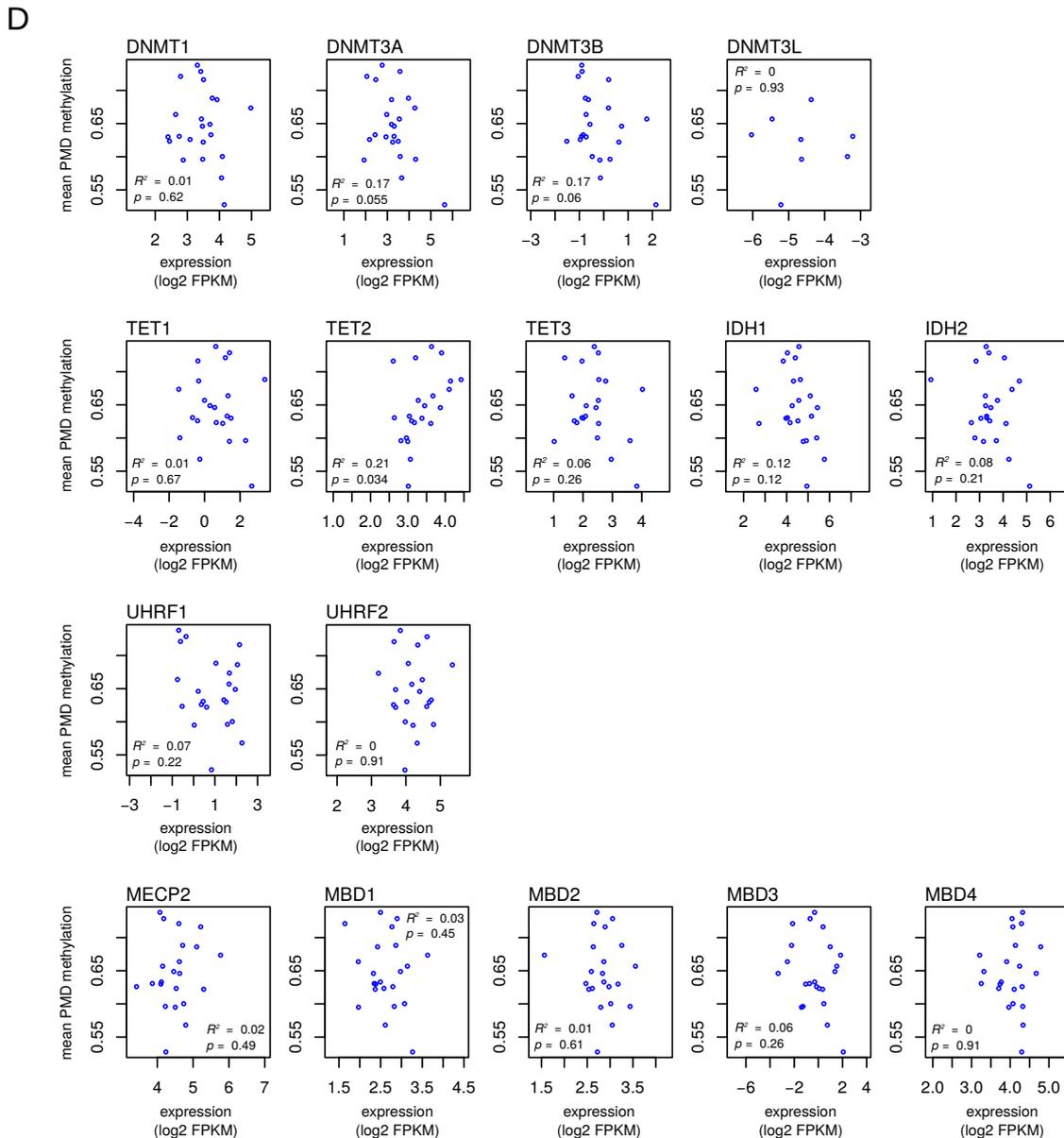
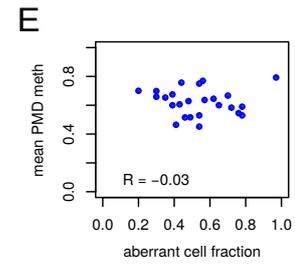
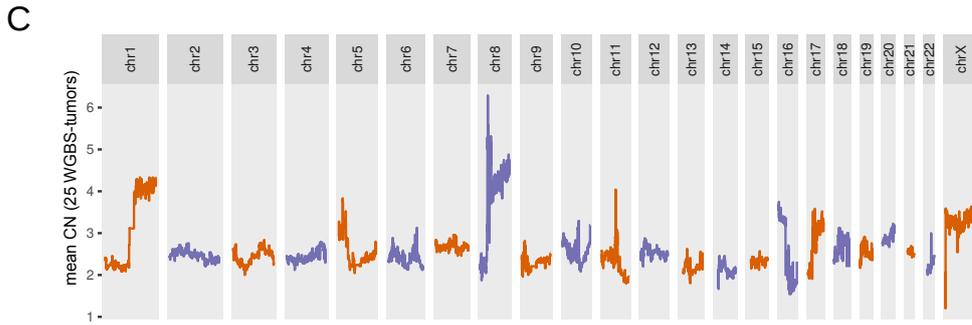
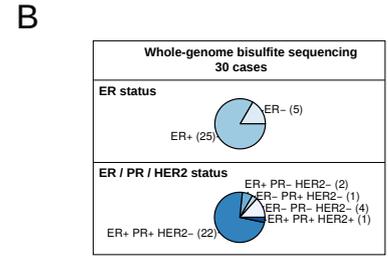
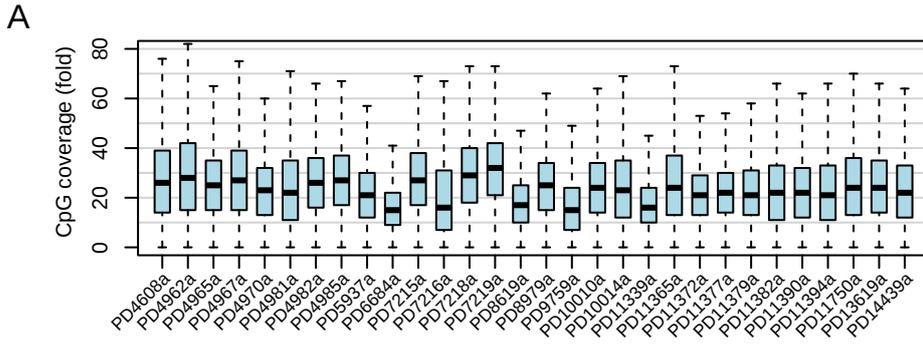


Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation

Arie B. Brinkman *et al.*

SUPPLEMENTARY INFORMATION

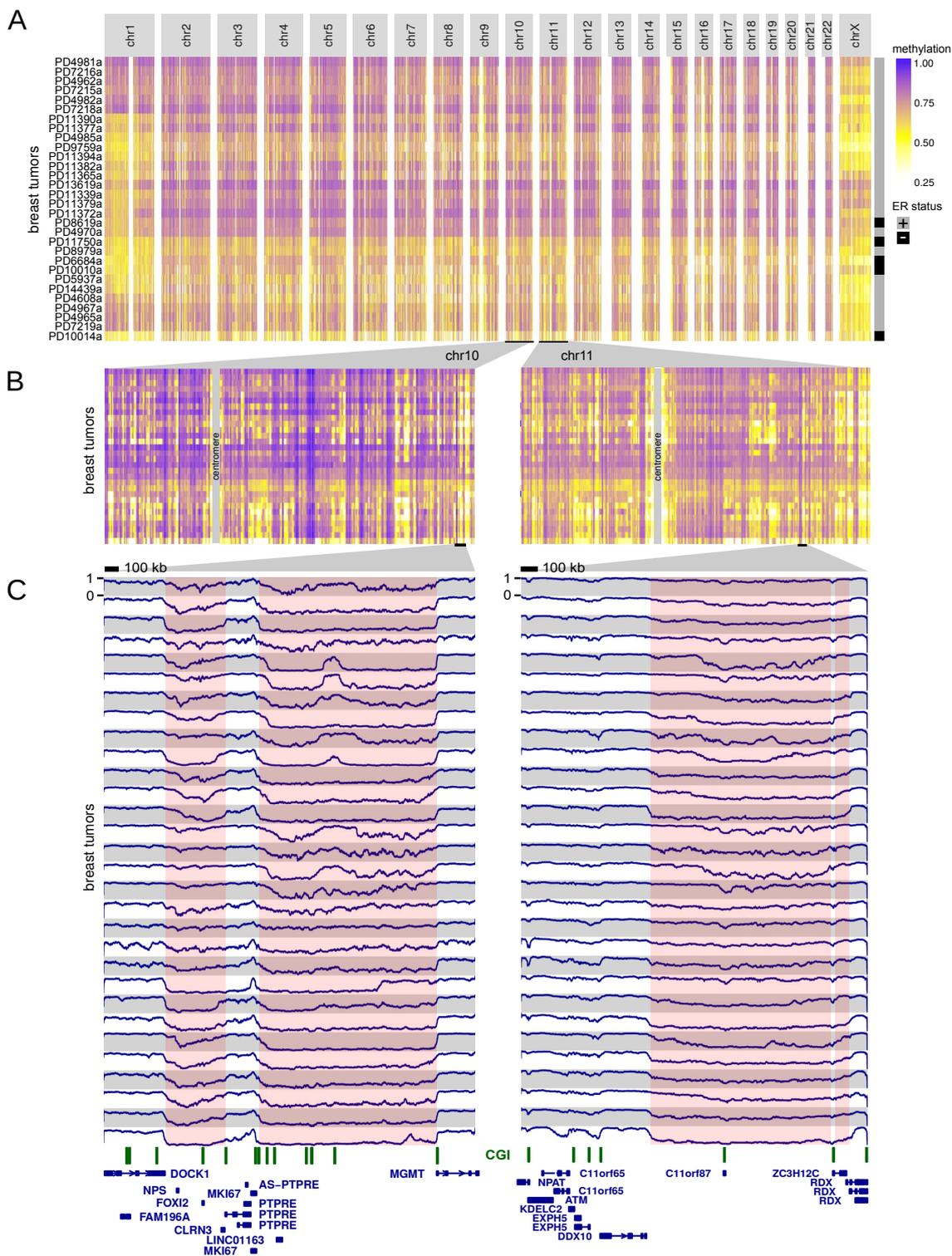


Supplementary Figure 1 | Sample characteristics: WGBS coverage, major pathological subtypes, copy-number profiles, expression of DNA methylation-related genes.

(A), CpG coverage in WGBS DNA methylation profiles of 30 breast tumor samples used in this study (see also Supplementary Table 1). Boxplots represent the median and 25th and 75th percentiles, whiskers 1.5 times the interquartile range, outliers are not shown. **(B)**, Clinicopathological features of the 30 tumor samples (see also Supplementary Table 2). **(C)**, Mean copy-number profiles of 25/30 tumor samples used in this study. Copy-number data was taken from our previous work¹. **(D)**, Association between mean PMD methylation and expression of genes involved in writing, erasing, or reading the 5-methylcytosine modification. Each dot represents one tumor sample. Linear regression was used to determine the variation explained (R^2) and the p-value of the association. Expression data was taken from our previous work². **(E)**, Mean PMD methylation (y-axis) is not associated with the fraction of aberrant cells (ASCAT³, x-axis).

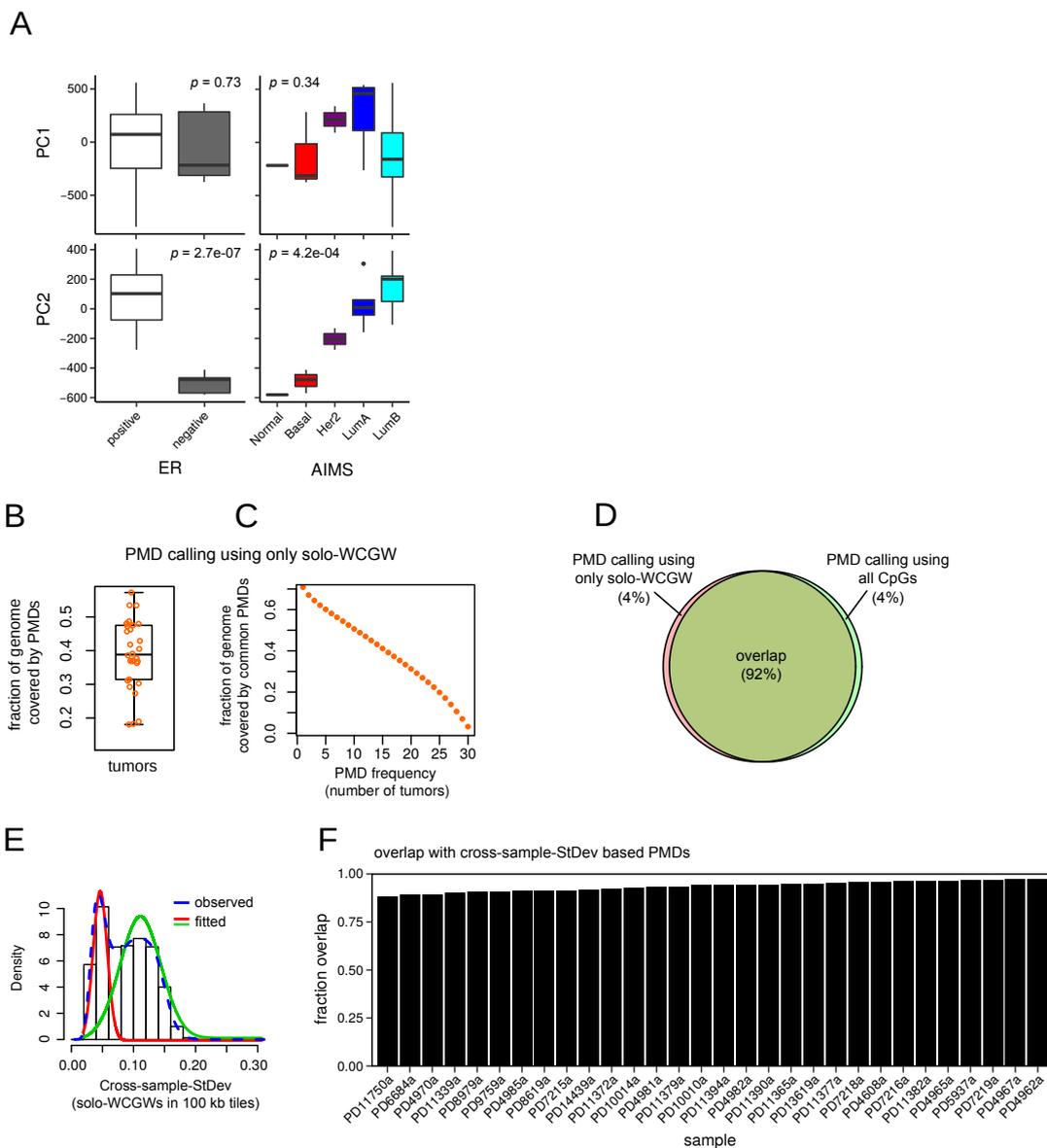
Supplementary Figure 2 | Visualization of inter-tumor variation at genome-wide scale, including non-tumor tissues.

Visualization of inter-tumor variation at genome-wide scale, as in main Figure 1, but including WGBS data from 72 additional, non-tumor tissues (Roadmap Epigenomics Project and ref.⁴). **(A)**, Genome-wide and **(B)**, chromosome-wide maps. Mean methylation is displayed in consecutive tiles of 10 kb (see Methods). For breast tumors of this study, the ER-status is indicated at the right (A).



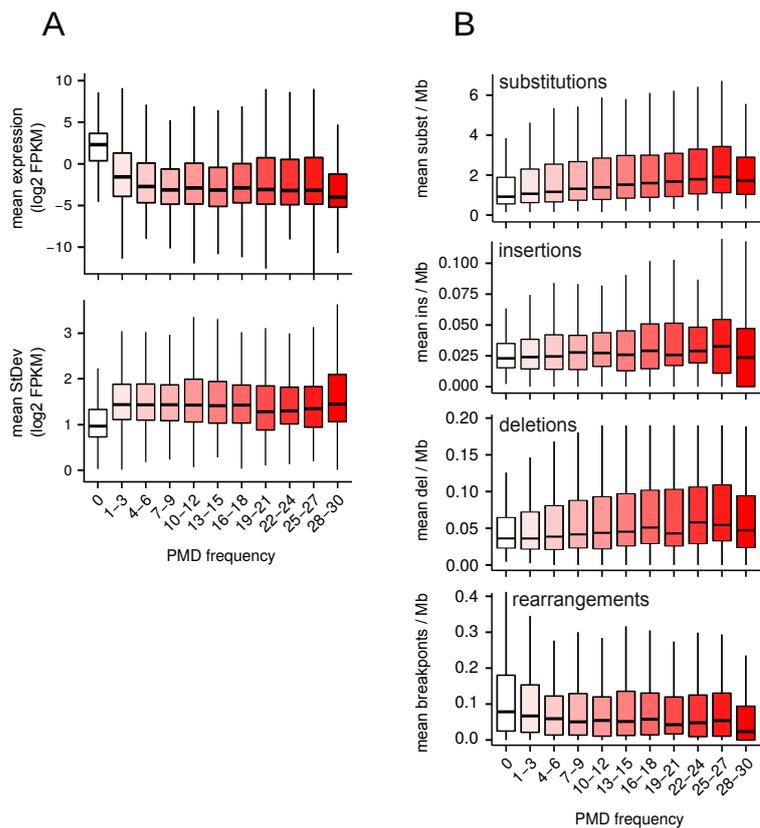
Supplementary Figure 3 | Visualization of inter-tumor variation at genome-wide scale, solo-WCGW CpGs only.

(A), Genome-wide and (B), chromosome-wide maps of WGBS DNA methylation profiles from 30 breast tumor samples. Exactly as in Fig. 1AB, but using only solo-WCGW CpGs⁵. Mean methylation is displayed in consecutive tiles of 10 kb (see Methods). Ordering of tumor samples is the same as in Fig. 1. (C), WGBS DNA methylation visualization at megabase-scale, exactly as in Fig. 1C, but using only solo-WCGW CpGs. Pink coloring indicates common methylation loss (PMDs) as in Fig. 1, although tumor-specific PMD borders vary. A scale bar (100 kb) is shown at the top of each panel. CpG islands are indicated in green.



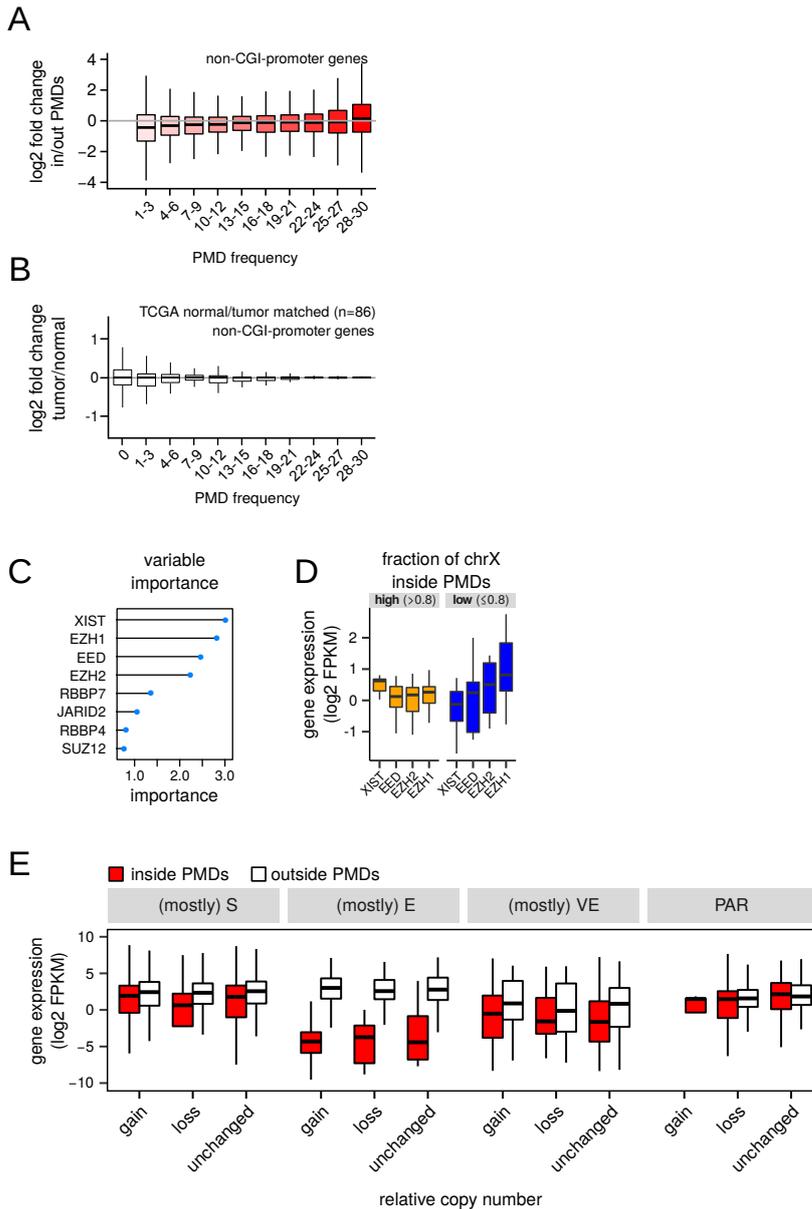
Supplementary Figure 4 | Association between DNA methylation variation and clinical features, PMD calling using solo-WCGW CpGs.

(A), Association between principal component 1 and 2 scores (PC1, PC2, see Fig. 1D) and major pathological subtypes. Significance testing was done as described in ref.⁶. Left panels, ER-status; right panels, AIMS intrinsic subtypes⁷. (B,C), The same analysis as in Fig. 2AB, but using PMDs detected with only solo-WCGW CpGs⁵. Boxplots represent the median and 25th and 75th percentiles, whiskers 1.5 times the interquartile range. (B), Fraction of the genome covered by PMDs. Each dot represents one tumor sample, the boxplot summarizes this distribution. (C), Fraction of the genome covered by PMDs that are common between breast tumors. PMD frequency: the number of tumors in which a genomic region is a PMD. (D), Venn-diagram showing the overlap between the union of all breast cancer PMDs ('all-CpGs') and the union of all breast cancer solo-WCGW PMDs. (E), Bimodal distribution of cross-sample standard deviation of mean methylation in 100 kb genomic windows. Only solo-WCGW CpGs were used to calculate window means. As described in⁵, a mixed gaussian was fitted to determine a cutoff for genome segmentation. (F), Overlaps between the cross-sample s.d. based PMDs (E) and PMDs called on individual samples in this study, using all CpGs.



Supplementary Figure 5 | Gene expression and somatic mutations inside breast cancer PMDs, full cohort.

(A), Gene expression as a function of PMD frequency, as in main Figure 2F, but here extended to all 266 cases of the breast tumor (RNA-seq) transcriptomes cohort². Top, gene expression; bottom, standard deviation. **(B)**, Somatic mutations plotted as a function of PMD frequency, as in main Figure 2G, but here extended to all 560 cases of the breast tumor full genomes cohort¹. Boxplots represent the median and 25th and 75th percentiles, whiskers 1.5 times the interquartile range, outliers are not shown.

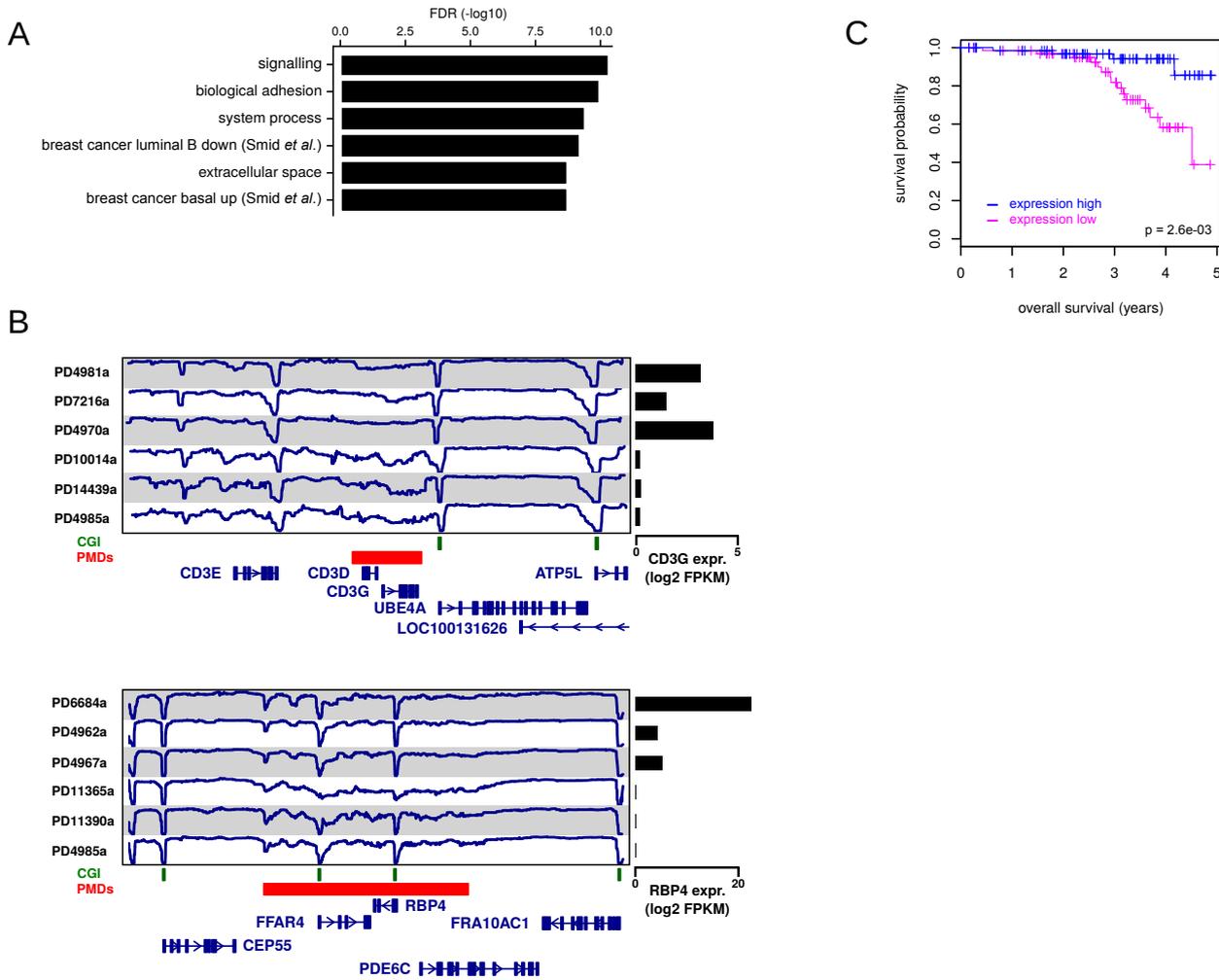


Supplementary Figure 6 | Expression change of non-CGI-promoter genes and X-linked genes inside PMDs, association of X-linked methylation loss with expression of PRC2 subunits.

(A), Expression change of non-CGI-promoter genes inside vs. outside of PMDs, as a function of PMD frequency. (B), Expression change of non-CGI-promoter genes of tumor vs. normal, as a function of PMD frequency. From the TCGA breast cancer dataset, matched tumor/normal pairs were selected. PMD frequency for each gene was taken from our own dataset. (C,D), Multivariate linear regression was performed with expression levels of genes involved in XCI as explanatory variables and PMD abundance on chrX as response variable. The variable importance of each XCI gene is plotted in (C), and their expression levels in two PMD abundance bins is plotted in (D). (E), Expression of X-linked genes when inside or outside PMDs. Genes were grouped according their consensus X-inactivation status (E, escape; S, subject to XCI; VE, variably escaping; PAR, pseudoautosomal region)⁸ and further stratified over their copy-number status (gain, loss, unchanged) as determined previously¹. All boxplots in this figure represent the median and 25th and 75th percentiles, whiskers 1.5 times the interquartile range, outliers are not shown.

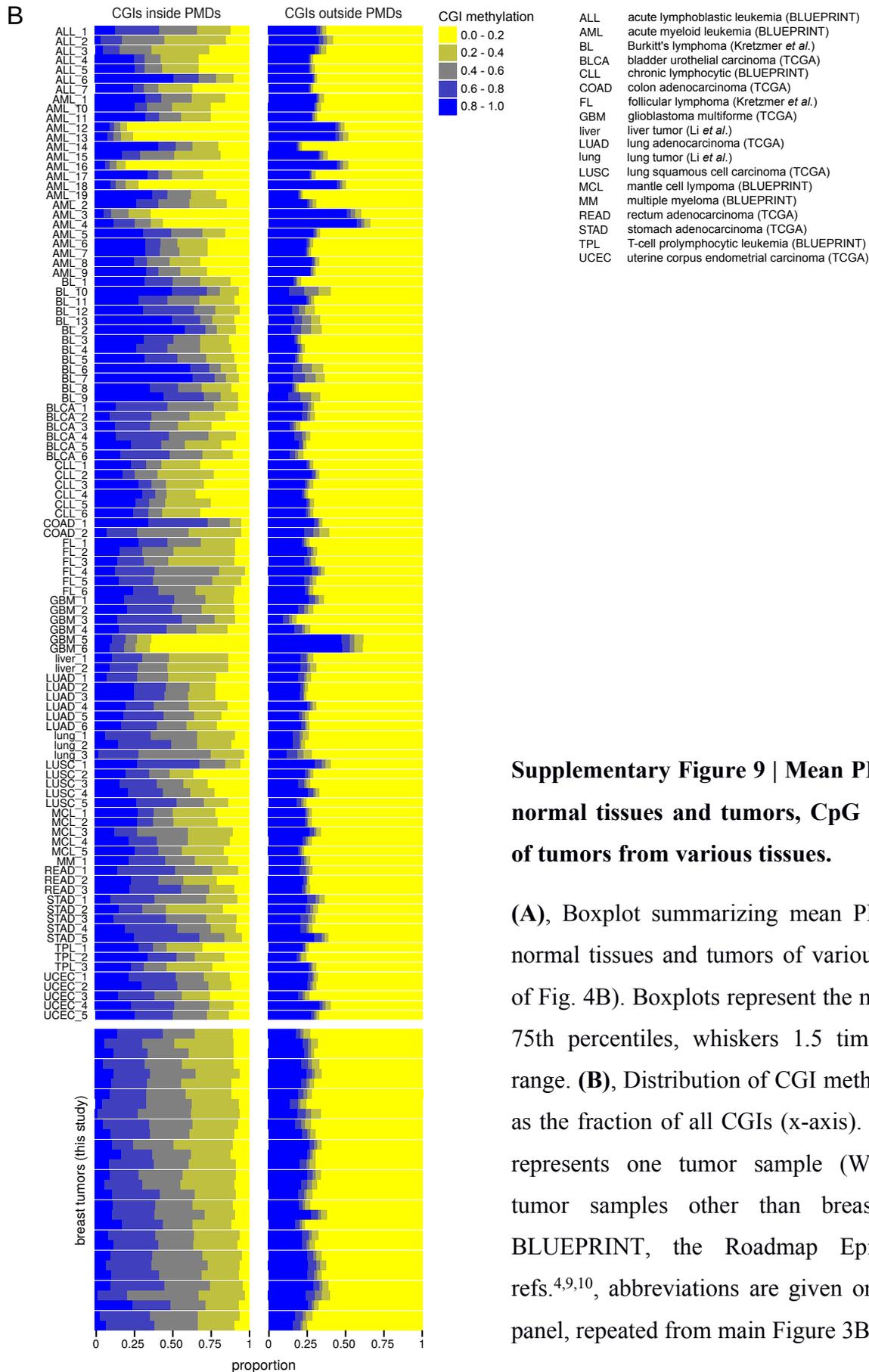
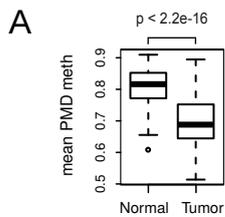
Supplementary Figure 7 | Expression change of tumor suppressor genes and breast cancer driver mutated genes inside PMDs.

(A), Expression change of TSGs/breast cancer driver mutated genes when inside PMDs. 31 of such genes are located inside PMDs in a subset of tumor samples. ‘TSGs all cancers’, genes annotated as TSGs regardless of cancer type; ‘TSGs breast cancer’, genes annotated as TSG in breast cancer; ‘Nik-Zainal breast cancer driver mutations’, genes with driver mutations in breast cancer¹. **(B)**, Examples of genes from panel (A) being repressed when inside PMDs. Blue line, DNA methylation (WGBS); green bars, CGIs; red bars, PMDs. Gene expression (RNA-seq) of the corresponding gene is represented at the right of each panel. **(C)**, Pearson correlation between CGI-promoter methylation and expression. Gene classes are indicated as in panel (A). **(D)**, Expression changes (RNA-seq) of genes in panel (B), breast tumor vs. normal. Data is from an independent cohort (TCGA). Left panels, non-matched normal (n=88) and tumor samples (n=769); right panels, matched normal/tumor samples (n=86). p-values were calculated using a *t*-test. Boxplots represent the median and 25th and 75th percentiles, whiskers 1.5 times the interquartile range.



Supplementary Figure 8 | Gene set enrichment analysis (GSEA) of genes downregulated when inside PMDs.

(A), Gene set enrichment analysis (GSEA) of genes downregulated when inside PMDs (>2.5 log₂-fold, 400 genes, Supplementary Table 4). (B), Examples of downregulated genes inside PMDs. CD3D encodes the gamma polypeptide of the T-cell receptor-CD3 complex (gene sets ‘signalling’, ‘adhesion’, and ‘breast cancer luminal B down’); RBP4 encodes retinol binding protein 4 (gene set ‘signalling’, and ‘breast cancer luminal B down’). Blue line, DNA methylation (WGBS); green bars, CGIs; red bars, PMDs. Gene expression (RNA-seq) of the corresponding gene is represented at the right of each panel. (C), Overall survival of patient groups stratified according expression of the 400 PMD-downregulated genes (see Methods).



Supplementary Figure 9 | Mean PMD methylation of normal tissues and tumors, CpG island methylation of tumors from various tissues.

(A), Boxplot summarizing mean PMD methylation of normal tissues and tumors of various tissues (summary of Fig. 4B). Boxplots represent the median and 25th and 75th percentiles, whiskers 1.5 times the interquartile range. (B), Distribution of CGI methylation, represented as the fraction of all CGIs (x-axis). Each horizontal bar represents one tumor sample (WGBS). Top panel, tumor samples other than breast cancer (TCGA, BLUEPRINT, the Roadmap Epigenomics Project, refs.^{4,9,10}, abbreviations are given on the right); bottom panel, repeated from main Figure 3B for comparison.

SUPPLEMENTARY REFERENCES

1. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 1–20 (2016).
2. Smid, M. *et al.* Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. *Nature Communications* **7**, 12910 (2016).
3. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences* **107**, 16910–16915 (2010).
4. Schultz, M. D. *et al.* Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–6 (2015).
5. Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nature Genetics* **50**, 591–602 (2018).
6. Le, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. of Statistical Software* **25**, 1–18 (2008).
7. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8418–8423 (2003).
8. Balaton, B. P., Cotton, A. M. & Brown, C. J. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biology of Sex Differences* **6**, 35 (2015).
9. Kretzmer, H. *et al.* DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. *Nature genetics* **47**, 1316–25 (2015).
10. Li, X., Liu, Y., Salz, T., Hansen, K. D. & Feinberg, A. Whole genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver. *Genome Research* **26**, 1730–1741 (2016).