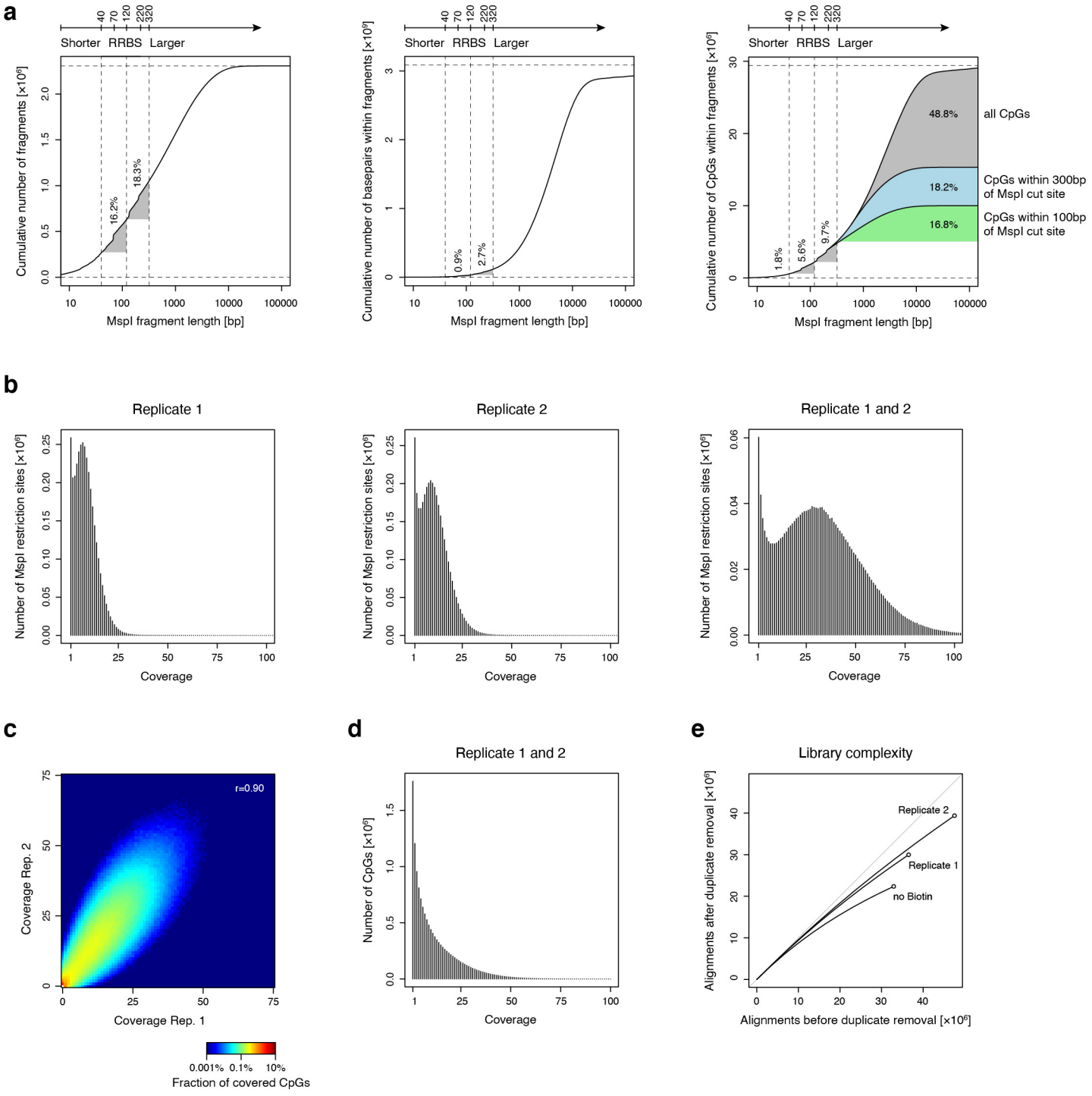
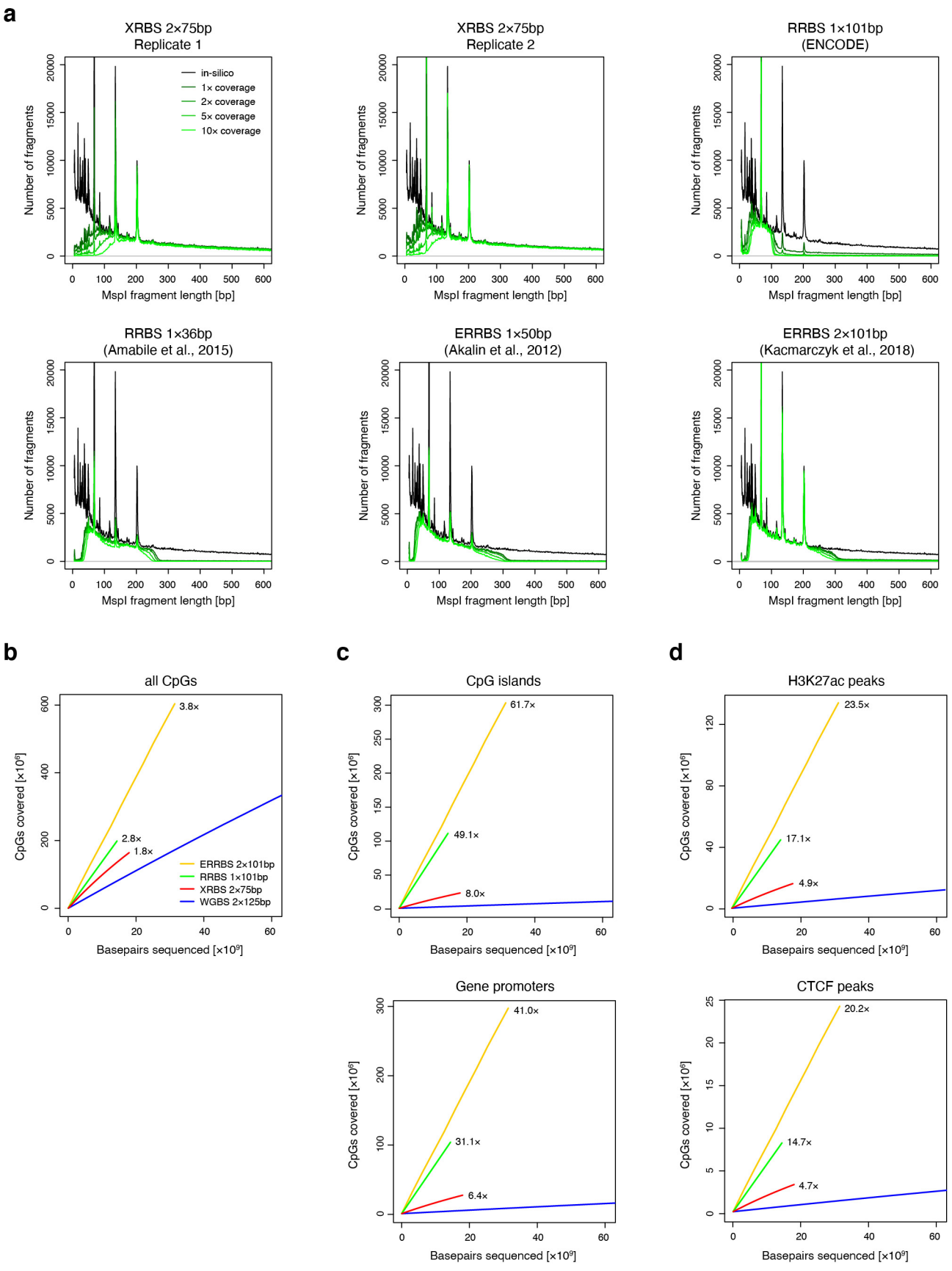
Supplementary Figure 1



Supplementary Fig. 1 | Evaluation of single MspI anchor design for methyl-CpG profiling

A) Plots show results from an in silico MspI restriction digest analysis of the human genome. The cumulative number of MspI fragments (total of 2.3 million, left), of basepairs (total of 3.1 billion, middle), and of CpGs (total of 29.4 million, right) is shown relative to increasing MspI fragment length. Vertical dotted lines show the size range of fragments captured in typical RRBS experiments. This analysis shows that RRBS of MspI fragments 40-120 bases in length covers only 0.9% of the genome, but enriches for 5.6% of genomic CpGs. Recent implementations of RRBS (e.g. enhanced RRBS; [14,15] that consider fragments up to 320 bases in length cover an additional 9.7% of CpGs. Approximately 35.0% of CpGs that are located within 300 bases of a single MspI site are not captured by these techniques.
B) Histogram shows coverage depth of MspI restriction sites for individual replicates of a 10ng XRBS library (left, middle), and both replicates combined (right).
C) Heatmap shows coverage depth of CpGs between replicates of a 10ng XRBS library (Pearson's r= 0.90).
D) Histogram shows coverage depth of CpGs in the combined dataset of both replicates (n=2 independently generated libraries).
E) Plot shows unique reads as a function of aligned reads in millions. With greater sequencing depth the fraction of unique reads decreases, as the chance of sampling a non-unique read (i.e. PCR duplicate) increases.

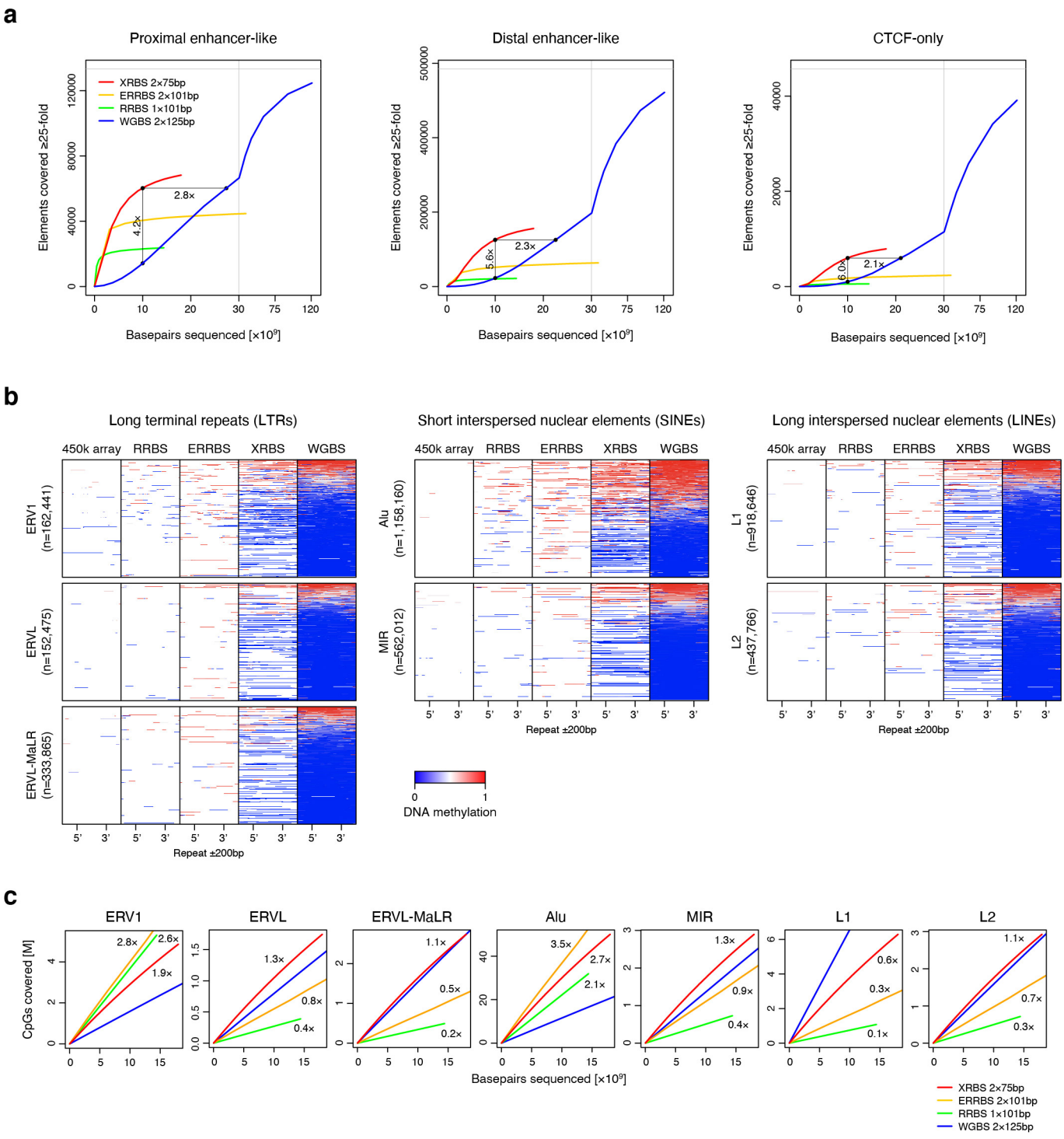
Supplementary Figure 2



Supplementary Fig. 2 | Comparison of MspI fragment detection and CpG coverage over regulatory elements

A) Plot shows number of detected fragments plotted as a function of calculated MspI fragment length from XRBS 10ng library replicates and from public RRBS and enhanced RRBS (ERRBS) datasets. Because of the random hexamer-primed second strand elongation step, XRBS efficiently detects fragments that exceed the selected fragment size range in RRBS (ENCODE; Amabile et al.: 40-220bp) and ERRBS (70-320bp). XRBS less efficiently captures short fragments (<70bp) compared to ERRBS and RRBS. Peaks in the graph correspond to three fragments commonly generated from Alu repetitive elements.
B) Plot compares CpG coverage as a function of sequencing depth (x-axis) for XRBS (red), WGBS (blue, ENCODE), ERRBS (orange; [52]) and RRBS (green, ENCODE).
C) Downsampling analysis plot as in panel B but restricted to CpGs within CpG islands (top) and gene promoters (bottom).
D) Downsampling analysis plot as in panel B but restricted to CpGs within H3K27ac peaks (top) and CTCF binding sites (bottom).

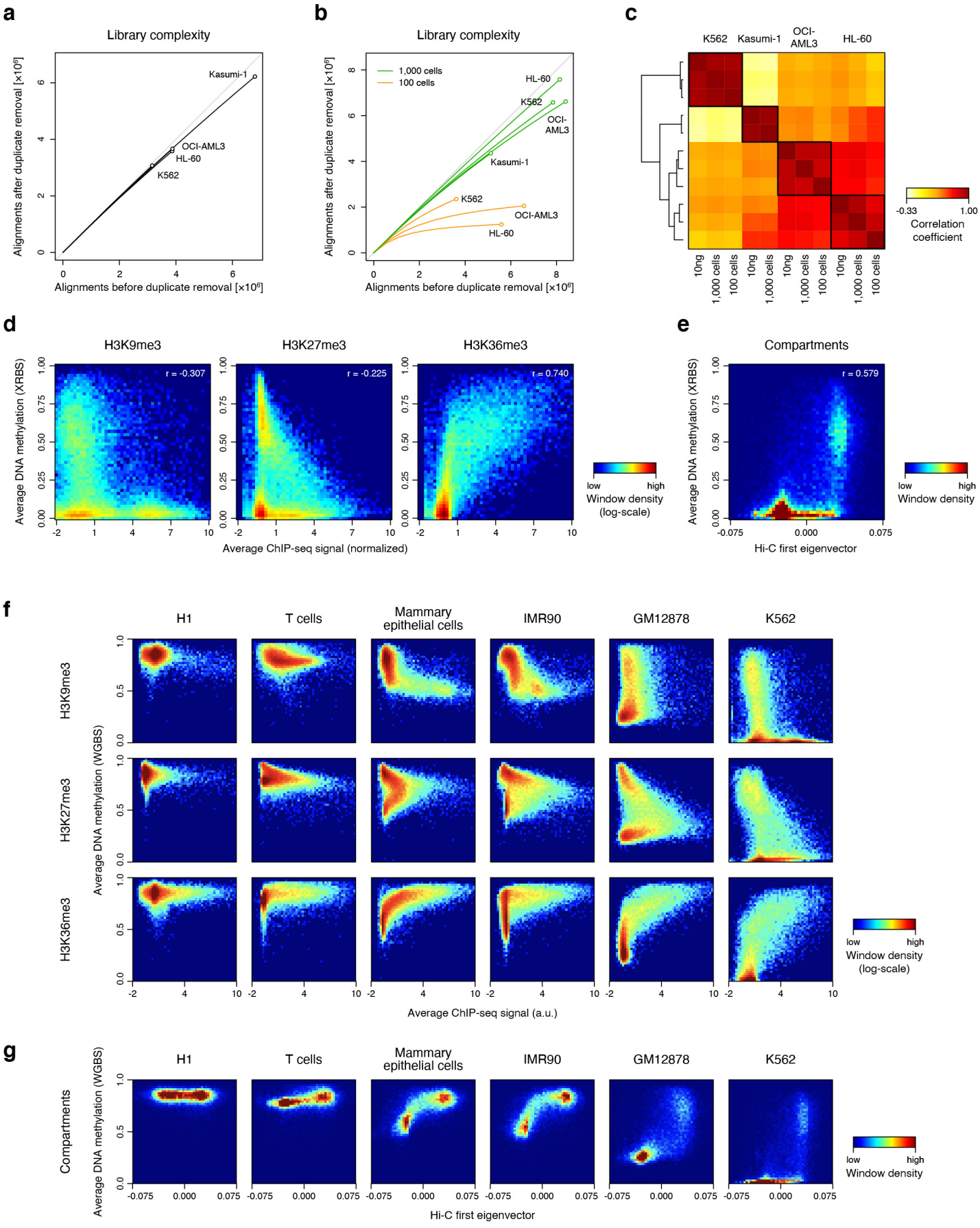
Supplementary Figure 3



Supplementary Fig. 3 | XRBS efficiently covers CpGs in regulatory elements and repetitive regions

A) Plots show the number of proximal enhancer-like, distal enhancer-like, and CTCF-only elements (as defined in the ENCODE SCREEN database) with at least 25-fold combined coverage as a function of sequencing depth for XRBS (red), WGBS (blue), ERRBS (orange), and RRBS (green). Enrichment for functional elements at a uniform sequencing depth of 10 billion base pairs is indicated. Vertical grey line indicates break in x-axis scale.
B) Heatmaps depict genomic regions (rows, n= 3,725,365 LTRs, SINEs, and LINEs) containing different repeat element families (as defined by RepeatMasker). Individual repeat elements are divided into 50 equally sized windows (5' and 3' position indicated). Upstream and downstream regions (±200bp) are divided into 25 equally sized windows. Panels from left to right show DNA methylation calls from 450k methylation array, RRBS, ERRBS, XRBS, and WGBS.
C) Plots compare CpG coverage within different repeat element families as a function of sequencing depth for XRBS (red), WGBS (blue), ERRBS (orange), and RRBS (green). CpG enrichment relative to WGBS is indicated. In comparison to RRBS, XRBS enriches for most repeat families, with the exception of Alu and ERV1 elements that frequently contain MspI restriction sites and are also efficiently captured (see also Supplementary Fig. 2a).

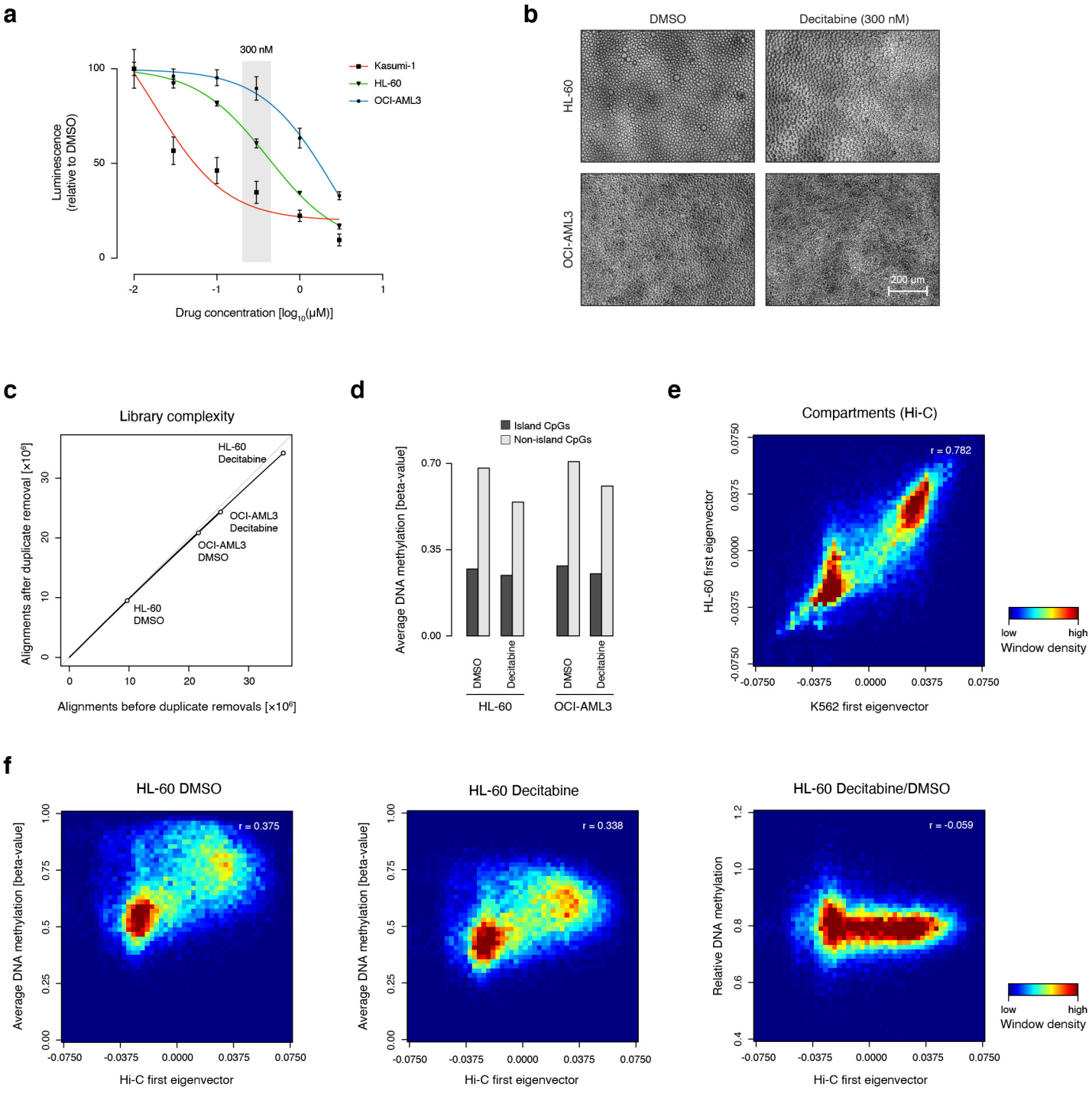
Supplementary Figure 4



Supplementary Fig. 4 | Correlation of DNA methylation with histone marks and compartment calls

A) Plot shows unique reads as a function of aligned reads in low-coverage XRBS libraries from K562, HL-60, OCI-AML3, and Kasumi-1 cells.
B) Plot shows unique reads as a function of aligned reads in low-coverage libraries from K562, Kasumi-1, HL-60, OCI-AML3 cells. Libraries were generated from 1,000 (green) and 100 (orange) cells sorted directly into lysis buffer. Libraries generated from 1,000 cells are comparable to libraries generated from 10ng of purified DNA (panel A), whereas 100 cell libraries show reduced complexity.
C) Heatmap shows Pearson correlation of XRBS methylation profiles of 100kb windows generated from 10ng gDNA, 1,000 or 100 sorted cells across four cell lines. Dendrogram derived from unsupervised clustering is indicated to the left. Sample grouping by DNA methylation is consistent with cell identity, indicating low technical variability between input material.
D) Heatmaps show correlation between average DNA methylation values and signal for H3K9me3 (left), H3K27me3 (center), and H3K36me3 (right) in 100kb-windows for K562 cells.
E) Heatmap shows correlation between DNA methylation and the Hi-C-derived first eigenvector indicating compartment A (positive values) and compartment B (negative values) in 100kb-windows for K562 cells.
F) Heatmap shows correlation between average DNA methylation values and ChIP seq signal for H3K9me3 (top), H3K27me3 (middle), and H3K38me3 (bottom) in 100kb-windows for human H1 embryonic stem cells, primary T cells and mammary epithelial cells, and cultured IMR90, GM12878 and K562 cells.
G) Heatmap as in panel F, but shows correlation between average DNA methylation values and the Hi-C-derived first eigenvector (x-axis). Positive values correspond to compartment A and negative values correspond to compartment B. While hypomethylation of compartment B is most pronounced in K562 cells, a similar trend is also observed in other cultured cell lines and in primary mammary epithelial cells.

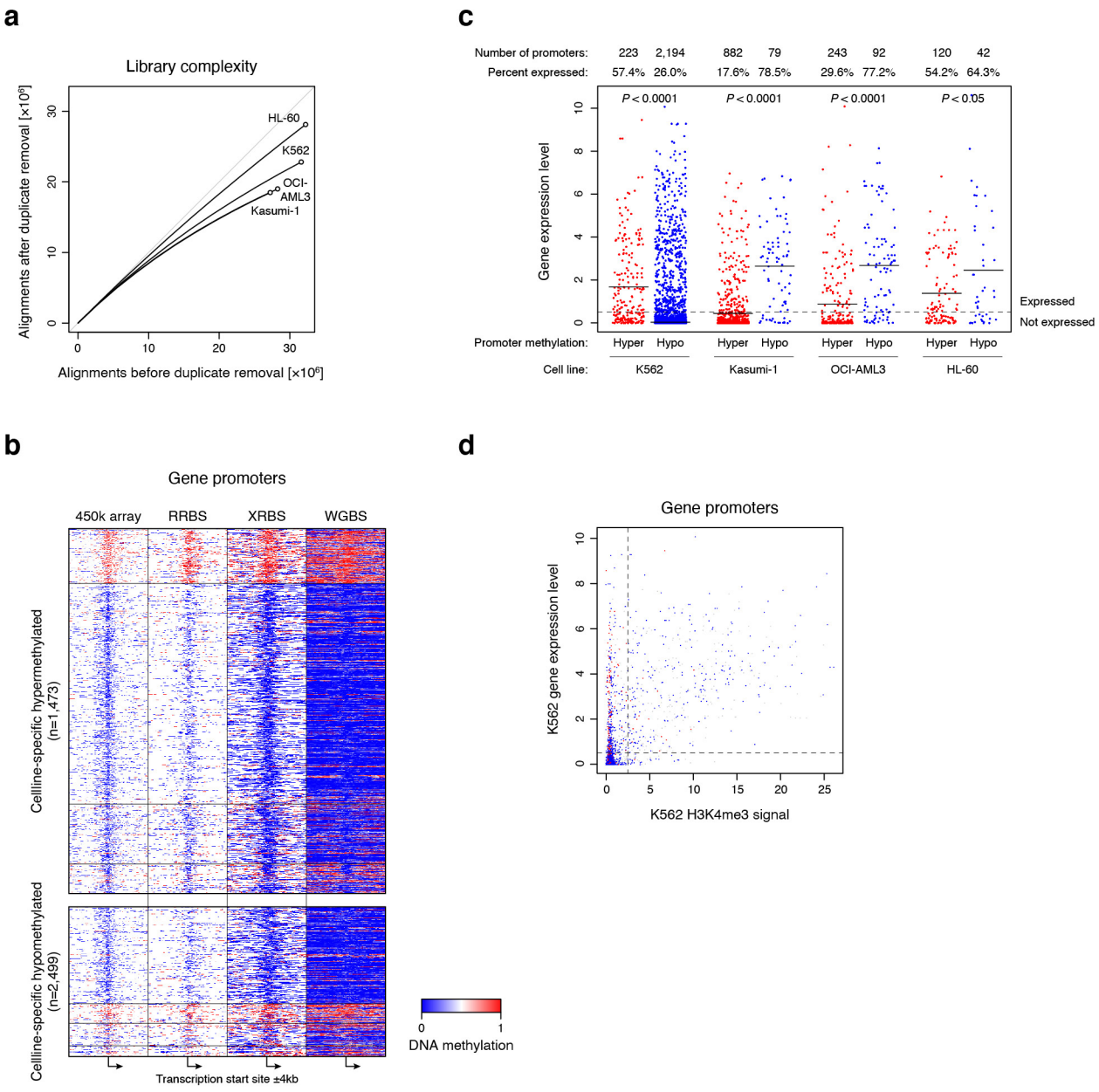
Supplementary Figure 5



Supplementary Fig. 5 | Characterization of decitabine treatment of cancer cell lines

A) Plot shows dose response curve for decitabine treatment of three cell lines Kasumi, HL-60, and OCI-AML3. Viability was measured using cell titer glo and is reported as luminescence relative to control DMSO treated cells (n=3 independently treated replicates, error bars represent standard deviation).
B) Images show HL60 and OCI-AML3 cells treated with 300 nM decitabine and a DMSO vehicle control. Morphology of decitabine treated cells similar to control, repeated three times. Scale bar is indicated and applies to all images.
C) Plot shows unique reads as a function of aligned reads in XRBS libraries from DMSO- and decitabine-treated HL-60 and OCI-AML3 cells.
D) Barplot shows average DNA methylation values across island (dark grey) and non-island (light grey) CpGs in DMSO- and decitabine-treated HL-60 and OCI-AML3 cells. For example, average methylation of non-island CpGs in HL-60 cells is reduced from 68.1% to 54.3% by decitabine treatment (20.2% reduction, n=1 library per treatment).
E) Heatmap shows correlation between Hi-C-derived first eigenvectors from K562 and HL-60 cell lines in 100kb-windows, indicating high agreement in compartment structure between both cell lines.
F) Heatmaps show correlation between average DNA methylation values and Hi-C-derived eigenvector in 100kb-windows for DMSO- (left) and decitabine-treated HL-60 cells (center). Heatmap on the right shows relative DNA methylation values of decitabine- and DMSO-treated cells. Despite compartment B showing lower methylation compared to compartment A at baseline, induced DNA hypomethylation with decitabine treatment affects compartment A and B equally.

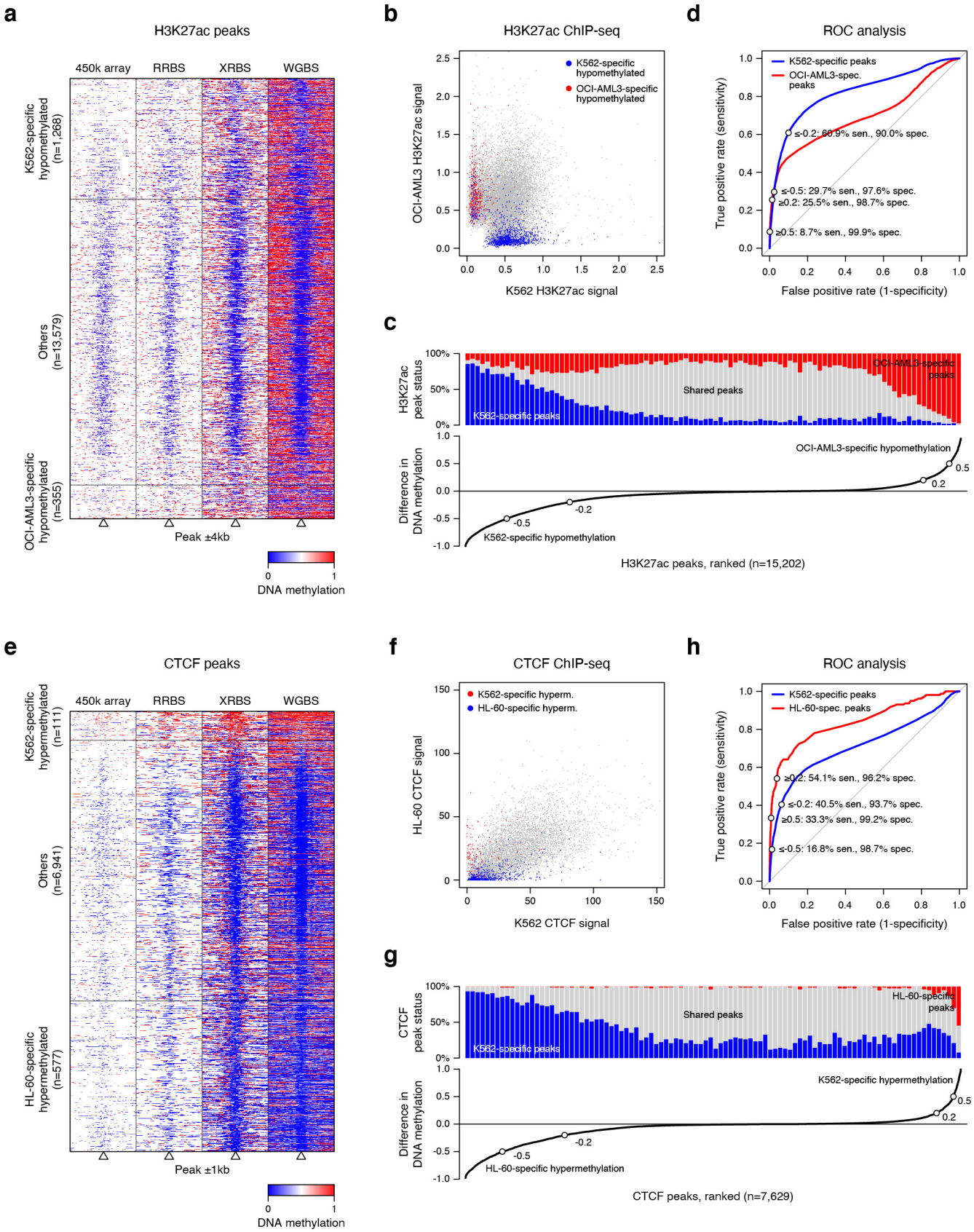
Supplementary Figure 6



Supplementary Fig. 6 | Differential DNA methylation of gene promoters

A) Plot shows unique reads as a function of aligned reads in 1,000 cell high-coverage libraries of four cell lines.
B) Heatmap depicts 8kb regions (rows, n=3,972 promoters) centered at transcription start sites that show cell line-specific hyper- or hypomethylation (as in Fig. 4a) and divided into 100 equally sized windows. Panels from left to right show methylation calls from 450k methylation array, RRBS, XRBS, and WGBS. All datasets except XRBS were retrieved from ENCODE [51].
C) Plot shows expression levels for genes that were associated with cell line-specific promoter hyper- and hypo-methylation. Genes with an expression level larger than 0.5 are considered as expressed. Average gene expression levels are indicated by horizontal lines. *P*-values were generated using a two-sided Mann-Whitney U test. In K562, the majority (74.0%) of hypomethylated promoters are associated with non-expressed genes, which is unique to this cell line, consistent with global hypo-methylation in K562.
D) Scatterplot compares gene expression level and H3K4me3 ChIP-seq signal for gene promoters that were identified as differentially methylated between all four cell lines. Individual promoters (dots) are colored if specifically hypermethylated (red) and hypermethylated (blue) in K562 cells. This analysis shows that promoters which are not expressed and specifically hypomethylated in K562 (n=1,624 promoters) are generally negative for the H3K4me3 histone mark (98.7%), whereas promoters that are hypomethylated and expressed (n=570) are more frequently positive for H3K4me3 (45.0%).

Supplementary Figure 7



Supplementary Fig. 7 | Evaluating the use of XRBS DNA methylation profiling to predict H3K27 acetylation and CTCF binding

A) Heatmap depicts 8kb regions (rows, n=15,202 peaks) centered on H3K27ac peaks, grouped into regions that are hypomethylated specifically in K562 or OCI-AML3 cells (as in Fig. 4b). Peaks that are not specifically hypomethylated ('Others') are downsampled for visualization. Regions are divided into 100 equally sized windows. Panels from left to right show: methylation calls from 450k methylation array, RRBS, XRBS, and WGBS. All datasets except XRBS were retrieved from ENCODE [51].

B) Scatterplot shows merged H3K27ac peaks from OCI-AML3 and K562 ChIP-seq datasets. Individual peaks (dots) are colored if specifically hypomethylated in K562 (blue) or OCI-AML3 (red) cells.

C) Line plot (bottom) shows difference in methylation between K562 and OCI-AML3 cells over merged H3K27ac peaks (n=15,202 peaks). Of these peaks, 7.5% and 2.1% are specifically hypomethylated in K562 (methylation difference ≤-0.5) and OCI-AML3 (≥0.5) cells, respectively. Bar plot (top) shows the fraction of cell line-specific H3K27ac peaks within 100 equally sized bins grouped by difference in methylation. Shared peaks are indicated in gray.

D) Receiver operating characteristic (ROC) curve shows performance of predicting cell line-specific H3K27ac peaks based on difference in DNA methylation over peaks that are covered by XRBS. Sensitivity and specificity are indicated at different thresholds (±0.2 and ±0.5, as in panel C).

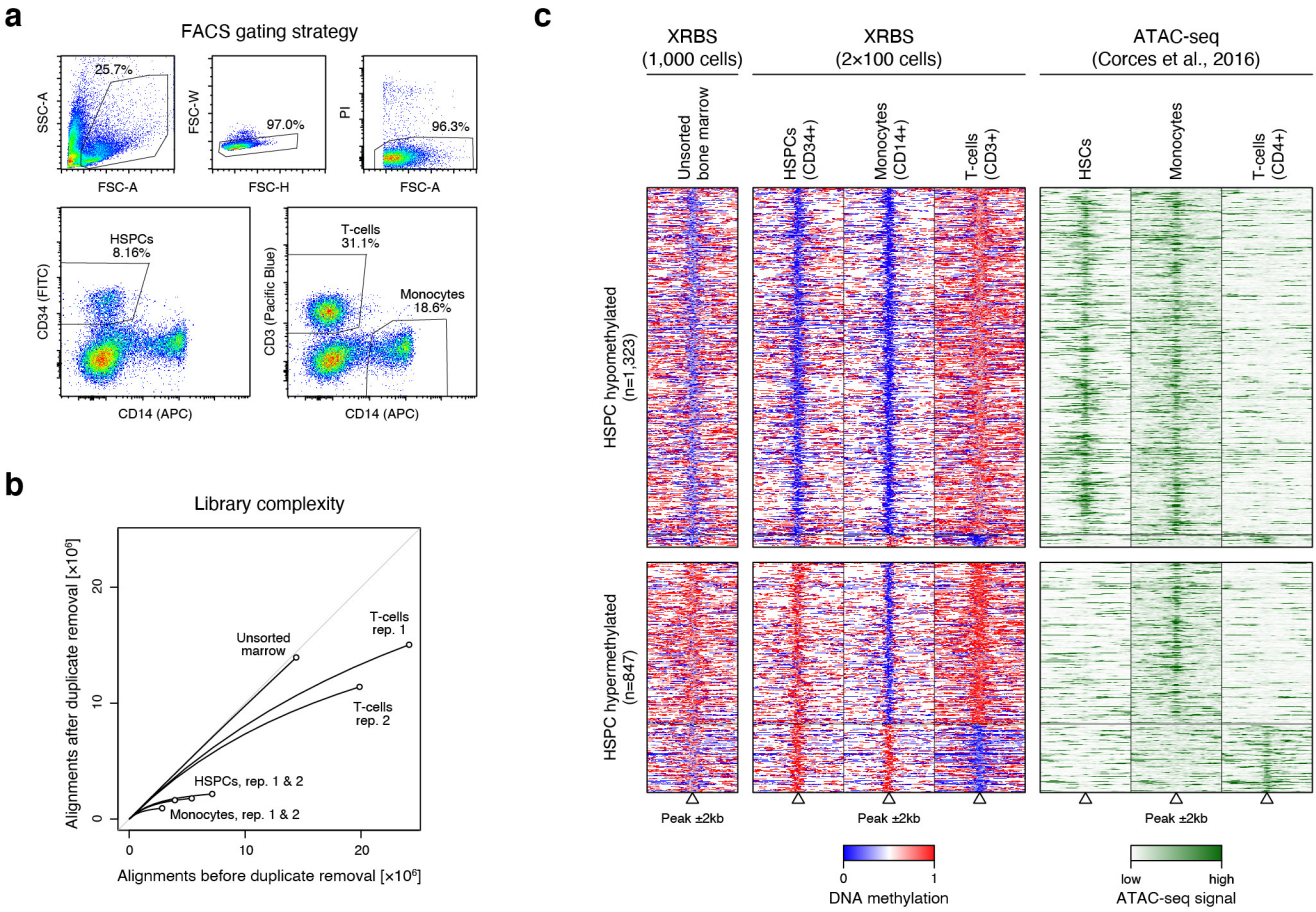
E) Heatmap depicts 2kb regions (rows, n=7,629 peaks) centered at merged CTCF peaks from K562 and HL-60 Chip-seq datasets. Individual peaks (dots) are colored if specifically hypermethylated in K562 or HL-60 cells (as in Fig. 4c). Peaks not specifically hypermethylated ('Others') are downsampled for visualization. Panels from left to right show methylation calls from 450k methylation arrays, RRBS, XRBS, and WGBS. All datasets except XRBS were retrieved from ENCODE [51].

F) Scatterplot shows merged CTCF peaks from K562 and HL-60 ChIP-seq datasets. Individual CTCF binding sites (dots) are colored if specifically hypermethylated in K562 (red) or HL-60 (blue) cells.

G) Line plot (bottom) shows difference in methylation between K562 and HL-60 cells over merged CTCF peaks (n=7,629 peaks). Bar plot (top) shows the fraction of cell line-specific CTCF peaks within 100 equally sized bins grouped by difference in methylation. Shared peaks are indicated in gray.

H) ROC curve shows performance of predicting cell line-specific CTCF peaks based on difference in DNA methylation over peaks that are covered by XRBS. Sensitivity and specificity are indicated at different thresholds (±0.2 and ±0.5, as in panel G).

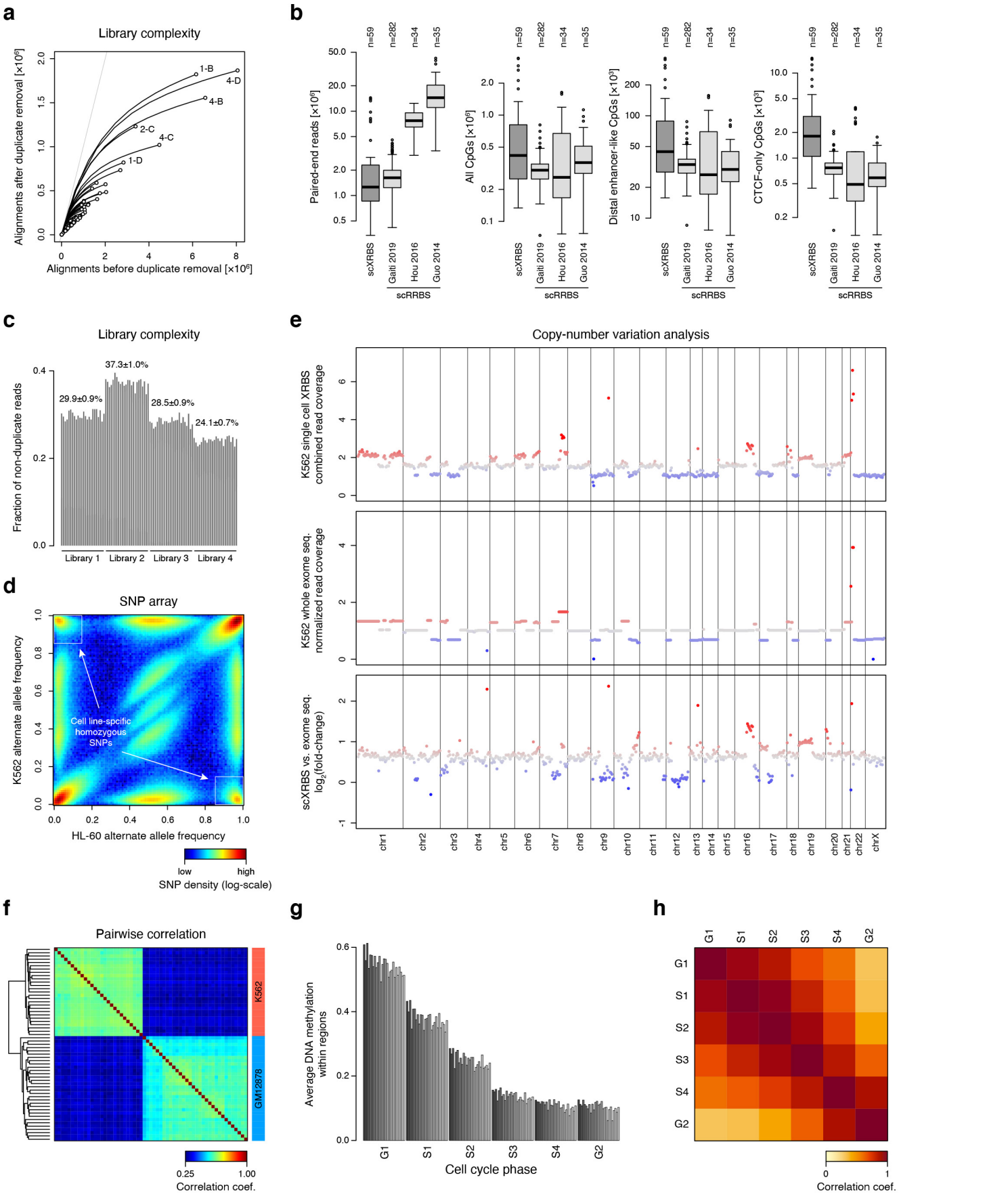
Supplementary Figure 8



Supplementary Fig. 8 | XRBS profiling of limited human bone marrow cell types

A) Plots show the gating strategy for fluorescence assisted cell sorting (FACS) of human bone marrow of CD34+ HSPCs, CD3+ T cells, and CD14+ monocytes. Singlets (FSC-W vs. -H) and viable cells (PI vs. FSC-A) were sorted based on cell surface marker signal.
B) Plot shows unique reads as a function of aligned reads in libraries from unsorted human bone marrow, HSPCs, monocytes, and T cells. Libraries were generated from 100 sorted cells. 1000 cells were used for the unsorted bone marrow library.
C) Heatmap depicts 4kb regions (rows, n=2,170 regions) centered over elements defined in the ENCODE SCREEN database. Only differentially methylated elements between monocytes and T cells are shown. Elements were stratified by their methylation status in HSPCs (hypomethylated: top; hypermethylated: bottom). Methylation levels of the unsorted bone marrow are shown for comparison (left). ATAC-seq signal for sorted hematopoietic stem cells (HSCs), monocyte, and CD4+ T cells (obtained from [36]

Supplementary Figure 9



Supplementary Fig. 9 | Evaluation of single cell XRBS profiles

A) Plot shows unique reads as a function of aligned reads in single cell XRBS profiles (n=96 cells). With greater sequencing depth the fraction of unique reads decreases, as the chance of sampling a non-unique read (i.e. PCR duplicate) increases.
B) Boxplots compare DNA methylation profiles from human scXRBS (n=59 cells) and three published scRRBS datasets generated from human cells: Chronic lymphocytic leukemia (n=282 cells) [49], hepatocellular carcinoma and HepG2 cells (n=34 cells) [45], and oocytes, sperm and pronuclei (n=35 cells) [48]. Single cells from Hou et al. were generated using the scTrio-seq protocol that in part resembles scRRBS. Only CpGs within 75 bases of an MspI cut site were considered for scRRBS libraries to adjust for differences in read lengths. Libraries from Gaiti et al. were sequenced at 2x51 bases. Left plot shows the number of paired-end reads sequenced for each cell. Other plots show the number of CpGs covered (≥1-fold) across all CpGs in the genome, CpGs within distal enhancer-like regions, and CpGs within 'CTCF-only' regions (SCREEN database). Both strands of a CpG dinucleotide are assessed individually. Although sequenced at the lowest depth, scXRBS libraries on average capture the most CpGs, particularly in CpG-sparse regions. Boxplots were generated in R using default settings: Bounds of box and thick horizontal line represent 25th, 75th, and 50th percentile of observations, whiskers represent minimum and maximum observations, and outliers are indicated as dots.
C) Barplot shows the fraction of unique reads (i.e. reads not representing PCR duplicates) per single cell library. Within the same PCR reaction, the duplicate rate was very similar, irrespective of the total number of aligned reads per single cell. Each bar plot represents a single cell XRBS library. Twenty four barcoded cells were in each of 4 independent libraries.
D) Heatmap compares alternate allele frequencies from SNP array data for K562 and HL-60 cell lines. Cell line-specific homozygous alleles are indicated by white boxes boxes and were used for single cell SNP analysis in Fig. 5d.
E) Plots show copy number variation calls from combined single cell XRBS profiles (top) and whole exome sequencing data (middle) for K562 cells. A number of chromosomes show differences in copy number between XRBS and whole exome sequencing (bottom). However, these differences likely represent true copy number variations between K562 cells used for these experiments.
F) Heatmap shows pairwise correlation coefficients of single cell methylation profiles. Dendrogram shows unsupervised clustering. Single cell XRBS profiles cluster by cell type.
G) Barplot shows K562 single cell average DNA methylation values within various early and late replicating regions. Each bar represents an individual K562 single cell library. There are 32 single cell libraries plotted for each cell cycle phase.
H) Heatmap shows pairwise correlation of average DNA methylation values within various early and late replicating regions. Late replicating regions (G2 phase) cluster separately. These results suggest that one source of single cell DNA methylation heterogeneity is decreased fidelity of maintenance DNA methylation in late replicating domains.