Supplementary Information

TABLE OF CONTENTS

Supplementary Tables

Supplementary Table 1. CpG and GpC methylation rates of samples treated with combinations of methylations.

Supplementary Table 2. Nanopore sequencing results of training and testing sets.

Supplementary Table 3. Confusion matrix of CpG and GpC methylation calling

Supplementary Table 4. Relative accuracy and call rates for notable genomic contexts

Supplementary Table 5. Individual nanopore sequencing run metrics

Supplementary Table 6. Summary of Single-read Combinatorial Clustering on Genes by Expression

Supplementary Table 7. Summary of Single-read Combinatorial Clustering on Genes by Histone Modification

Supplementary Table 8. Summary of allele-specific differential epigenetic regions by genomic context in X chromosome

Supplementary Table 9. Summary of allele-specific differential epigenetic regions by genomic context in autosomes

Supplementary Table 10. Summary of structural variations detected in breast cell lines.

Supplementary Data

Supplementary Data 1. nanoNOMe accessibility peaks in GM12878

Supplementary Data 2. CTCF binding sites in GM12878

Supplementary Data 3. Estimated protein-bound regions near a subset of gene TSS in GM12878

Supplementary Data 4. Protein binding stratified by promoter epigenetic signatures

Supplementary Data 5. Allele-specific DMRs and DARs in GM12878

Supplementary Data 6. Gene promoter regions with allele-specific DMRs and DARs in GM12878

Supplementary Data 7. Heterozygous structural variations in GM12878

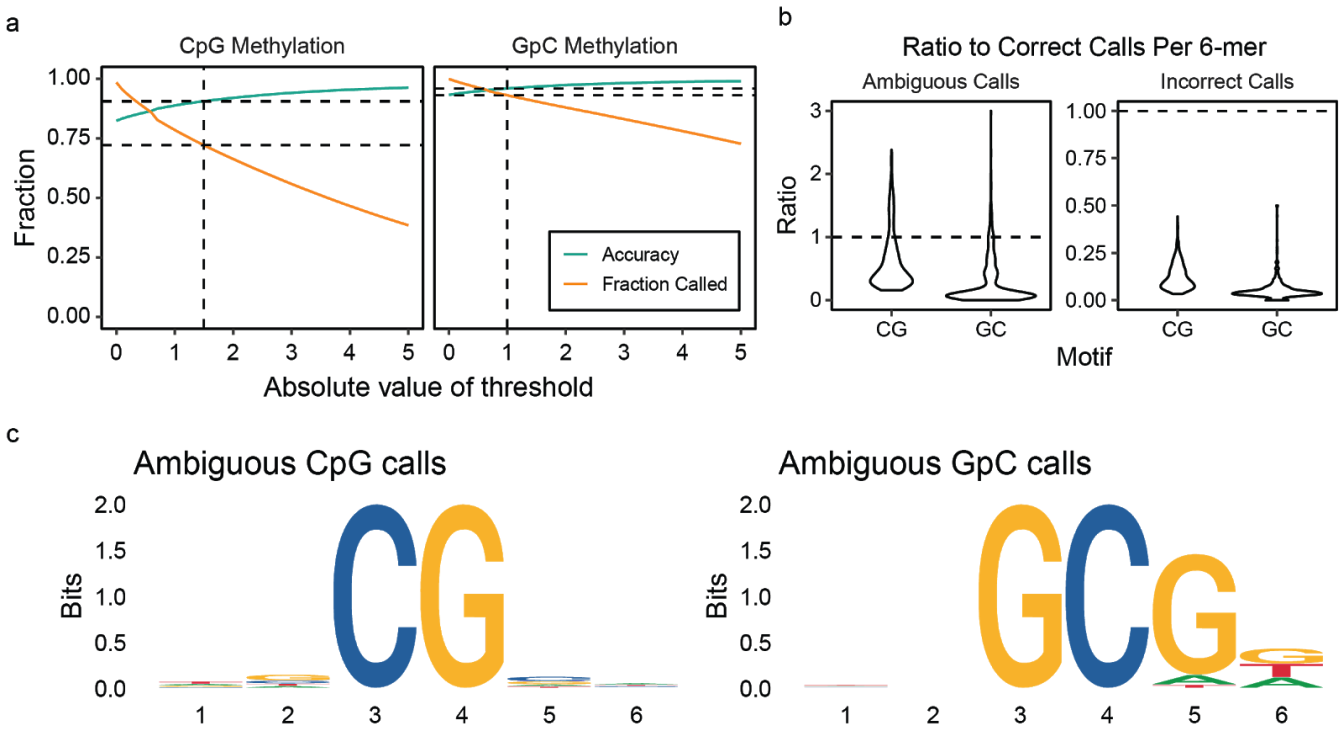
Supplementary Data 8. DMRs and DARs in MCF-7 and MDA-MB-231 in comparison to MCF-10A

Supplementary Data 9. Summary of DMRs and DARs with respect to genomic contexts and structural variations

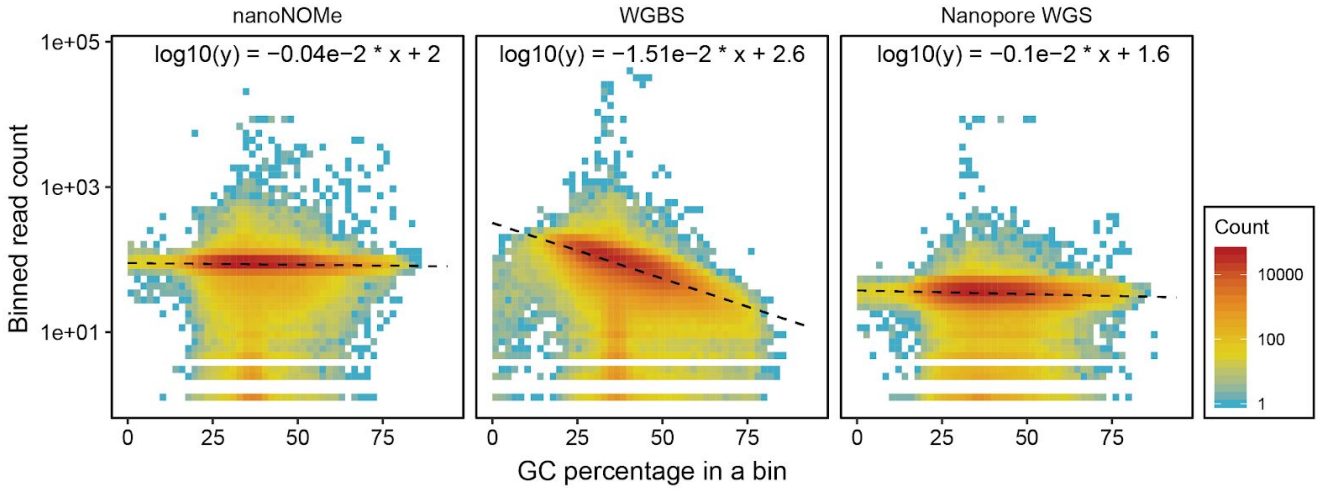
Supplementary Data 10. Structural variations in MCF-10A, MCF-7, and MDA-MB-231

Supplementary Data 11. Promoter epigenetic signatures of differentially expressed genes in MCF-10A, MCF-7, and MDA-MB-231

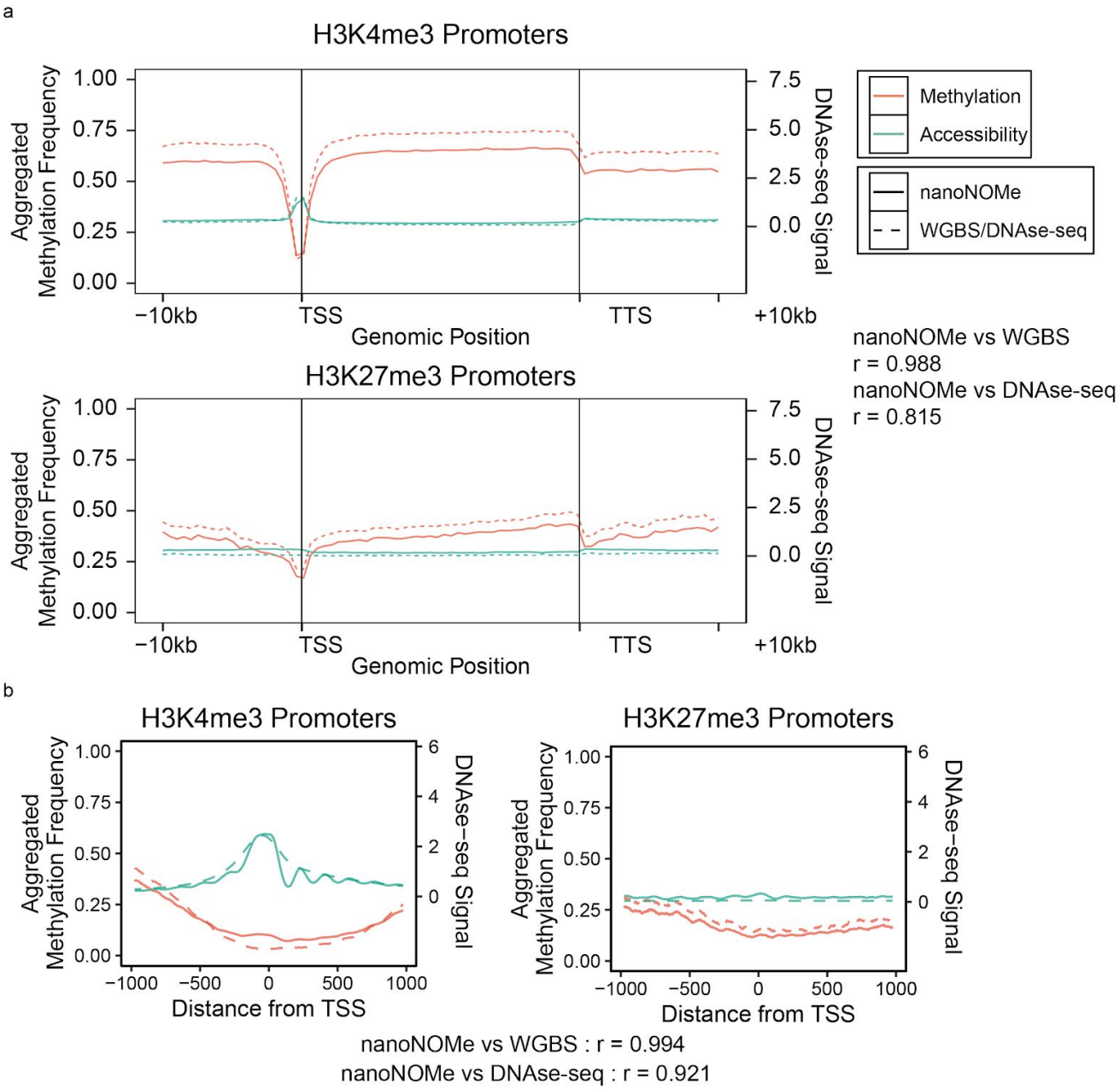
Supplementary Data 12. Protein binding regions near differentially expressed genes in MCF-10A, MCF-7, and MDA-MB-231



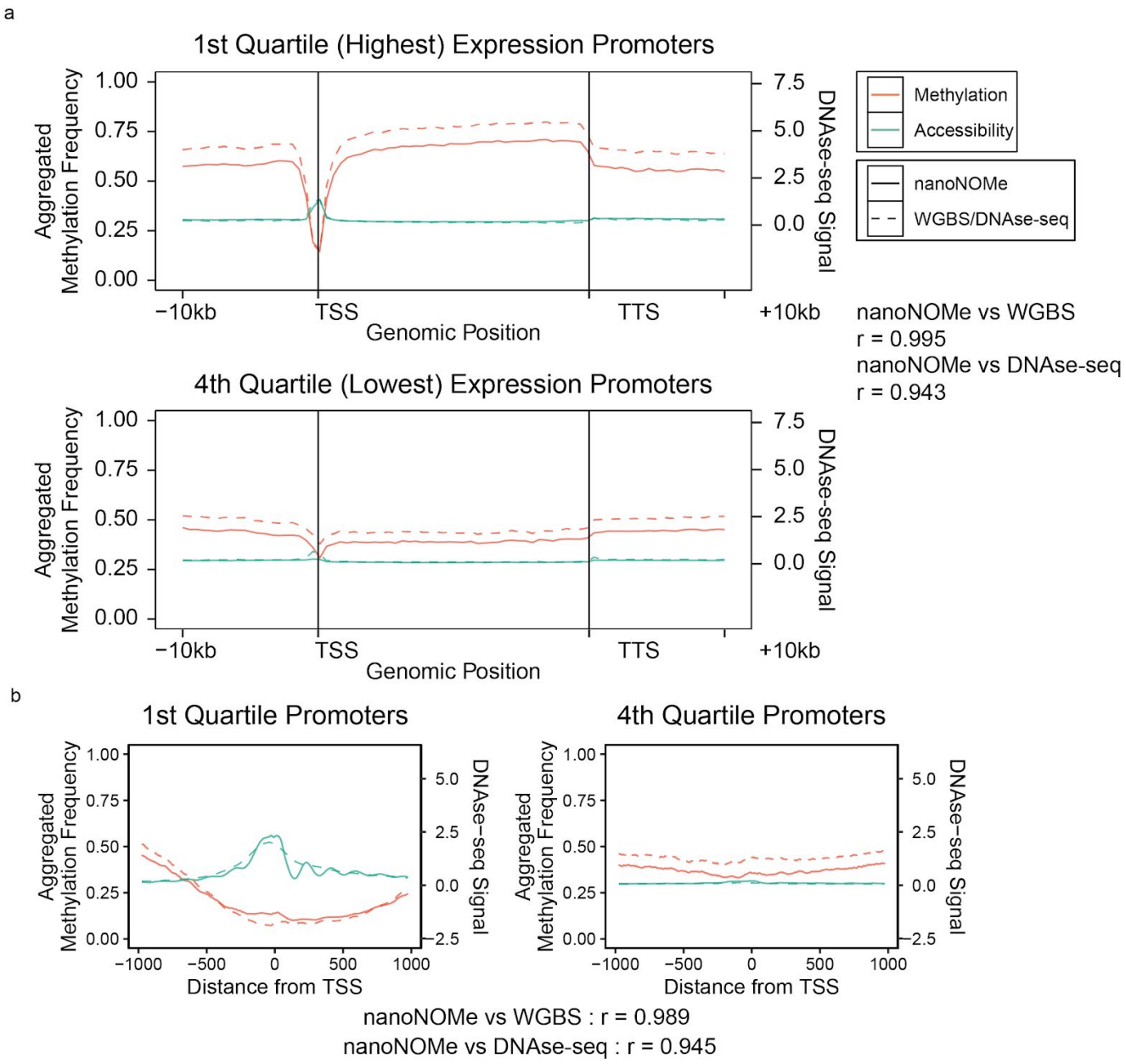
Supplementary Figure 1. Performance of CpG and GpC dual methylation calling. (a) Across a range of absolute values of methylation calling thresholds, the fraction of k-mers passing the threshold filter and the fraction of k-mers from which methylation was correctly called were calculated to decide the optimal threshold. **(b)** The ratio of ambiguous and incorrectly called calls to correctly called calls for every 6-mer with the methylation motif in the center (3th position for CpG and 4th position for GpC), and **(c)** motif analysis of enriched ambiguous 6-mers from (d) in (left) CpG calls and (right) GpC calls, showing enrichment in GCG motifs in GC calling, which are removed in our pipeline.



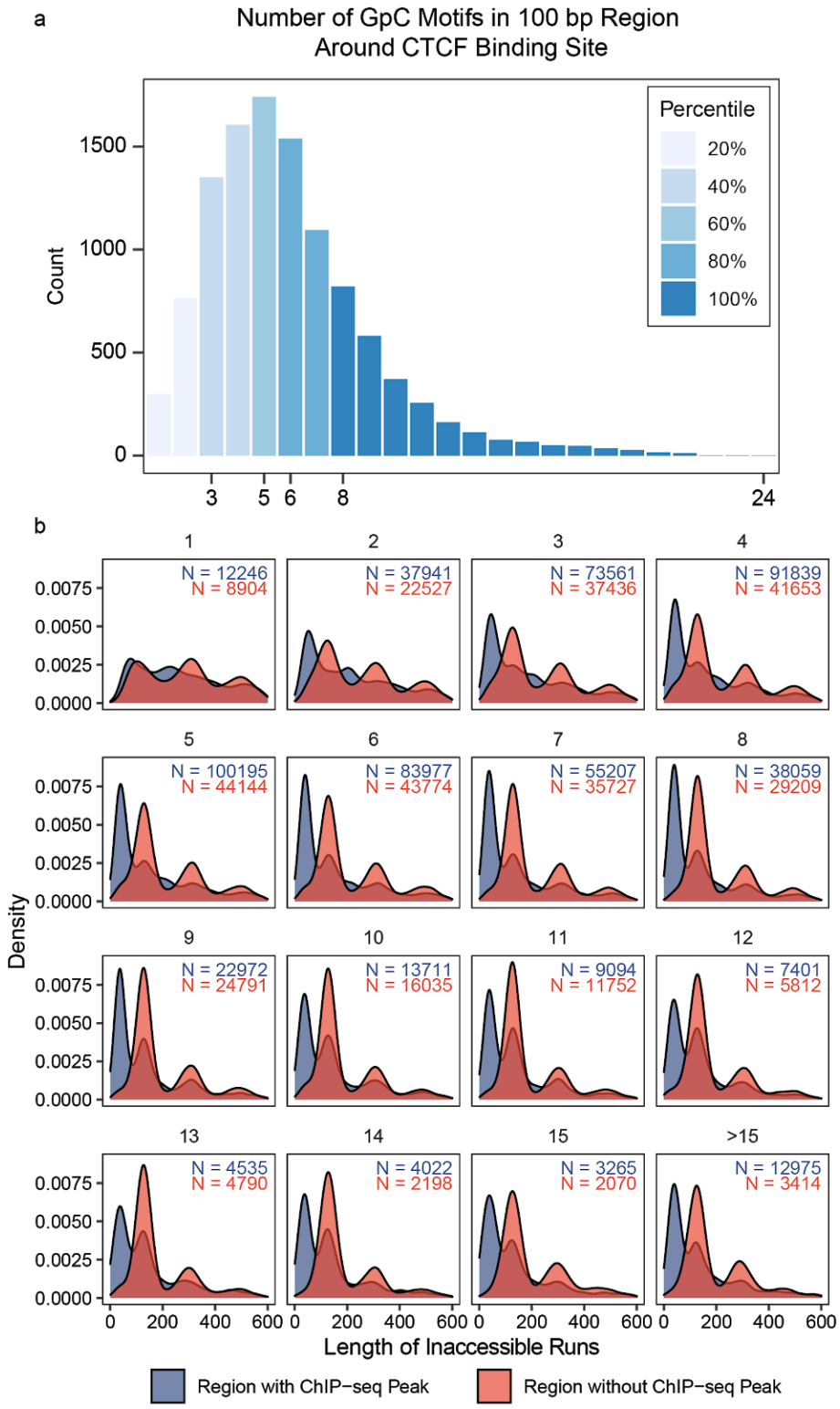
Supplementary Figure 2. GC-content bias of coverage in sequencing methods. Binned coverage versus GC percentage in corresponding bins of nanoNOMe, WGBS (ENCODE), and WGS by nanopore sequencing, along with regression models representing the degree of dependence of coverage on GC-content in the form of the slope of the regression model.



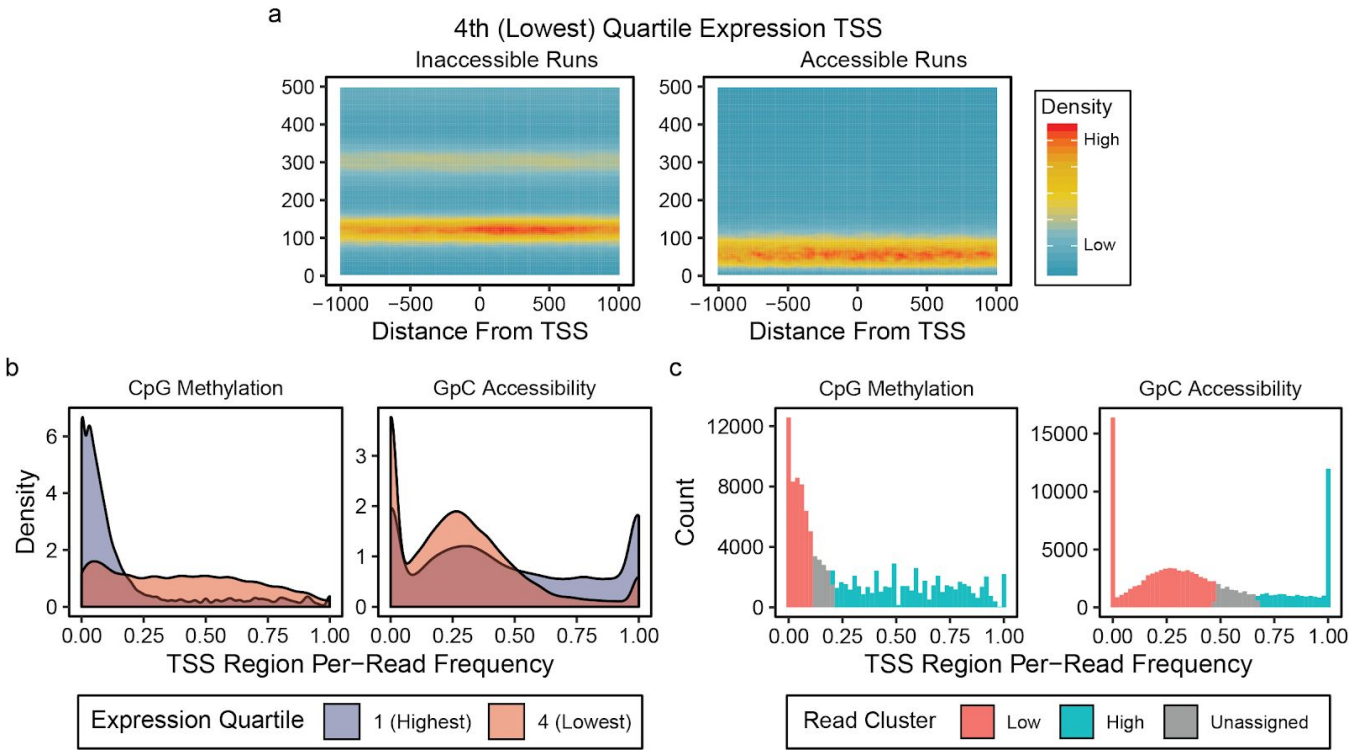
Supplementary Figure 3. Bulk NanoNOMe profiles in Genes With Histone Marks. Metaplots at TSS with euchromatic (H3K4me3) and heterochromatic (H3K27me3) histone modifications **(a)** across the lengths of the genes and **(b)** in the promoter regions.



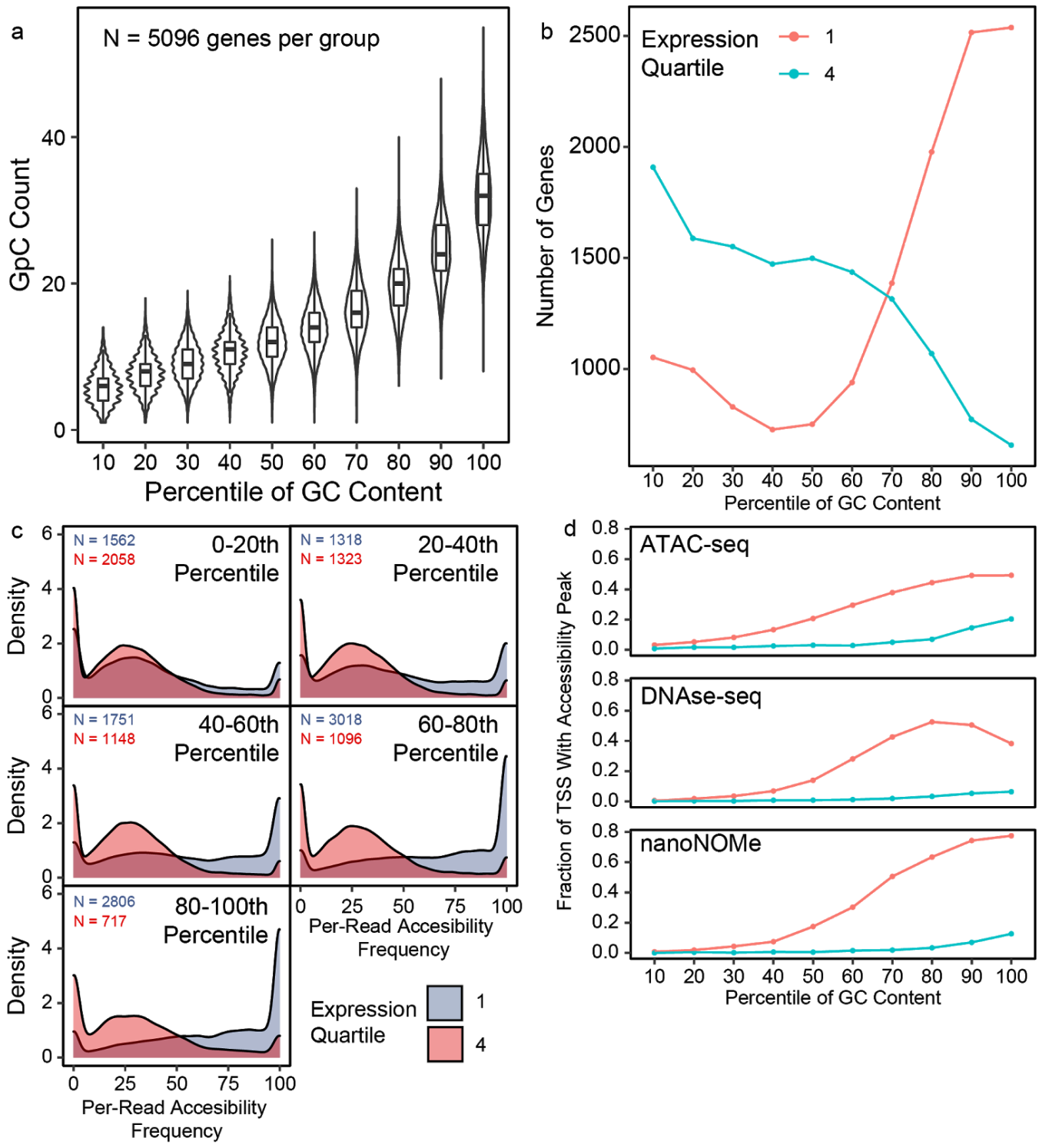
Supplementary Figure 4. Bulk NanoNOMe profiles in Genes by Expression Quartile. Metaplots at TSS with highest (1st quartile) and lowest (4th quartile) expression **(a)** across the lengths of the genes and **(b)** in the promoter regions.



Supplementary Figure 5. Footprint detection stratified by number of GpCs. a) Histogram of number of GpC motifs in the 100bp window centered at potential CTCF binding sites in the genome. **b)** Density distributions of inaccessible runs at the CTCF binding sites as in Supplemental Figure 5b, stratified by GpC number in the 100bp window; note that above 3 GpCs the distributions show the same behavior.

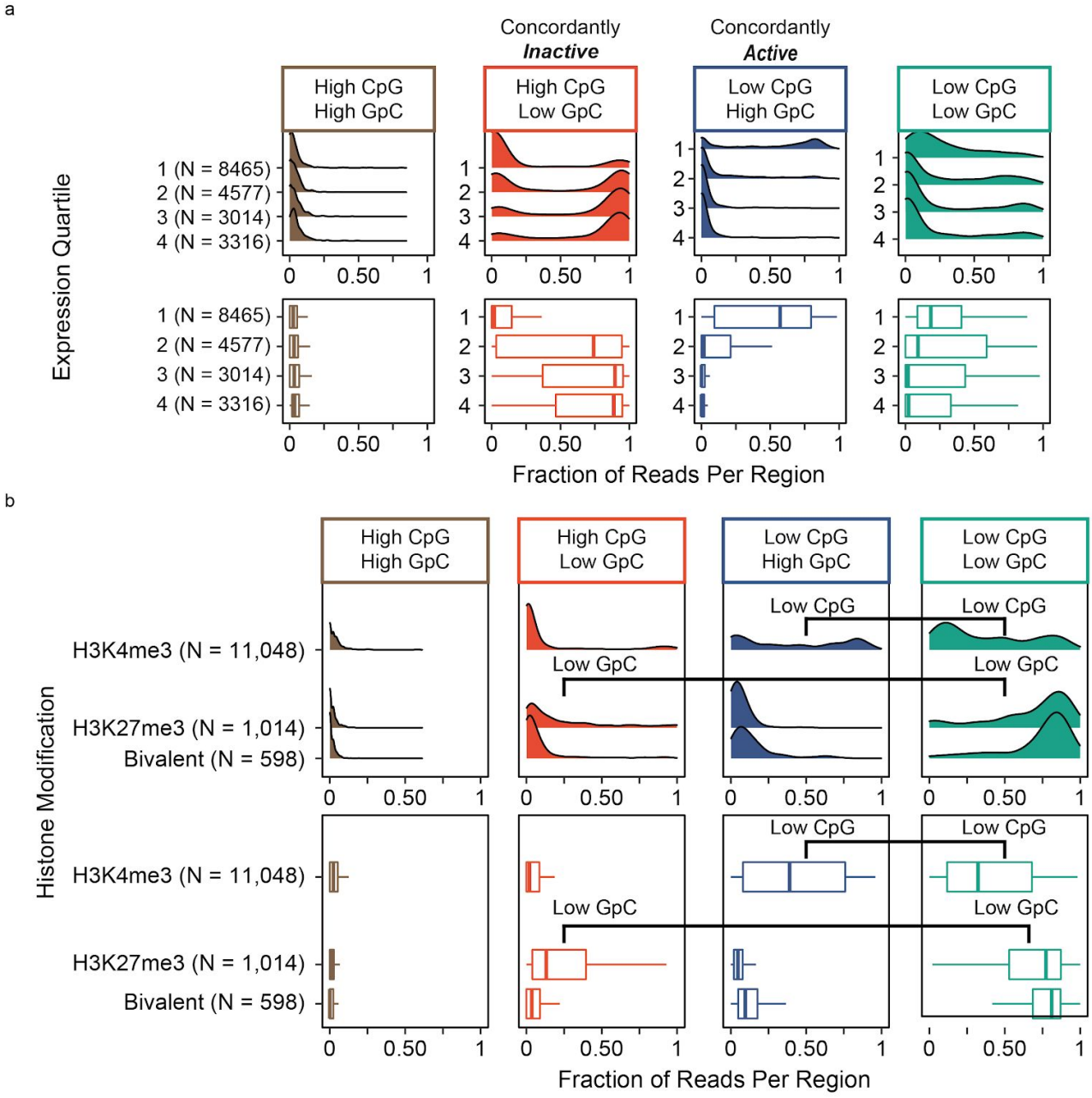


Supplementary Figure 6. Single-read epigenetic assessment on transcription start sites. (a) Heatmaps of lengths of inaccessible and accessible runs on individual reads with respect to the distance to transcription start sites of lowly expressed genes. **(b)** Distributions of per-read CpG methylation frequency in 1kb region around TSS and GpC accessibility frequency in 200 bp region around TSS, stratified by expression, showing that more reads are demethylated and accessible with an increase in expression. **(c)** We used the windows in (b) to cluster the reads based on methylation around TSS into two groups for each feature, high frequency (blue) and low frequency (red), resulting in four possible combinatorial signatures for each read at TSS.

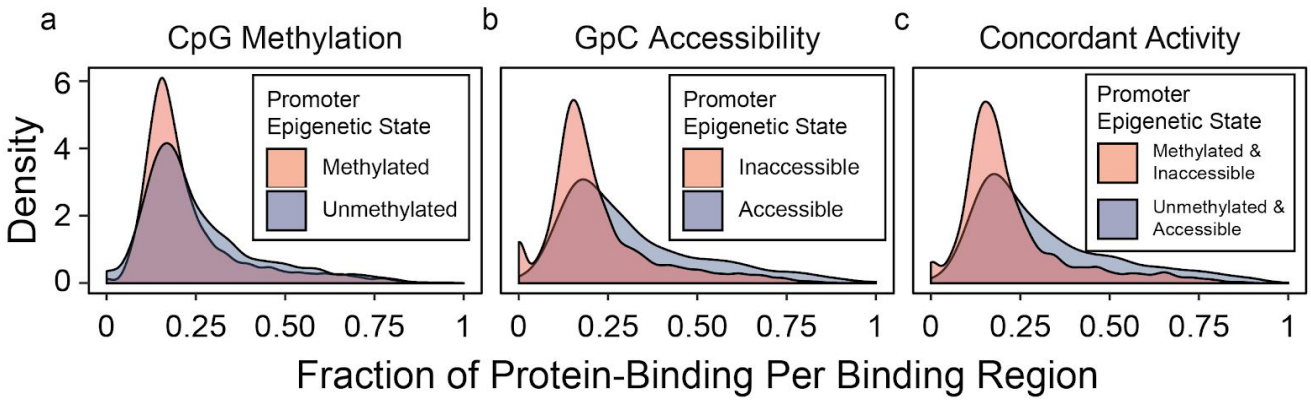


Supplementary Figure 7. TSS Regions stratified by GC density. a) Plot of number of GpCs versus GC percentile in the 200bp window centered at transcriptional start sites. Data are presented as median values, interquartile range (IQR), and 1.5X IQR, as well as density distributions.
b) Number of genes in highest or lowest expression quartile versus GC percentile in the TSS region. **c)** Density distribution of GpC accessibility frequency in 200 bp region around TSS, as in Supplementary Figure 7b, with color representing expression quartile. Results are faceted by GC percentile in the

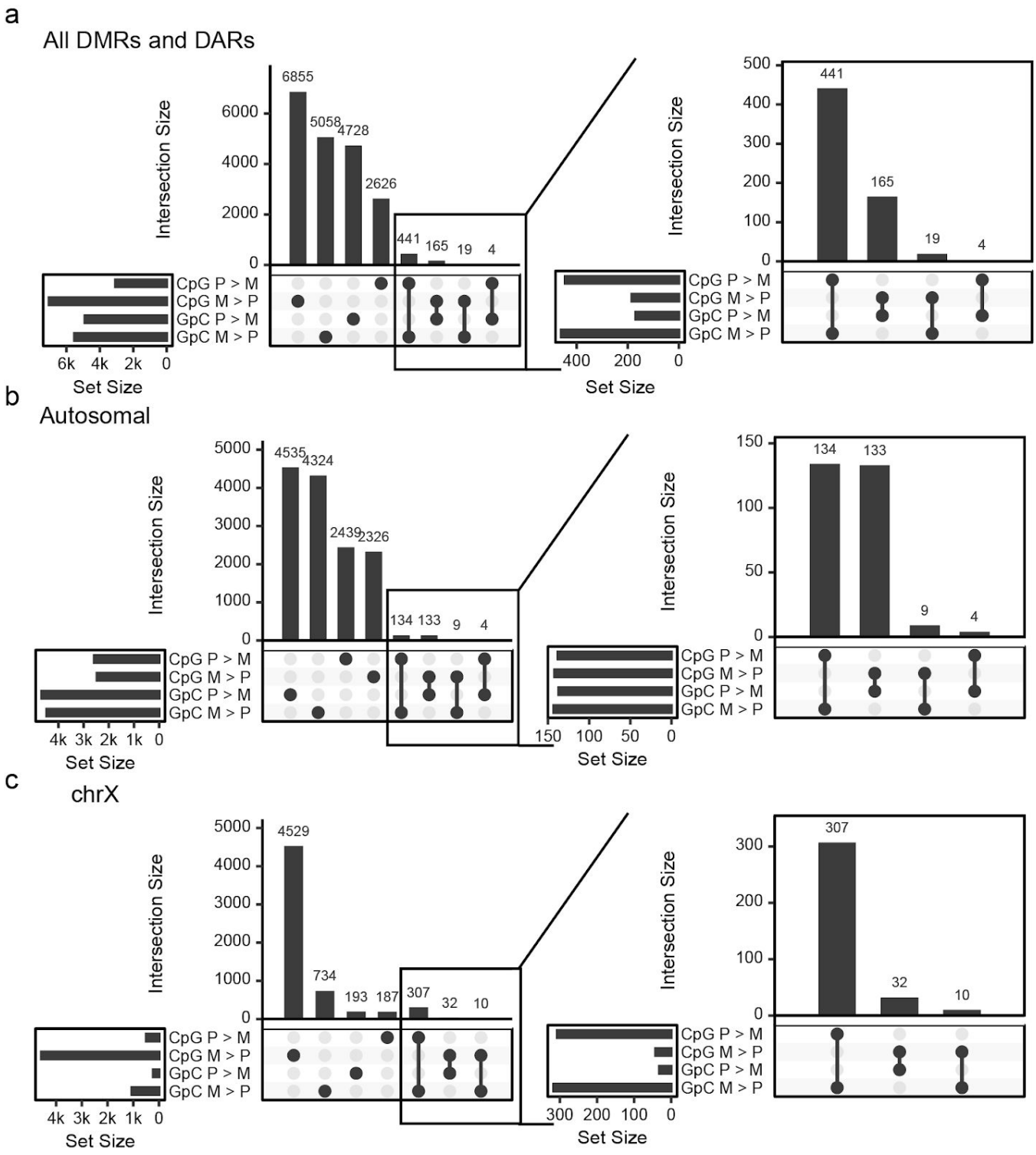
200bp window. **d)** Number of ATAC-seq, DNAse-seq or nanoNOMe accessibility peaks in TSS region for either highest or lowest expression quartile compared to GC% in TSS region.



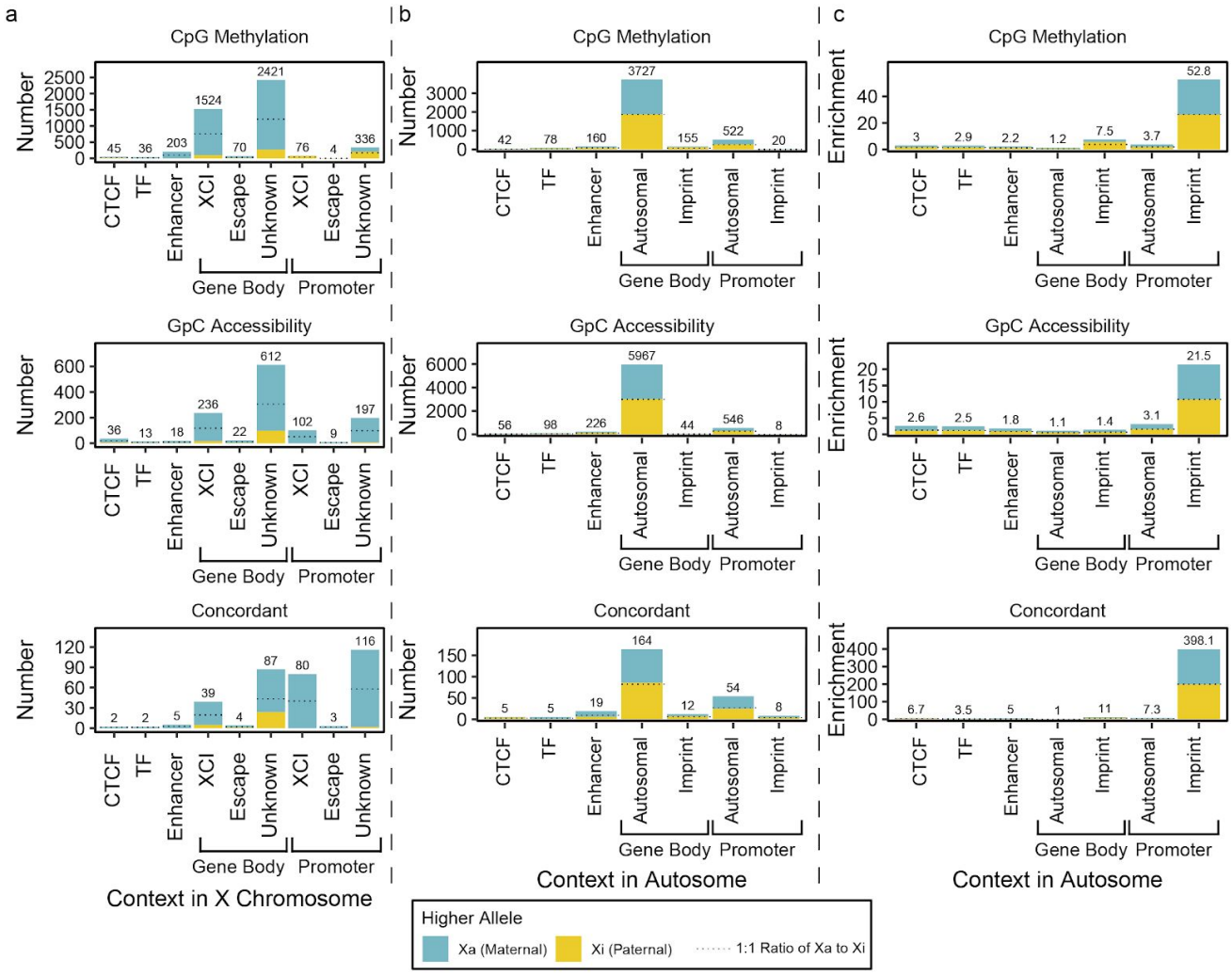
Supplementary Figure 8. Assessment of read-level combinatorial epigenetic signatures of TSS. Fractions of read-level combinatorial epigenetic signatures at TSS were calculated for each TSS in a subset of 1,000 genes per group and compared **(a)** by expression quartiles, showing that with increasing expression more genes have higher fraction of combinatorially active reads and less fraction of inactive reads (Supplementary Table 6), and **(b)** by promoter histone modification, showing that reads at euchromatic H3K4me3 genes are demethylated and reads at heterochromatic H3K27me3 genes are inaccessible (Supplementary Table 7). Data are presented as median values, interquartile range (IQR), and 1.5X IQR, as well as density distributions.



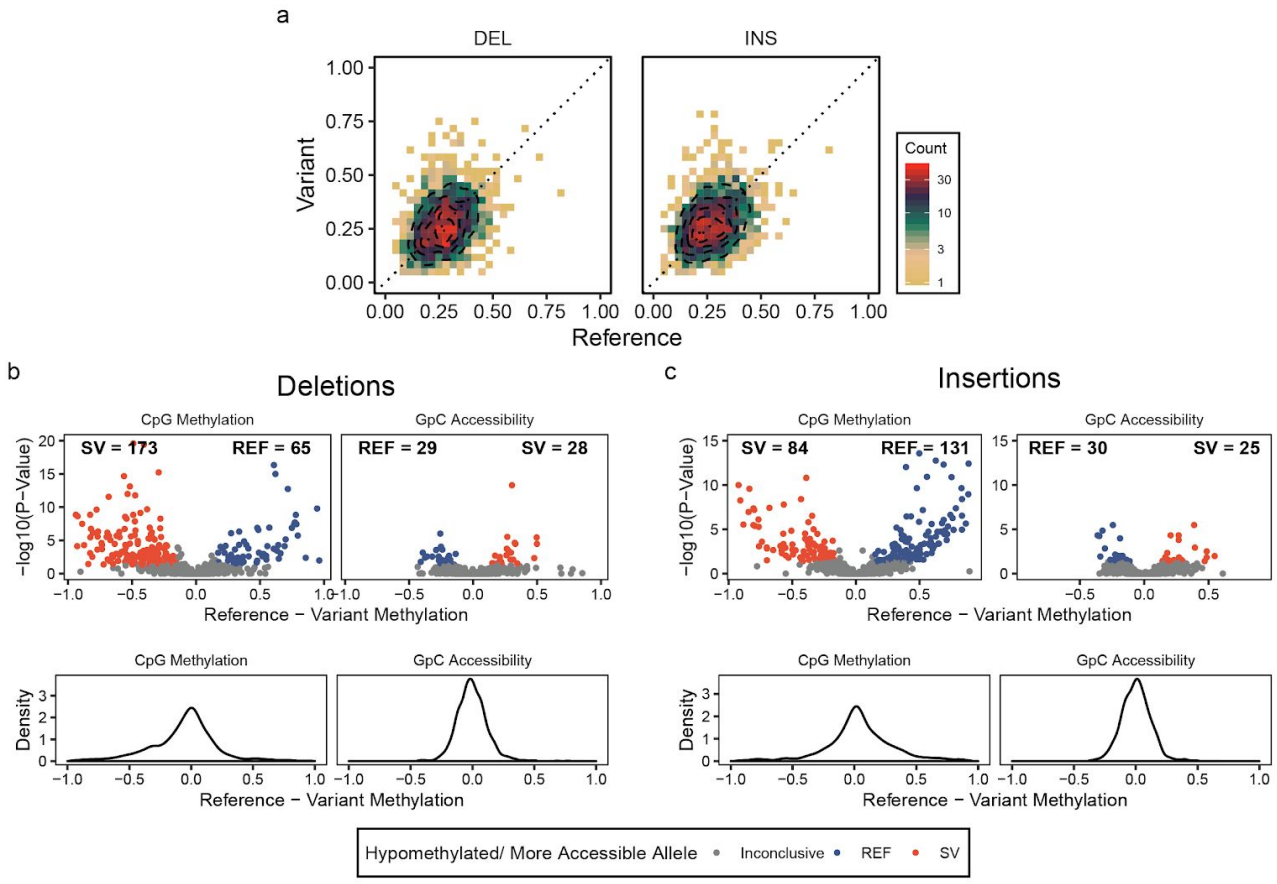
Supplementary Figure 9. Comparisons of predicted TF-binding with respect to promoter epigenetic signatures. For each of the 1,000 randomly selected genes, reads were split into two groups based on their epigenetic signature: (left) CpG methylation, (center) accessibility (right) concordance of both. Then protein-binding regions were predicted within 10kb of the TSS and the fraction of protein-occupied reads was calculated (see **Methods**). The relationship between protein-binding and promoter epigenetic signature is assessed by the distribution of fractions of protein-bound reads between read groups based on (**a**) CpG methylation, (**b**) GpC accessibility, and (**c**) concordant epigenetic activity



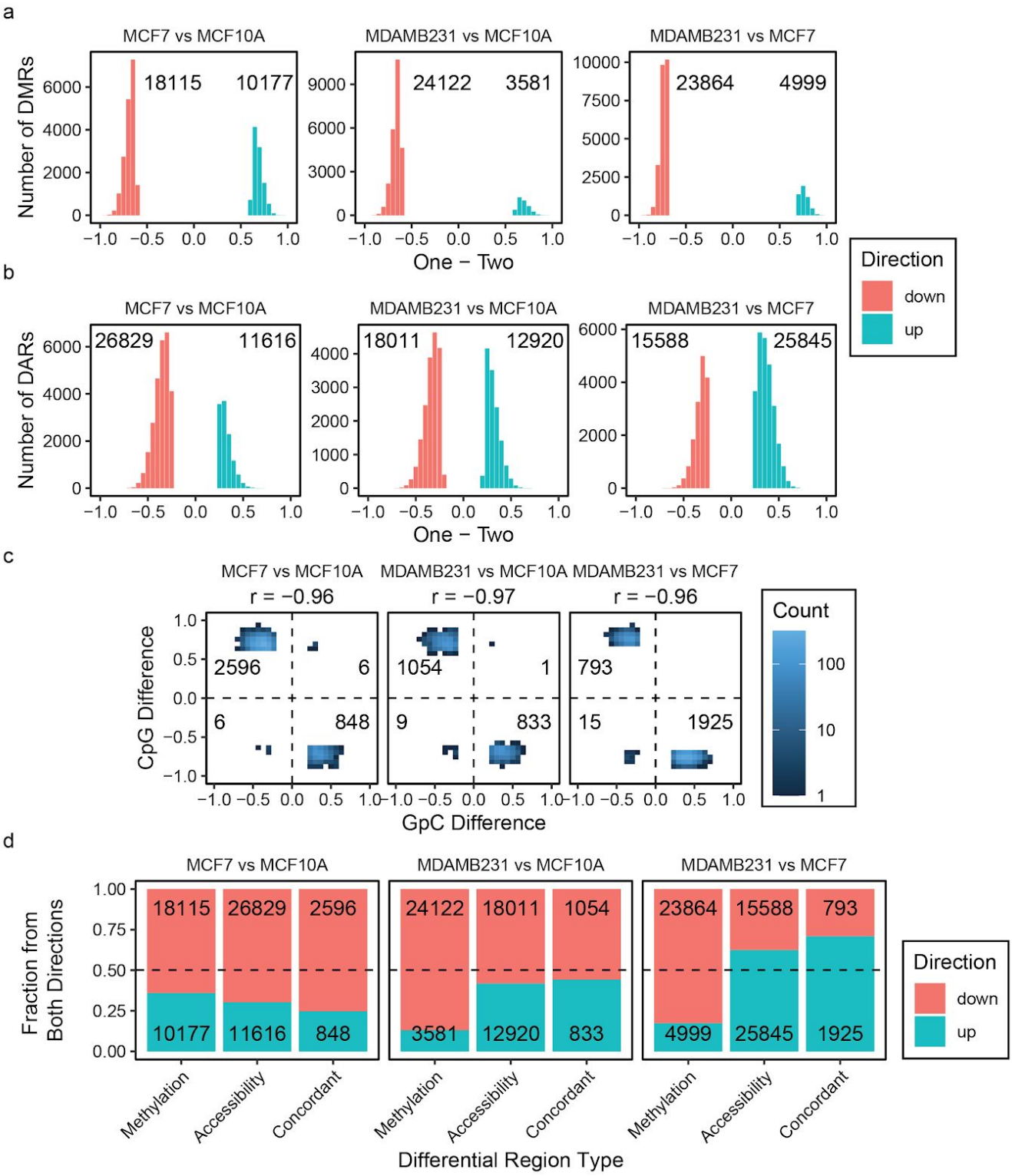
Supplementary Figure 10. Upset plots of differentially methylated and differentially accessible regions between alleles in GM12878. Directions of DMRs/DARs as well as their overlaps were observed using upset plots, across the genome (left), as well as separated by autosomes (middle) and X chromosome (right) **(a)** across the entire genome, **(b)** in autosomes, and **(c)** in the X chromosome..



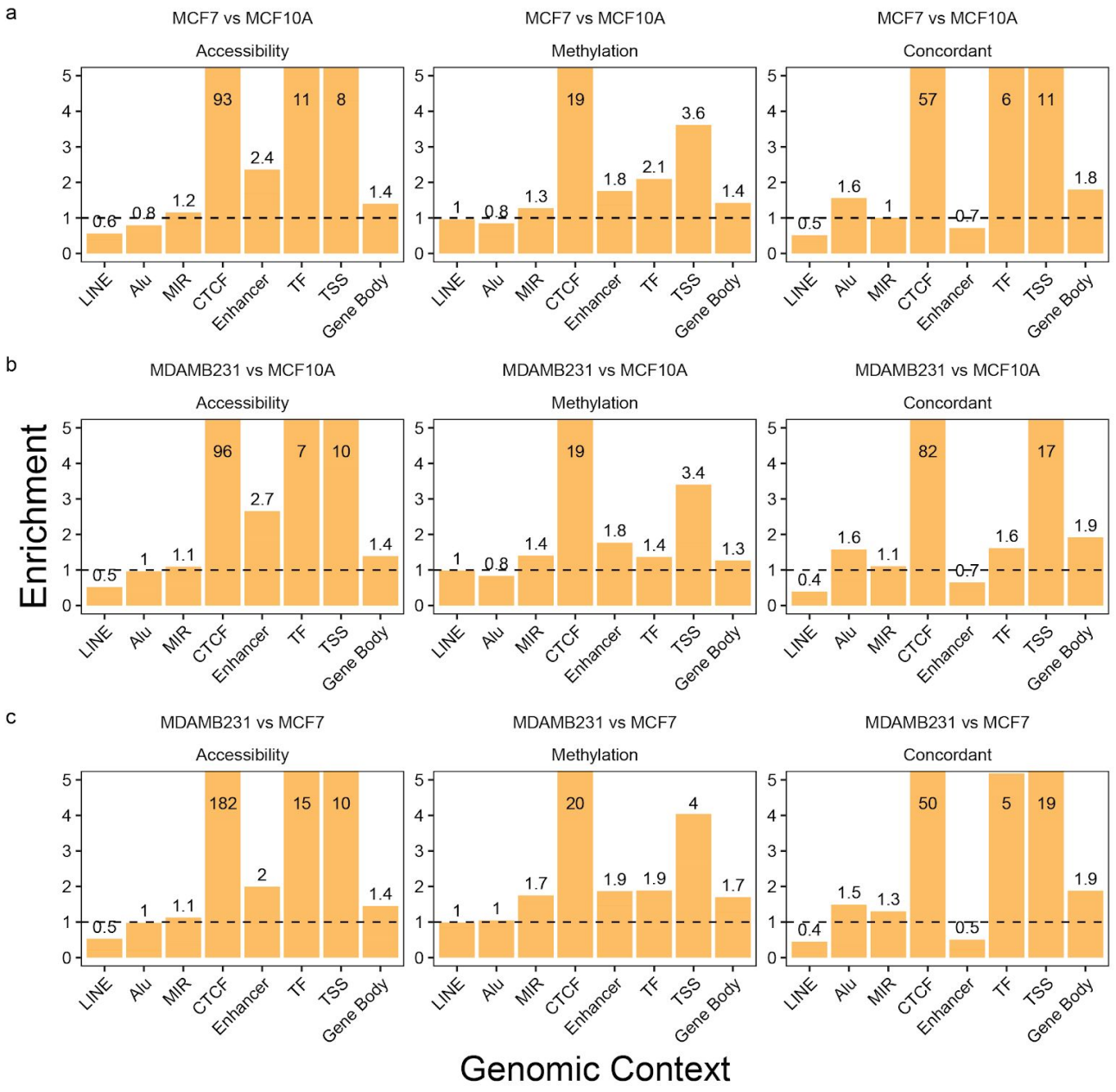
Supplementary Figure 11. Genome-context assessment of allele-specific DMRs and DARs. (a) Numbers of differential regions in each genomic context in X chromosome, **(b)** Numbers of differential regions in genomic contexts in autosomes, and **(c)** enrichments of differential regions in autosomes. All numbers and enrichment values are also provided in Supplementary Tables 8 and 9.



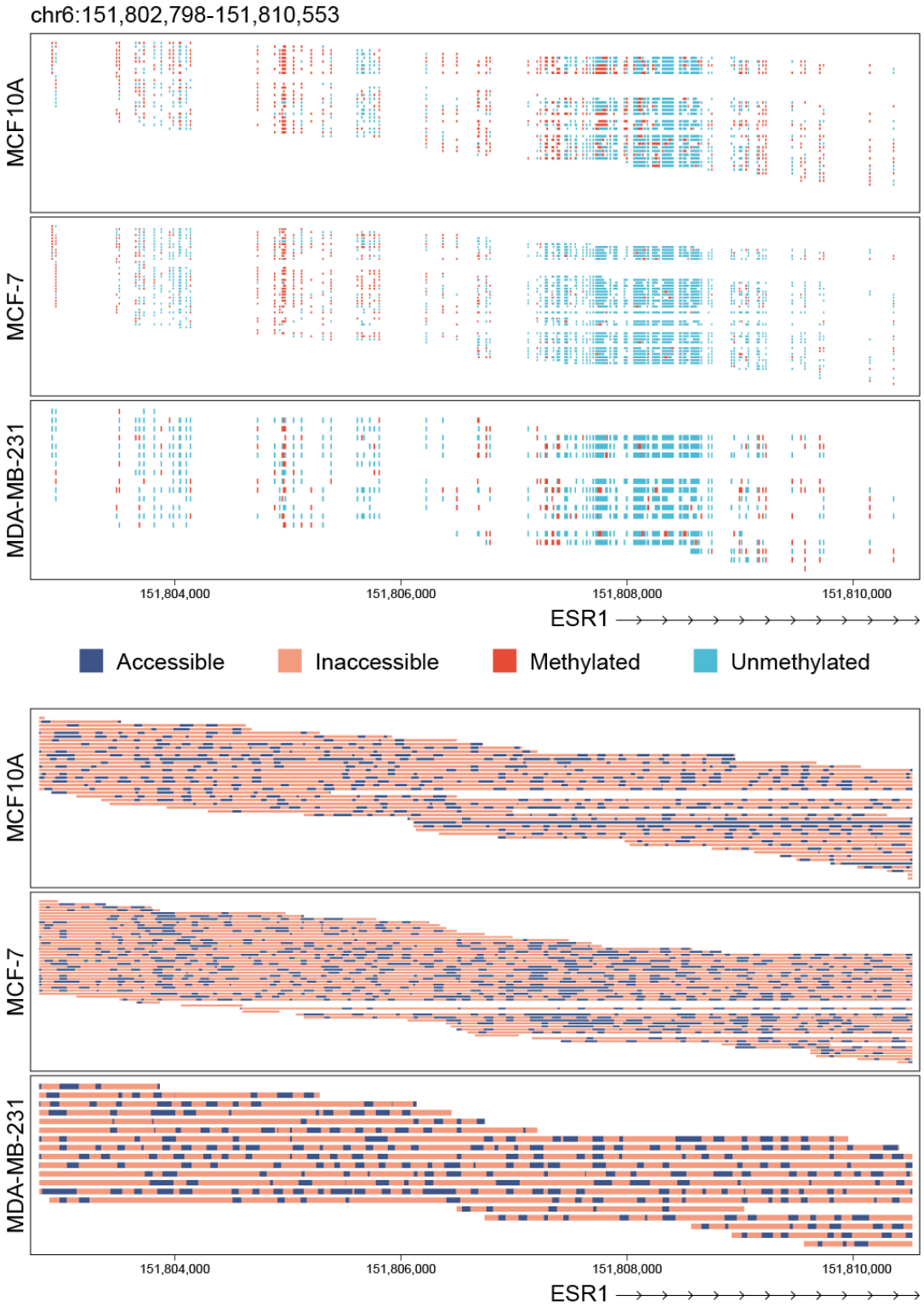
Supplementary Figure 12. Epigenetic comparison of heterozygous structural variations (a) Pair-wise comparison of GpC methylation around breakpoints of heterozygous SVs, with the allele with the SV on the y-axis and the allele without the SV on the x-axis. The difference of CpG methylation and GpC methylation in variant alleles of SVs in comparison to the reference alleles, with one-sided Fisher's exact tests and presented as volcano plots for **(b)** heterozygous deletions and **(c)** heterozygous insertions.



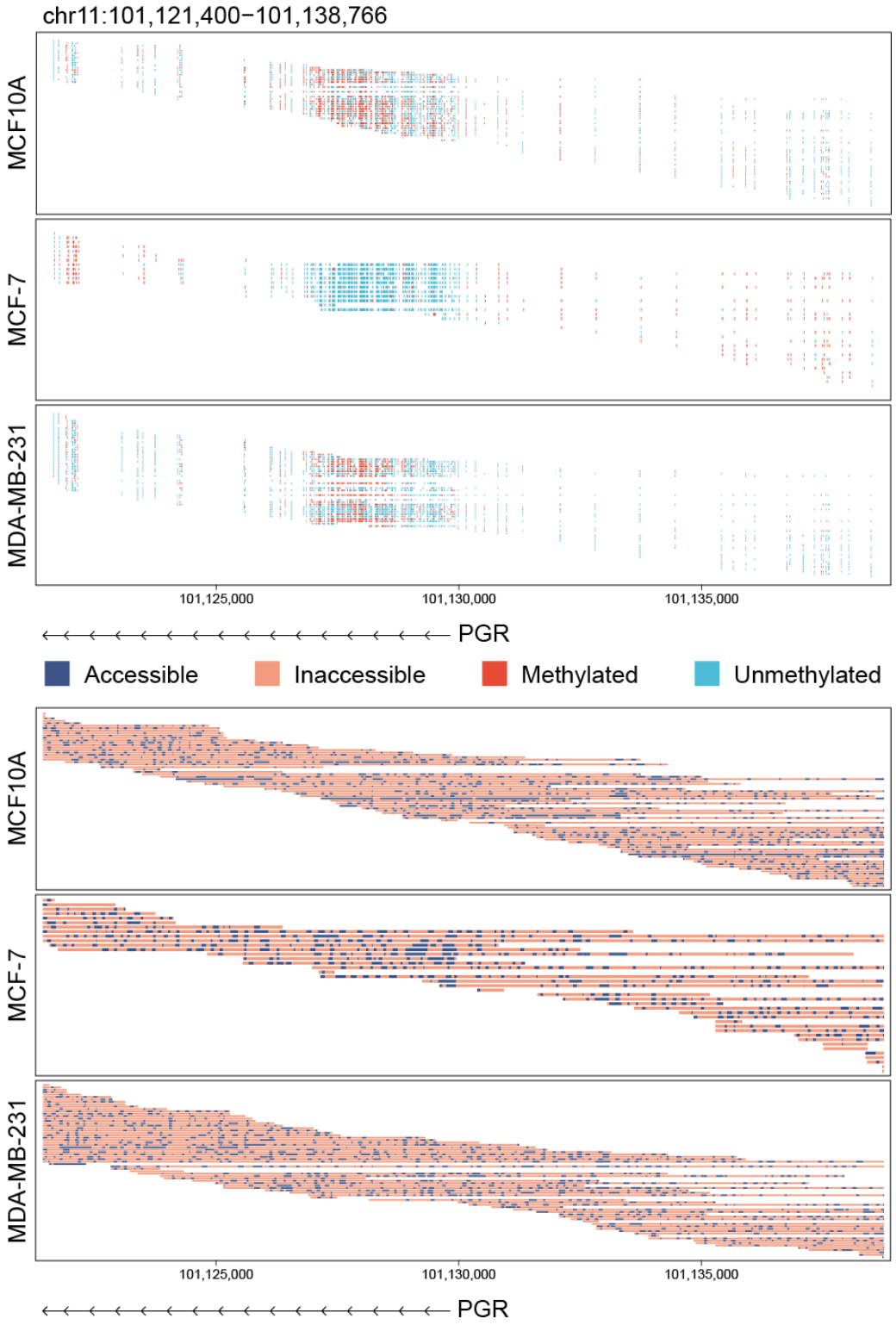
Supplementary Figure 13. Bulk genome-wide differential methylation and accessibility analysis on breast cancer models. Histograms of the difference between the three breast cell lines in **(a)** in differentially methylated regions and **(b)** in differentially accessibility regions. **(c)** Comparison of average methylation to average accessibility in concordantly differential regions. **(d)** Comparisons of the directions of DMRs, DARs, and concordantly differential regions. Dotted line is the 1:1 ratio.



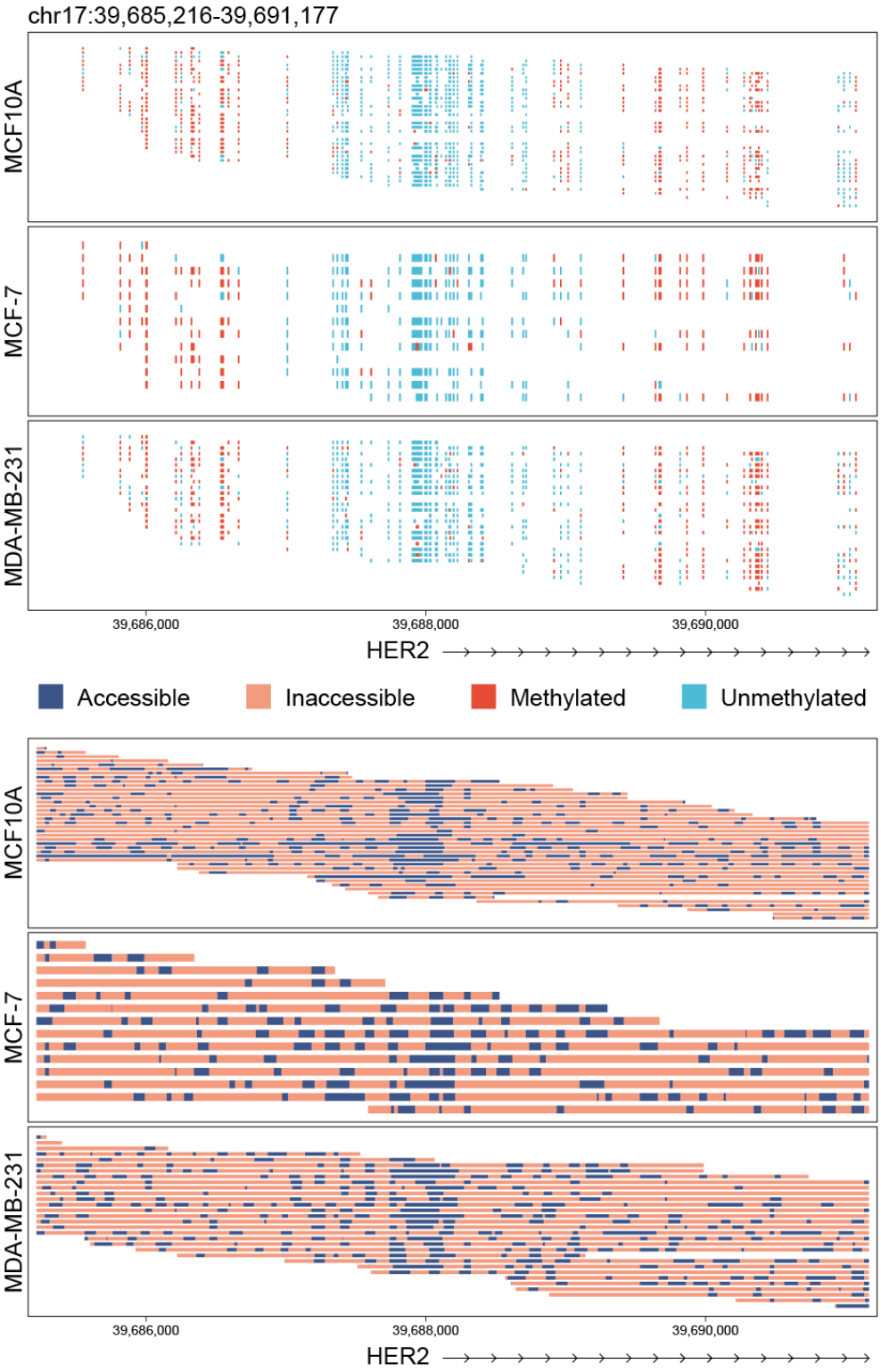
Supplementary Figure 14. Enrichment of differential epigenetic regions in various genomic contexts. Enrichment of DMRs, DARs, and concordantly differential regions are calculated for each genomic context against the abundance across the genome for comparisons of **(a)** MCF-7 vs. MCF-10A, **(b)** MDA-MB-231 vs. MCF-10A, and **(c)** MDA-MB-231 vs. MCF-7. These values, along with the numbers and directions and including structural variation junctions as additional genomic contexts, can be found in Supplementary Data 10.



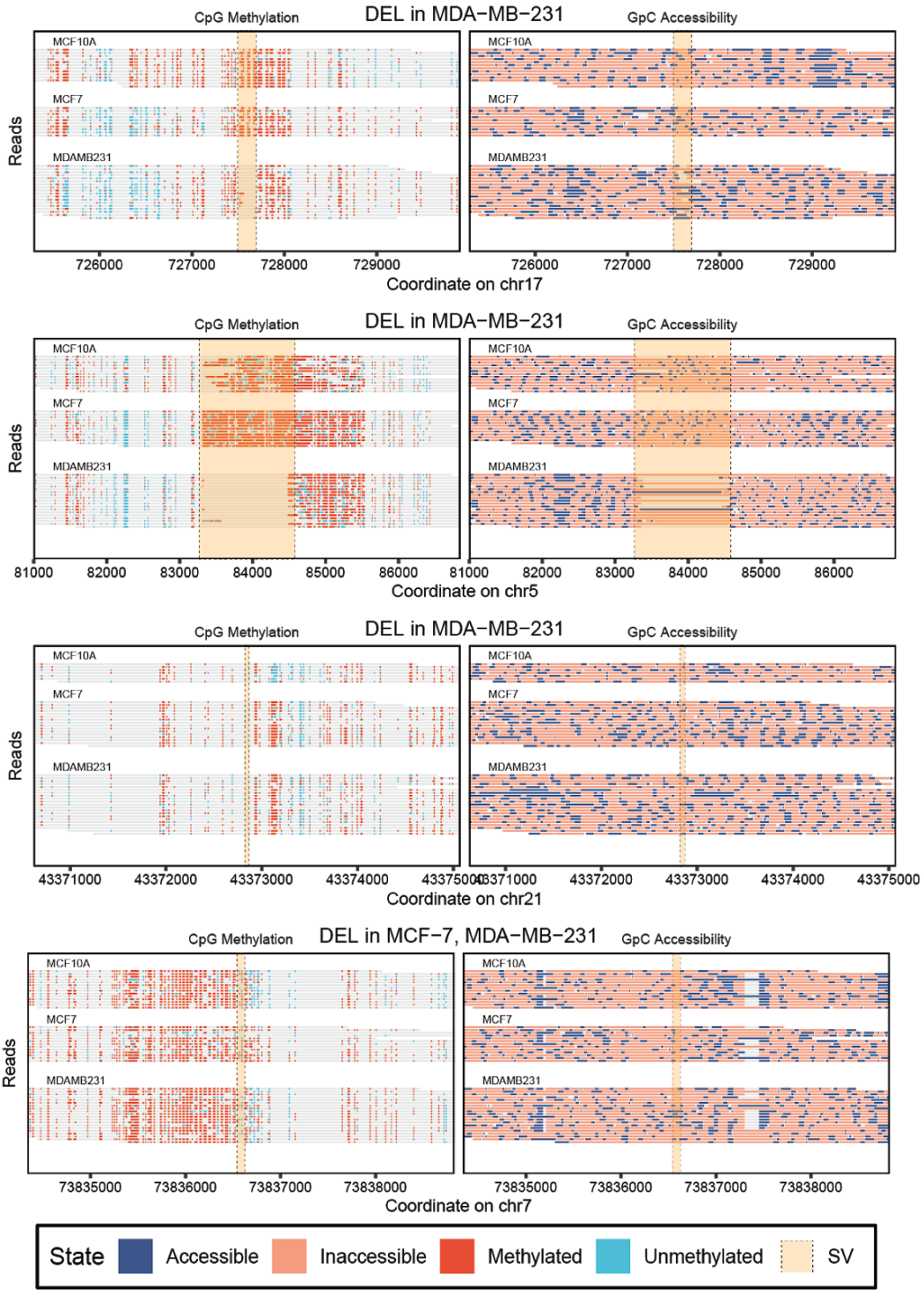
Supplementary Figure 15. Single read level plot of methylation and accessibility proximal to *ER* TSS. DNA methylation (top) and accessibility (bottom) single read plots for the 3 breast cancer cell lines (MCF-10A, MCF-7, MDA-MB-231)



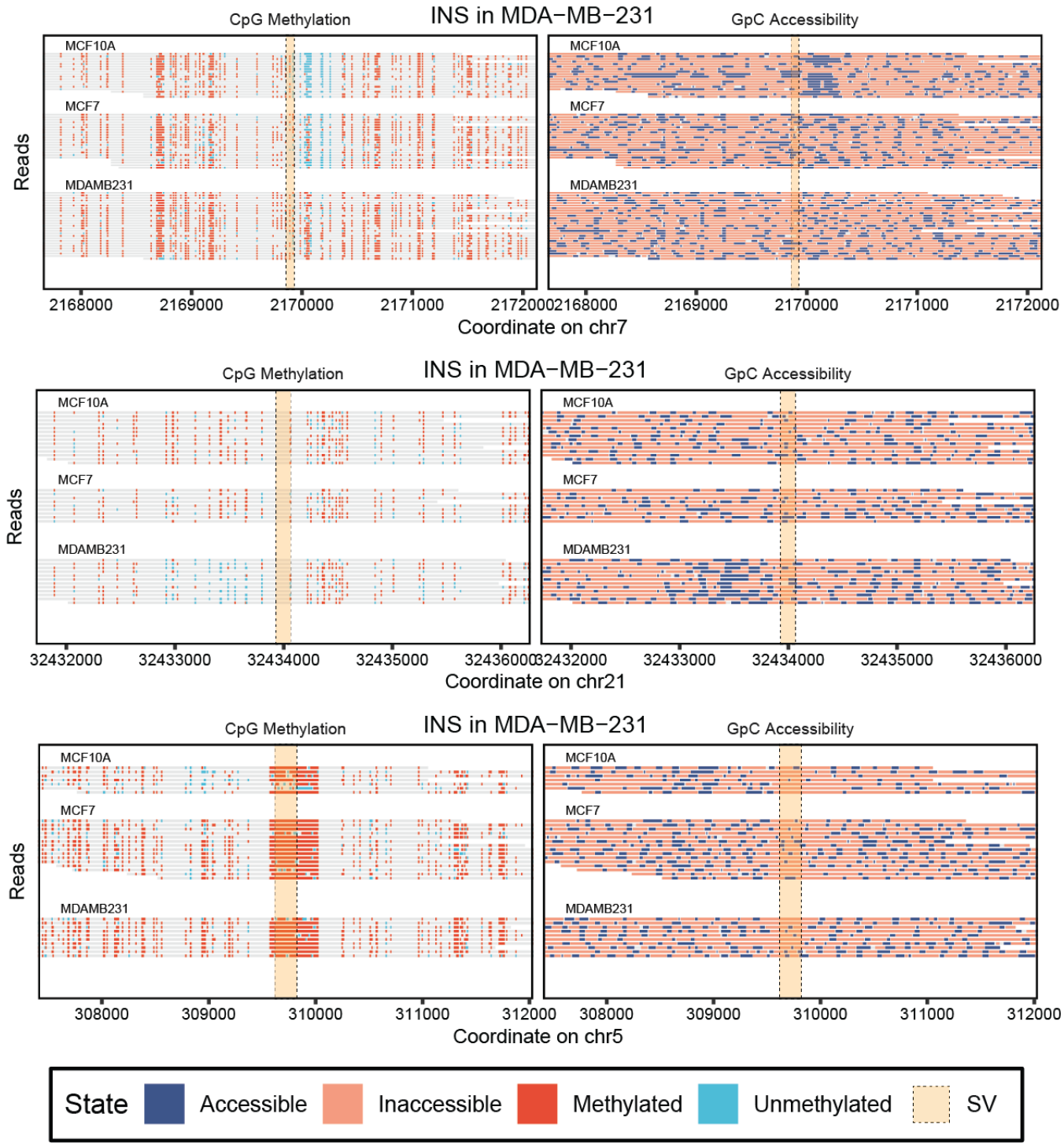
Supplementary Figure 16. Single read level plot of methylation and accessibility proximal to *PR* TSS. DNA methylation (top) and accessibility (bottom) single read plots for the 3 breast cancer cell lines (MCF-10A, MCF-7, MDA-MB-231)



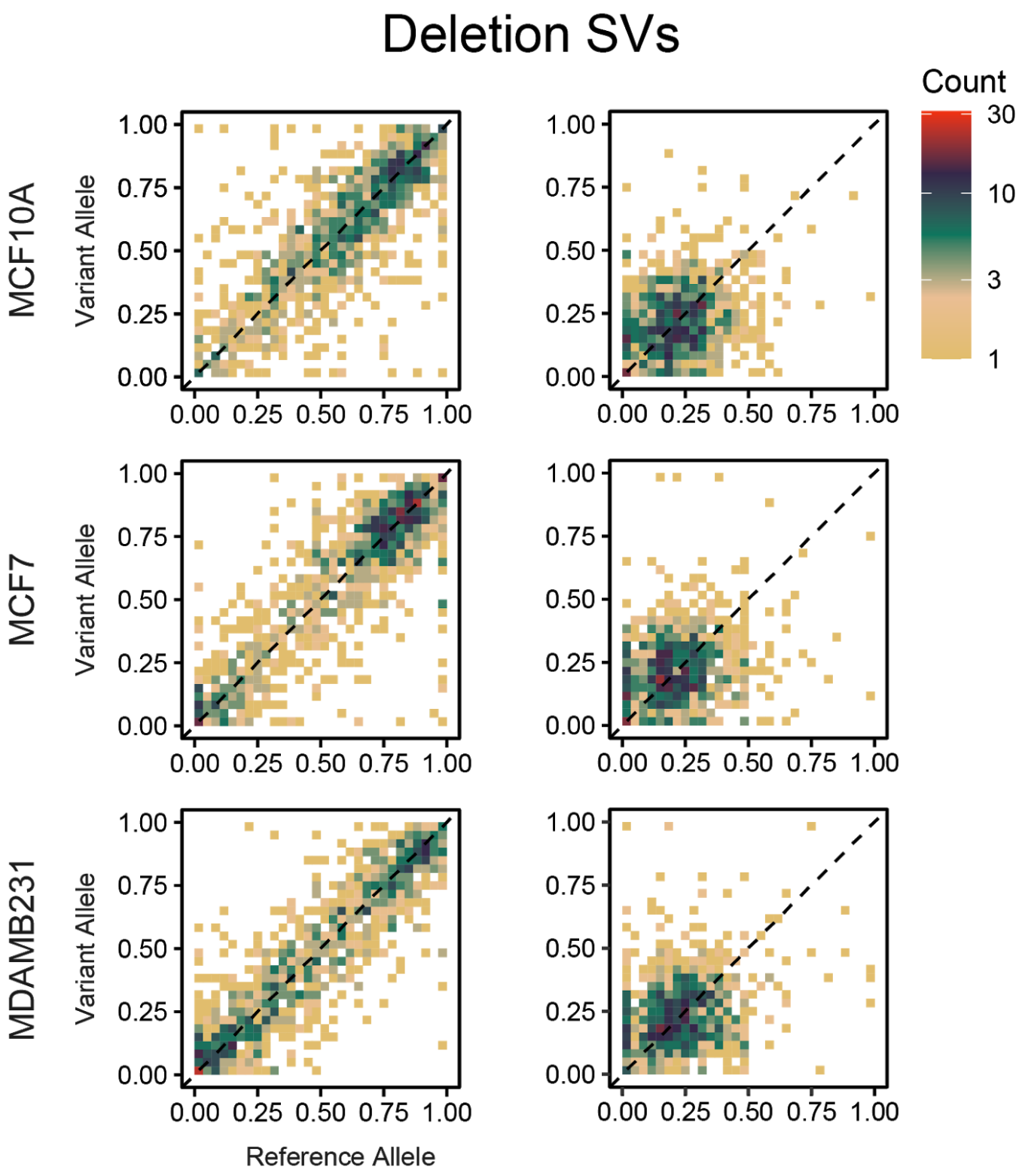
Supplementary Figure 17. Single read level plot of methylation and accessibility proximal to *HER2* TSS. DNA methylation (top) and accessibility (bottom) single read plots for the 3 breast cancer cell lines (MCF-10A, MCF-7, MDA-MB-231)



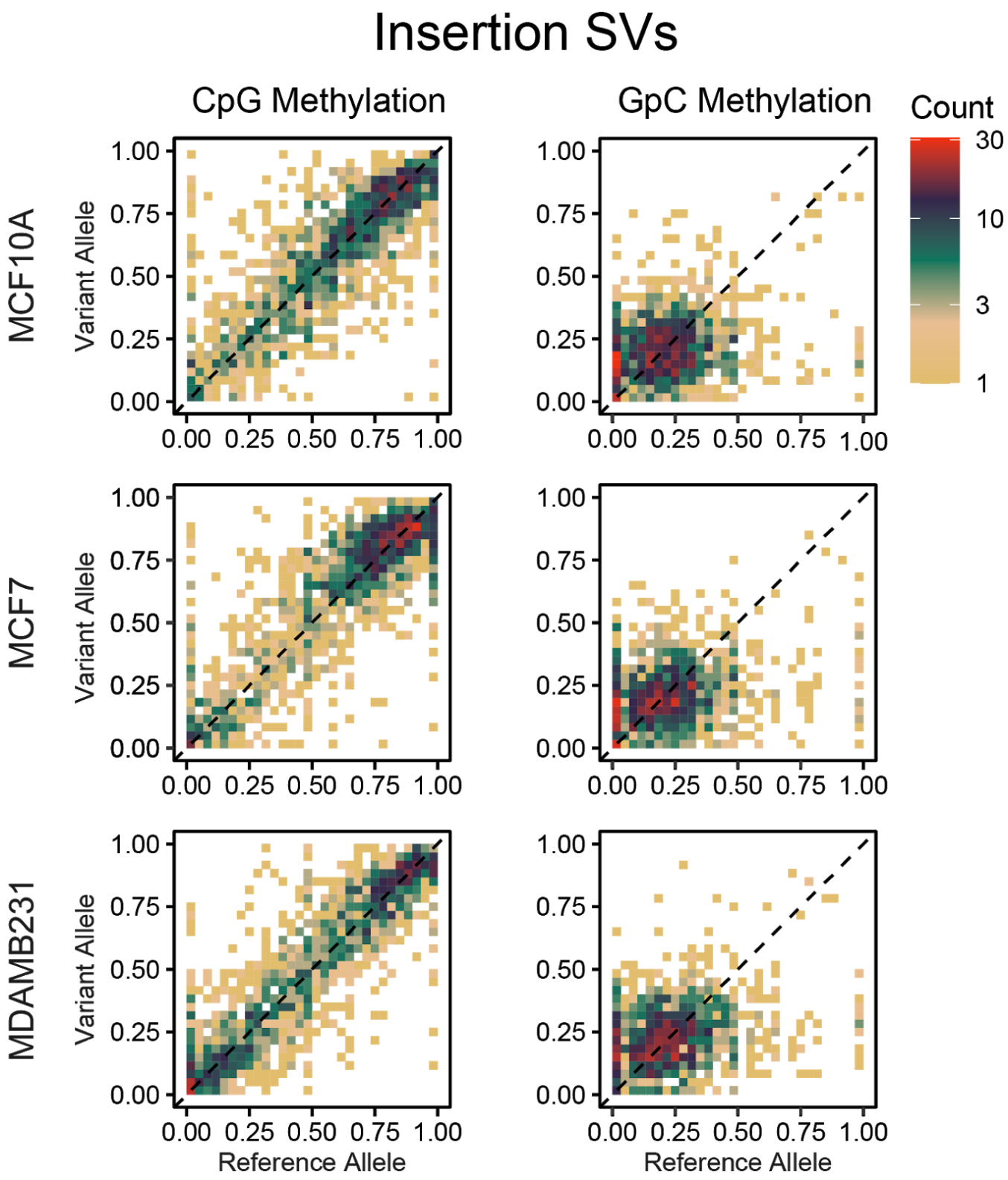
Supplementary Figure 18. Structural variations and differential epigenetics: Deletions. Single-read methylation and accessibility plots of regions that had a deletion SV in the cancer subtypes and not in MCF-10A, as well as differential methylation and accessibility in the cancer subtypes in comparison to MCF-10A.



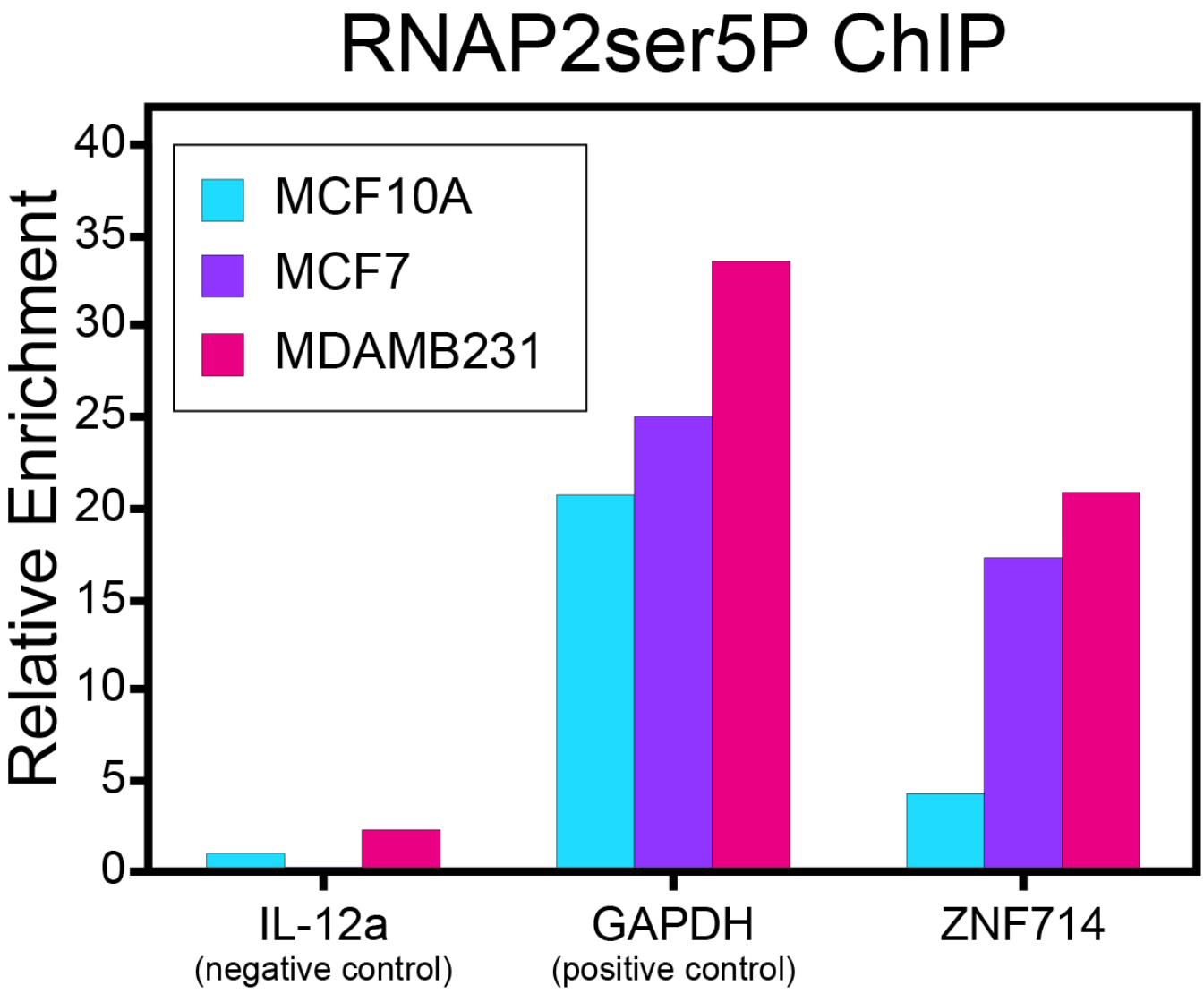
Supplementary Figure 19. Structural variations and differential epigenetics: Insertion. Single-read methylation and accessibility plots of regions that had an insertion SV in the cancer subtypes and not in MCF-10A, as well as differential methylation and accessibility in the cancer subtypes in comparison to MCF-10A.



Supplementary Figure 20. Epigenetic comparison at heterozygous deletions. Using a window 1kb upstream and downstream of the breakpoint for CpG methylation and 100bp upstream/downstream of the breakpoint for GpC methylation (accessibility) bp we compared the epigenetic state at heterozygous SVs detected in the breast cancer cell lines between the reference allele and the variant allele.



Supplementary Figure 21. Epigenetic comparison at heterozygous insertions. Using a window 1kb upstream and downstream of the breakpoint for CpG methylation and 100bp upstream/downstream of the breakpoint for GpC methylation (accessibility) bp we compared the epigenetic signature at heterozygous SVs detected in the breast cancer cell lines between the reference allele and the variant allele.



Supplementary Figure 22. ChIP-qPCR at *ZNF714* Transcription Start Site. We performed ChIP against RNAPolII Ser5P (the state associated with promoter-proximal pausing) to interrogate whether the subnucleosomal footprint represented a transcription complex binding. We found dramatic enrichment in binding to the *ZNF714* TSS in MCF7 and MDAMB231 as compared to the MCF10A cell line. For comparison we also examined the TSS of *IL-12a* (negative control) and *GAPDH* (positive control). Results are normalized to 5% input and plotted relative to the negative control (*IL-12a*) in MCF10A.

Origin	Methylation	Detected methylation	Methylated loci	Unmethylated loci	Methylation %
Ecoli	CpG	CpG	278724	5446	**98**
Ecoli	CpG	GpC	10629	485977	**2**
Ecoli	CpGGpC	CpG	235883	4051	**98**
Ecoli	CpGGpC	GpC	430714	7089	**98**
Ecoli	GpC	CpG	12080	172313	**7**
Ecoli	GpC	GpC	359733	5412	**99**
Ecoli	Unmethylated	CpG	4132	158333	**3**
Ecoli	Unmethylated	GpC	3529	292160	**1**
GM12878	CpG	CpG	210033	8703	**96**
GM12878	CpG	GpC	19158	1866755	**1**
GM12878	CpGGpC	CpG	179562	8364	**96**
GM12878	CpGGpC	GpC	1628962	42718	**97**
GM12878	GpC	CpG	10838	243231	**4**
GM12878	GpC	GpC	2374964	48817	**98**
GM12878	Unmethylated	CpG	2478	195918	**1**
GM12878	Unmethylated	GpC	13620	1739225	**1**

Supplementary Table 1. CpG and GpC methylation rates from bisulfite sequencing of samples treated with combinations of methyltransferases. Total number of methylated motifs (CG or GC depending on treatment) and unmethylated loci for each sample were tabulated to calculate the percent methylation per sample.

Sample	Methylation	Total Reads	Reads with MAPQ >= 20	Fraction	Yield (Gb)	Coverage
E. coli	Unmethylated	724,887	707,199	0.98	4.14	891.24
E. coli	CpG	1,396,674	1,355,969	0.97	7.39	1592.70
E. coli	GpC	1,251,301	1,209,030	0.97	7.15	1540.85
E. coli	CpGGpC	1,084,568	1,048,904	0.97	6.41	1381.27
NA12878	Unmethylated	2,999,723	2,803,707	0.93	10.28	3.18
NA12878	CpG	2,951,563	2,748,099	0.93	8.47	2.62
NA12878	GpC	2,941,260	2,723,240	0.93	9.28	2.87
NA12878	CpGGpC	2,848,032	2,631,771	0.92	9.14	2.83

Supplementary Table 2. Nanopore sequencing results of training and testing sets. Total number of methylated loci and unmethylated loci for each sample was tabulated to calculate the percent methylation per sample.

CpG Methylation

Truth	Prediction: M	Prediction: U
M	44 %	6 %
U	5 %	45 %

GpC Methylation

Truth	Prediction: M	Prediction: U
M	47 %	3 %
U	2 %	48 %

M : Methylated
U : Unmethylated

Supplementary Table 3. Confusion matrix of CpG and GpC methylation calling. Subsets of 100,000 sites in unmethylated and methylated samples were methylation called using a log-likelihood ratio threshold of 1.5 for CpG and 1 for GpC. The number of resulting categories of calls is summarized in the table.

		Genomic Contexts				
Motif	Metric	CGI	Genes	LINE	Promoters	SINE
CG	Accuracy	0.99	1.00	1.00	1.00	1.00
CG	Call Rate	1.09	1.01	1.01	1.04	1.01
GC	Accuracy	0.99	1.00	1.00	1.00	1.00
GC	Call Rate	0.97	1.01	1.01	1.00	1.01

Supplementary Table 4. Relative accuracy and call rates for notable genomic contexts. Accuracy and call rate with respect to genomic contexts in comparison to overall accuracy and call rate (context accuracy/call rate divided by overall accuracy/call rate). A relative value of 1 represents the same value in the context in comparison to overall value; No particular decrease in accuracy or call rate was observed based on genomic context.

Cell	Experiment number	Flowcell type	Number of flowcells	Number of raw reads (M)	Total raw bases (Gb)	Aligned reads (M)	Aligned bases (Gb)	N50 length
GM12878	1	FLO-MIN106	2	1.64	12.95	1.30	11.20	11,475
GM12878	2	FLO-MIN106	2	1.90	14.86	1.56	12.68	10,736
GM12878	3	FLO-MIN106	2	0.81	6.53	0.68	5.66	10,526
GM12878	4	FLO-MIN106	2	1.03	11.10	0.89	9.79	17,346
GM12878	5	FLO-MIN106	2	1.95	14.77	1.55	12.11	11,635
GM12878	6	FLO-MIN106	2	2.09	15.42	1.69	13.55	11,156
GM12878	7	FLO-PROM002	1	11.51	72.84	8.90	62.62	9,791
GM12878	8	FLO-PROM002	1	5.40	85.51	4.80	74.94	20,850
GM12878	9	FLO-PROM002	1	5.66	64.28	5.03	54.37	20,626
MCF10A	1	FLO-MIN106	2	0.86	6.52	0.60	5.70	11,428
MCF10A	2	FLO-MIN106	3	3.59	33.11	3.05	29.64	11,888
MCF10A	3	FLO-MIN106	4	4.96	41.96	4.09	37.06	11,215
MCF7	1	FLO-MIN106	1	1.17	9.71	1.03	8.77	11,210
MCF7	2	FLO-MIN106	2	0.78	5.29	0.52	4.72	11,652
MCF7	3	FLO-MIN106	5	5.42	44.25	4.48	39.61	12,462
MCF7	4	FLO-MIN106	3	1.62	17.51	1.46	16.05	18,532
MDAMB231	1	FLO-MIN106	1	0.89	8.22	0.78	7.39	11,531
MDAMB231	2	FLO-MIN106	2	2.10	19.01	1.82	17.30	11,488
MDAMB231	3	FLO-MIN106	3	3.09	33.06	2.68	30.14	14,583
MDAMB231	4	FLO-MIN106	3	1.87	22.10	1.67	20.06	15,783

Supplementary Table 5. Individual nanopore sequencing run metrics. Nanopore sequencing run statistics for individual sequencing experiments performed, before pooling them by cell line to achieve the final yields

Cluster	Expression Quartile	Number of Genes	Total Reads	Fraction Q1	Fraction Mean	Fraction Median	Fraction Q3
High CpG ; High GpC	1	8465	17348	0	0.041	0.022	0.048
High CpG ; High GpC	2	4577	8861	0	0.036	0.02	0.048
High CpG ; High GpC	3	3014	6444	0	0.037	0.026	0.05
High CpG ; High GpC	4	3316	7378	0.013	0.037	0.027	0.053
High CpG ; Low GpC	1	8465	55038	0	0.106	0.019	0.047
High CpG ; Low GpC	2	4577	100415	0.02	0.355	0.091	0.862
High CpG ; Low GpC	3	3014	103723	0.077	0.553	0.746	0.938
High CpG ; Low GpC	4	3316	128635	0.135	0.607	0.843	0.943
Low CpG ; High GpC	1	8465	232465	0.236	0.507	0.588	0.767
Low CpG ; High GpC	2	4577	43884	0	0.182	0.054	0.265
Low CpG ; High GpC	3	3014	11711	0	0.075	0	0.058
Low CpG ; High GpC	4	3316	9427	0	0.053	0	0.038
Low CpG ; Low GpC	1	8465	155850	0.137	0.346	0.262	0.524
Low CpG ; Low GpC	2	4577	105164	0.048	0.427	0.4	0.785
Low CpG ; Low GpC	3	3014	54033	0.014	0.335	0.15	0.727
Low CpG ; Low GpC	4	3316	58403	0	0.303	0.095	0.66

Supplementary Table 6. Summary of Single-read Combinatorial Clustering on Genes by Expression. Each read on gene promoters was clustered as one of the four combinatorial epigenetic signatures, and the fractions of reads in each cluster were calculated for each gene promoter. The distributions of the cluster fractions stratified by expression quartile were plotted in Supplementary Figure 14a. Shown in this table are summary statistics showing for each epigenetic signature and expression quartile the number of genes, total mapped reads, and - for the fraction of reads in each signature - the 25% percentile (Q1), mean, median, and 75% percentile (Q3).

Cluster	Histone Modification	Number of Genes	Total Reads	Fraction Q1	Fraction Mean	Fraction Median	Fraction Q3
High CpG ; High GpC	Bivalent	598	399	0	0.013	0	0.02
High CpG ; High GpC	H3K27me3	1014	947	0	0.018	0.013	0.024
High CpG ; High GpC	H3K4me3	11048	21244	0	0.04	0.02	0.046
High CpG ; High GpC	None	6712	17441	0.016	0.042	0.033	0.058
High CpG ; Low GpC	Bivalent	598	2497	0.015	0.084	0.033	0.075
High CpG ; Low GpC	H3K27me3	1014	12513	0.043	0.236	0.121	0.345
High CpG ; Low GpC	H3K4me3	11048	47123	0	0.08	0.019	0.047
High CpG ; Low GpC	None	6712	325678	0.667	0.749	0.912	0.955
Low CpG ; High GpC	Bivalent	598	4215	0.039	0.128	0.076	0.158
Low CpG ; High GpC	H3K27me3	1014	2540	0.015	0.045	0.034	0.062
Low CpG ; High GpC	H3K4me3	11048	283683	0.182	0.477	0.525	0.746
Low CpG ; High GpC	None	6712	7049	0	0.021	0	0.015
Low CpG ; Low GpC	Bivalent	598	26114	0.727	0.776	0.845	0.898
Low CpG ; Low GpC	H3K27me3	1014	39665	0.587	0.701	0.806	0.893
Low CpG ; Low GpC	H3K4me3	11048	235844	0.167	0.404	0.326	0.64
Low CpG ; Low GpC	None	6712	71827	0	0.188	0.023	0.238

Supplementary Table 7. Summary of Single-read Combinatorial Clustering on Genes by Histone Modification. Each read on gene promoters was clustered as one of the four combinatorial epigenetic signatures, and the fractions of reads in each cluster were calculated for each gene promoter. The distributions of the cluster fractions stratified by histone modification were plotted in Supplementary Figure 14b and the summary values are shown in this table

Type	Context	Enrichment in Xa (Maternal)	Enrichment in Xi (Paternal)	Number in Xa (Maternal)	Number in Xi (Paternal)
Concordant	CTCF	2.125	0	2	0
Concordant	Gene_Body_Escape	0.712	0.237	3	1
Concordant	Gene_Body_XCI	1.324	0.213	31	5
Concordant	Gene_Body_unknown	0.536	0.195	66	24
Concordant	TF	1.679	0	2	0
Concordant	TSS_Escape	28.344	0	3	0
Concordant	TSS_XCI	102.148	0	66	0
Concordant	TSS_unknown	18.13	0.283	128	2
Concordant	Enhancer	1.828	0	5	0
DMR	CTCF	1.817	1.454	25	20
DMR	Gene_Body_Escape	0.94	0.194	58	12
DMR	Gene_Body_XCI	4.145	0.3	1421	103
DMR	Gene_Body_unknown	1.19	0.153	2146	275
DMR	TF	1.952	0.115	34	2
DMR	TSS_Escape	0.646	1.939	1	3
DMR	TSS_XCI	0.529	7.516	5	71
DMR	TSS_unknown	1.239	2.013	128	208
DMR	Enhancer	5.023	0.05	201	2
DAR	CTCF	10.987	2.652	29	7
DAR	Gene_Body_Escape	1.441	0.32	18	4
DAR	Gene_Body_XCI	3.008	0.233	219	17
DAR	Gene_Body_unknown	1.176	0.227	513	99
DAR	TF	2.976	0.893	10	3
DAR	TSS_Escape	44.564	5.57	8	1
DAR	TSS_XCI	91.21	0	102	0
DAR	TSS_unknown	10.993	0.465	189	8
DAR	Enhancer	1.626	0.325	15	3

Supplementary Table 8. Summary of allele-specific differential epigenetic regions by genomic context in X chromosome. The numbers and enrichments of differentially methylated regions (DMRs), differentially accessible regions (DARs), and concordantly differential regions in X chromosome, divided by the types of genomic contexts. Enrichments and counts are graphically shown in Figure 3b and Supplementary Figure 20a, respectively.

Type	Context	Enrichment in Maternal	Enrichment in Paternal	Number in Maternal	Number in Paternal
Concordant	CTCF	1.331	5.325	1	4
Concordant	Gene_Body_Autosomal	0.474	0.523	77	85
Concordant	Gene_Body_imprint	4.576	6.407	5	7
Concordant	TF	2.791	0.698	4	1
Concordant	TSS_Autosomal	3.903	3.365	29	25
Concordant	TSS_imprint	199.028	199.028	4	4
Concordant	Enhancer	3.689	1.845	14	7
DMR	CTCF	1.273	1.697	18	24
DMR	Gene_Body_Autosomal	0.603	0.614	1847	1880
DMR	Gene_Body_imprint	1.748	5.777	36	119
DMR	TF	1.295	1.591	35	43
DMR	TSS_Autosomal	1.716	2.017	240	282
DMR	TSS_imprint	26.378	26.378	10	10
DMR	Enhancer	1.036	1.204	74	86
DAR	CTCF	1.394	1.208	30	26
DAR	Gene_Body_Autosomal	0.554	0.567	2950	3017
DAR	Gene_Body_imprint	0.712	0.712	22	22
DAR	TF	1.209	1.26	48	50
DAR	TSS_Autosomal	1.505	1.644	261	285
DAR	TSS_imprint	10.728	10.728	4	4
DAR	Enhancer	0.893	0.909	112	114

Supplementary Table 9. Summary of allele-specific differential epigenetic regions by genomic context in autosomes. The numbers and enrichments of differentially methylated regions (DMRs), differentially accessible regions (DARs), and concordantly differential regions in autosomes, divided by the types of genomic contexts. Enrichments and counts are graphically shown in Supplementary Figure 20b,c.

Sample	DEL	INS	DUP	INV	TRA	Total
MCF10A	2245	1381	188	37	57	3908
MCF7	2297	1139	225	63	110	3834
MDAMB231	2605	1754	305	44	37	4745
MCF7 + MDAMB231	1020	669	92	9	15	1805
MCF10A + MDAMB231	1562	1124	156	14	11	2867
MCF10A + MCF7	1032	655	69	19	21	1796
Total	10761	6722	1035	186	251	18955

Inclusive counts

MCF10A	4839	3160	413	70	89	8571
MCF7	4349	2463	386	91	146	7435
MDAMB231	5187	3547	553	67	63	9417

Supplementary Table 10. Summary of structural variations detected in breast cell lines. Structural variations types are deletions (DEL), translocations (TRA), duplications (DUP), inversions (INV), and insertions (INS), and are grouped by uniquely occurring (first three lines), commonly occurring in any combination of two cell lines (second three lines), and a summary of inclusive counts for each cell line. SVs of < 50bp were filtered out.

Supplementary Data 1. nanoNOMe accessibility peaks in GM12878
Peaks of accessibility were detected from nanoNOMe GpC methylation frequencies across the genome. The significance of each accessible region was determined by performing a binomial test of the raw frequency of accessibility, with overall accessibility frequency as the null probability. The probabilities were corrected for multiple testing using Benjamini-Hochberg correction, and accessible regions with adjusted p-values less than 0.01 and widths greater than 50 bps were determined to be accessibility peaks.

Chromosome: Chromosome of peak region
Start: 1-based start coordinate of peak region
End: 1-based end coordinate of peak region
idxStart: Start GpC site index of peak region
idxEnd: End GpC site index of peak region
Num_Sites: Number of sites
Observations: Total number of GpC observations in the region
Methylated: Total number of methylated GpCs in the region
Average_Accessibility: Average accessibility of the region
Maximum_Accessibility: Peak accessibility of the region
p.value: P-value of the binomial test of the accessibility using the median accessibility across the genome as the hypothetical probability
adjusted.pval: FDR adjusted p-value
Region_Width: Width of the region

Supplementary Data 2. CTCF binding sites in GM12878
BED file of CTCF binding sites determined by overlapping computationally predicted CTCF binding sites (from CTCFBSDB 2.0) with conservative IDR peaks in ChIP-seq of CTCF on GM12878 (ENCODE accession ENCSR000AKB) and removing peaks that fell within 2kb of known TSS. Coordinates are in bed format :

Column 1: Chromosome
Column 2: 0-based start coordinate
Column 3: 1-based end coordinate
Column 4: Presence of ChIP-seq peak in the binding site - Bound: ChIP-seq peak present in GM12878, Unbound: ChIP-seq peak not present in GM12878

Supplementary Data 3. Estimated protein-bound regions near a subset of gene TSS in GM12878
Estimated protein-binding regions were detected in regions 10kb upstream and downstream of the TSS of a randomly selected subset of 1,000 genes from each expression quartile. Sites that have 10 or more reads with short inaccessible runs, suggesting protein binding, in a window smaller than 80 bp were selected as estimated protein-binding regions.

Chromosome: Chromosome of estimated protein binding region
Start: 1-based start coordinate
End: 1-based end coordinate
Width: Width of the region

Bound: Number of reads suggesting protein binding
Unbond: Number of reads not suggesting protein binding
Bound_Fraction: Fraction of reads suggesting protein binding

Supplementary Data 4. Protein binding stratified by promoter epigenetic signatures
Reads were grouped by the promoter epigenetic signature, then the fraction of reads that are bound in a nearby protein binding region(s) was calculated for each group and each protein binding region.

Chromosome: Chromosome of estimated protein binding region
Start: 1-based start coordinate
End: 1-based end coordinate
TxID: Transcript ID of the nearby TSS
Transcription_Start_Site: TSS of the gene
Symbol: Gene Symbol
Gene_Strand: Strand of the gene
Bound_High_CpG_Low_GpC : Number of bound reads in High CpG/Low GpC (concordantly inactive) group
Bound_High_CpG_High_GpC: Number of bound reads in High CpG/High GpC group
Bound_Low_CpG_Low_GpC: Number of bound reads in Low CpG/Low GpC group
Bound_Low_CpG_High_GpC: Number of bound reads in Low CpG/High GpC (concordantly active) group
Unbound_High_CpG_Low_GpC: Number of unbound reads in High CpG/Low GpC (concordantly inactive) group
Unbound_High_CpG_High_GpC: Number of unbound reads in High CpG/High GpC group
Unbound_Low_CpG_Low_GpC: Number of unbound reads in Low CpG/Low GpC group
Unbound_Low_CpG_High_GpC: Number of unbound reads in Low CpG/High GpC (concordantly active) group
Bound_Fraction_High_CpG_Low_GpC: Fraction of bound reads in High CpG/Low GpC (concordantly inactive) group
Bound_Fraction_High_CpG_High_GpC: Fraction of bound reads in High CpG/High GpC group
Bound_Fraction_Low_CpG_Low_GpC: Fraction of bound reads in Low CpG/Low GpC group
Bound_Fraction_Low_CpG_High_GpC: Fraction of bound reads in Low CpG/High GpC (concordantly active) group

Supplementary Data 5. Allele-specific DMRs and DARs in GM12878
Methylation and accessibility were compared between the paternal and maternal alleles to obtain differentially methylated regions and differentially accessible regions. To find differentially methylated regions (DMRs) between two samples without replicates, the difference of methylation between the two samples was calculated for each CpG site. Then, continuous regions with differences greater than 99th percentile of the differences were selected as candidates for hypermethylation, and regions with differences less than the 1st percentile were selected as candidates for hypomethylation. Similarly, for DARs, we performed a one-sided Fisher's Exact test on raw counts of methylated and unmethylated calls on each candidate DAR. P-values were corrected using Benjamini-Hochberg correction, and regions with adjusted p-values less than 0.01 and widths greater than 100 bps were determined to be significant DMRs.

Chromosome: Chromosome of DMR/DAR
Start: 1-based start coordinate of region
End: 1-based end coordinate of region
idxStart: Start CpG/GpC site index of region
idxEnd: End CpG/GpC site index of region
Num_Sites: Number of sites
Paternal_Observations: Total number of observations in the region in the paternal allele
Maternal_Observations: Total number of observations in the region in the maternal allele
Paternal_Methylated: Total number of methylated CpG/GpCs in the region in the paternal allele
Maternal_Methylated: Total number of methylated CpG/GpCs in the region in the maternal allele
Paternal_Average: Average methylation/accessibility of the region in the paternal allele
Maternal_Average: Average methylation/accessibility of the region in the maternal allele
Hypermethylated/More_Accessible : Allele that has higher methylation/accessibility
Mean_Difference: Difference (Maternal - Paternal) of average methylation/accessibility in the region
Max_Difference: Maximum per-site pairwise difference (Maternal - Paternal) of methylation/accessibility in the region
p.value: P-value of one-sided Fisher's exact test of the difference between the alleles
adjusted.pval : FDR adjusted p-value
Width: Width of the region

Supplementary Data 6. Gene promoter regions with allele-specific DMRs and DARs in GM12878
Genes with an allele-specific DMR and/or DAR within 500 bp of the TSS were identified. In the cases where both a DMR and a DAR were near the TSS, concordant activity was determined for the allele that was hypomethylated and more accessible

Ensembl_ID: Ensembl gene ID
Chromosome: Chromosome of the gene
Transcription_Start_Site: TSS of the gene
Strand: strand of the gene
Symbol: Gene symbol
Higher_Methylation: Allele that had a higher methylation in a DMR nearby
Higher_Accessibility: Allele that had a higher accessibility in a DAR nearby
Concordant_Activity: Allele that was more concordantly active

Supplementary Data 7. Heterozygous structural variations in GM12878
After splitting GM12878 nanoNOMe reads into the maternal and paternal alleles, SVs were detected using sniffles on the two alleles separately. The SVs were processed using SURVIVOR pipeline, resulting in a merged vcf file of SVs from the two alleles. The data follows a standard VCF format from sniffles/SURVIVOR, with last two columns indicating the allele :

Paternal_Allele: SV information in the paternal allele
Maternal_Allele: SV information in the maternal allele

Supplementary Data 8. DMRs and DARs in MCF-7 and MDA-MB-231 in comparison to MCF-10A
Methylation and accessibility were compared between the cancer subtypes of breast cell lines - MCF-7 and MDA-MB-231 - and the normal MCF-10A cell line to obtain differentially methylated regions and differentially accessible regions. To find differentially methylated regions (DMRs) between two samples without replicates, the difference of methylation between the two samples was calculated for each CpG site. Then, continuous regions with differences greater than 99th percentile of the differences were selected as candidates for hypermethylation, and regions with differences less than the 1st percentile were selected as candidates for hypomethylation. Similarly, for DARs, we performed a one-sided Fisher's Exact test on raw counts of methylated and unmethylated calls on each candidate DAR. P-values were corrected using Benjamini-Hochberg correction, and regions with adjusted p-values less than 0.01 and widths greater than 100 bps were determined to be significant DMRs.

Chromosome: Chromosome of DMR/DAR
Start: 1-based start coordinate of region
End: 1-based end coordinate of region
idxStart: Start CpG/GpC site index of region
idxEnd: End CpG/GpC site index of region
Num_Sites: Number of sites
One: Sample One
Two: Sample Two
Comparison: Comparison that was performed
One_Observations: Total number of observations in the region in Sample 1
Two_Observations: Total number of observations in the region in Sample 2
One_Methylated: Total number of methylated CpG/GpCs in the region in Sample 1
Two_Methylated: Total number of methylated CpG/GpCs in the region in Sample 2
One_Average: Average methylation/accessibility of the region in the Sample 1
Two_Average: Average methylation/accessibility of the region in Sample 2
Higher: Sample that has higher methylation/accessibility
Mean_Difference: Difference of average methylation/accessibility in the region
Max_Difference: Maximum per-site pairwise difference of methylation/accessibility in the region
p.value: P-value of one-sided Fisher's exact test of the difference between the alleles
adjusted.pval: FDR adjusted p-value
Width: Width of the region

Supplementary Data 9. Summary of DMRs and DARs with respect to genomic contexts and structural variations.
DMRs, DARs, and concordantly differential regions were counted for each genomic context and SV, and the enrichments were determined with respect to expected number of the differential regions given the size of the given context.

What : Type of differential region
One : Sample 1
Two : Sample 2
Direction : Direction of differential signal in Sample 1 compared to Sample 2
Context : Genomic context
Number : Number of differential regions
Number_per_Mb : Number of differential regions per megabase of the genomic context
Enrichment : Enrichment of the differential regions compared to the genome-wide abundance

Supplementary Data 10. Structural variations in MCF-10A, MCF-7, and MDA-MB-231
SVs were detected using sniffles on each of the breast cell line nanoNOMe samples. The SVs were processed using SURVIVOR pipeline, resulting in a merged VCF file of SVs from all three samples. The data follows a standard VCF format from sniffles/SURVIVOR, with last three columns indicating the sample:

MCF10A_nanoNOMe: SV information in MCF-10A
MCF7_nanoNOMe: SV information in MCF-7
MDAMB231_nanoNOMe: SV information in MDA-MB-231

Supplementary Data 11. Promoter epigenetic signatures of differentially expressed genes in MCF-10A, MCF-7, and MDA-MB-231

Promoter epigenetic signatures of individual reads on genes that are differentially expressed in the same direction (up or downregulated) in both MCF-7 and MDA-MB-231 in comparison to MCF-10A

Chromosome: Chromosome of the gene
Ensembl_ID: Ensemble ID of the gene
Symbol: Gene Symbol
Gene_Strand: Strand of the gene
Transcription_Start_Site: TSS of the gene
Expression_in_MCF7_MDAMB231: Direction of differential expression in MCF-7 and MDA-MB-231 in comparison to MCF-10A
MCF10A_High_CpG_High_GpC: Number of High CpG/High GpC reads in MCF-10A
MCF10A_High_CpG_Low_GpC: Number of High CpG/Low GpC (concordantly inactive) reads in MCF-10A
MCF10A_Low_CpG_Low_GpC: Number of Low CpG/Low GpC reads in MCF-10A
MCF10A_Low_CpG_High_GpC: Number of Low CpG/High GpC (concordantly active) reads in MCF-10A
MCF7_High_CpG_High_GpC: Number of High CpG/High GpC reads in MCF-7
MCF7_High_CpG_Low_GpC: Number of High CpG/Low GpC (concordantly inactive) reads in MCF-7
MCF7_Low_CpG_Low_GpC: Number of Low CpG/Low GpC reads in MCF-7
MCF7_Low_CpG_High_GpC: Number of Low CpG/High GpC (concordantly active) reads in MCF-7
MDAMB231_High_CpG_High_GpC: Number of High CpG/High GpC reads in MDA-MB-231
MDAMB231_High_CpG_Low_GpC: Number of High CpG/Low GpC (concordantly inactive) reads in MDA-MB-231
MDAMB231_Low_CpG_Low_GpC: Number of Low CpG/Low GpC reads in MDA-MB-231
MDAMB231_Low_CpG_High_GpC: Number of Low CpG/High GpC (concordantly active) reads in MDA-MB-231

Supplementary Data 12. Protein binding regions near differentially expressed genes in MCF-10A, MCF-7, and MDA-MB-231

Protein binding regions were estimated by finding regions near differentially expressed genes with >10 overlapping reads that suggest protein binding. Then the number of protein-bound reads in each sample was tabulated. Each row represents a different putative protein binding region.

Chromosome: Chromosome of estimated protein binding region
Start: 1-based start coordinate
End: 1-based end coordinate
Ensembl_ID: Ensembl ID of the nearby TSS
Transcription_Start_Site: TSS of the gene
Symbol: Gene Symbol
Gene_Strand: Strand of the gene
Expression_in_MCF7_MDAMB231: Direction of differential expression in MCF-7 and MDA-MB-231 in comparison to MCF-10A
Bound_MCF10A: Number of bound reads in MCF-10A
Bound_MCF7: Number of bound reads in MCF-7
Bound_MDAMB231: Number of bound reads in MDA-MB-231
Unbound_MCF10A: Number of unbound reads in MCF-10A
Unbound_MCF7: Number of unbound reads in MCF-7
Unbound_MDAMB231: Number of unbound reads in MDA-MB-231
Bound_Fraction_MCF10A: Number of unbound reads in MCF-10A
Bound_Fraction_MCF7: Number of unbound reads in MCF-7
Bound_Fraction_MDAMB231: Number of unbound reads in MDA-MB-231