Supplementary Figures

Supplementary Fig. 1 | Fraction of CpG sites analyzed by informME. Percentage of CpG sites that were analyzed by informME in each WGBS sample. Notably, the percentage of CpG sites with data analyzed by informME was consistent across all cancer, normal, and cell-line (Reh) samples.

Supplementary Fig. 2 | *ERG* demonstrates increased methylation stochasticity in *ETV6-RUNX1* ALL. Observed WGBS data at twelve contiguous CpG sites at a locus inside *ERG* [chr21: 39,830,065 - 39,830,570] demonstrate consistent methylation in CD19+ cells but increased methylation stochasticity in ALL (red marks: methylated CpG sites; blue marks: unmethylated CpG sites). Results are displayed using the Integrated Genomics Viewer v2.8.6 (Broad Institute) set to its 'CG bisulfite mode' with default settings.

Supplementary Fig. 3 | Comparison of two approaches to DNA methylation analysis. Examples of heat maps of mean methylation level (MML) and normalized methylation entropy (NME), computed from the probability distribution $P_X(\boldsymbol{x})$ of the methylation patterns (left column) and the probability distribution $P_L(l)$ of the methylation levels (right column). An analysis region is considered here to contain 4 CpG sites that exhibit constant CpG density ρ and distance d, whereas methylation is characterized by a potential energy landscape with parameters $-3 \leq A = (\alpha + \beta\rho) \leq 3$ and $0 \leq B = \frac{\gamma}{d} \leq 3$. The resulting MML values are identical, whereas the NME values are comparable, thus capturing a similar behavior in methylation stochasticity. Note that smaller negative values of parameter A result in decreased mean methylation level, indicating that fewer CpG sites are methylated on the average, whereas larger positive values of A result in increased mean methylation level, indicating that more CpG sites are methylated on the average. In both cases, this behavior is associated with increasingly reduced normalized methylation entropy and is exacerbated with larger values of parameter B due to increased correlation.

Supplementary Fig. 4 | Distributions of methylation statistics demonstrate global properties of methylation stochasticity in *ETV6-RUNX1* ALL. a,b, Distributions of mean methylation level (MML), normalized methylation entropy (NME), and methylation sensitivity (MSI) values in all normal control and *ETV6-RUNX1* samples: genome-wide, in (**a**), and within selected genomic features, in (**b**). Center lines, median; boxes, interquartile range (IQR); whiskers, 1.5 × IRQ. Enhancer and promoter annotations were chosen to be among ChromHMM annotations[65] with the highest median observed methylation discordance (Supplementary Table 15).

Supplementary Fig. 5 | Classifications of the methylation state show details of mean methylation level and normalized methylation entropy differences between *ETV6-RUNX1* and normal control samples. a,b, Percentages of analysis regions in the *ETV6-RUNX1* and normal control samples classified in terms of mean methylation level (MML), in (**a**), and normalized methylation entropy (NME), in (**b**), within selected genomic features. Classification uses a previously proposed method[15] based on the shape of the probability distribution of methylation level. Genomic regions classified as 'variably methylated' exhibit high methylation variability, including bistability.

Supplementary Fig. 6 | Quantifying methylation discordance in ALL. Distribution of the number N of analysis regions with respect to their Jensen-Shannon distance (JSD) and absolute differential mean methylation level (dMML) values, computed from all *ETV6-RUNX1*/CD19-1 comparisons over selected genomic features. Many analysis regions that exhibit similar differences in mean methylation level are associated with a wide range of Jensen-Shannon distance values demonstrating that the mean methylation level is not the only factor influencing methylation discordance in ALL.

Supplementary Fig. 7 | Methylation discordance mapping to *ERG* in *ETV6-RUNX1* ALL. UCSC genome browser images of Jensen-Shannon distance values and corresponding DMRs within a chromosomal region associated with *ERG* in all ten *ETV6-RUNX1* ALL samples when compared to CD19-1. Darker colors indicate more statistically significant DMRs.

Supplementary Fig. 8 | Examples of methylation discordance mapping to regulatory regions in *ETV6-RUNX1* ALL. a,b, UCSC genome browser images of two chromosomal regions corresponding

to *EBF2*, in (**a**), and *NR2F2*, in (**b**), exhibiting significant Jensen-Shannon distance values in an *ETV6-RUNX1*/CD19 comparison (ALL-45 vs. CD19-1) and associated with consistent hypermethylation, increased normalized methylation entropy (dNME tracks), and loss in methylation sensitivity (dMSI tracks) in ALL. Observed methylation discordance is localized at the promoters of *EBF2* and *NR2F2*, as well as and at binding sites for EZH2 and SUZ12 in the case of *NR2F2*.

Supplementary Fig. 9 | Effect of cell purity on single-cell analysis. A UMAP plot[79] of the single-cell RNA sequencing data (Supplementary Table 6a) shows that the four samples (ALL-45, ALL-289, CD19-4, and CD19-5) used for single-cell analysis are relatively homogenous in terms of gene expression, thus affirming the absence of a significant effect of cell purity on the analysis. Notably, gene expression in the ALL-289 sample is more similar to gene expression in the CD19-4 and CD19-5 samples and less similar to gene expression in the ALL-45 sample.

Supplementary Fig. 10 | Gene expression means and variances in ALL. Scatter plots of gene expression means and variances corresponding to two *ETV6-RUNX1* samples (ALL-45, ALL-289) compared to normal CD19 controls. Blue dots represent genes. Means and variances were obtained by bulk and single-cell RNA sequencing, respectively.

Supplementary Fig. 11 | Distributions of methylation statistics in three cytogenetic subtypes of ALL: hyperdiploid, *TCF3-PBX1*, and dicentric chromosome (9;20). a,b, Distributions of mean methylation level (MML), normalized methylation entropy (NME), and methylation sensitivity (MSI) values: genome-wide, in (**a**), and within selected genomic features, in (**b**). Center lines, median; boxes, interquartile range (IQR); whiskers, 1.5 × IRQ. Enhancer and promoter annotations were chosen to be among ChromHMM annotations[65] with the highest median observed methylation discordance (Supplementary Table 15).

Supplementary Fig. 12 | Distributions of differential methylation statistics in four cytogenetic subtypes of ALL. Distributions of Jensen-Shannon distance (JSD), differential mean methylation level (dMML), differential normalized methylation entropy (dNME), and differential methylation sensitivity (dMSI) values genome-wide and within selected genomic features in an *ETV6-RUNX1*/CD19 comparison (ALL-45 vs. CD19-1), in a hyperdiploid/CD19 comparison (ALL-337 vs. CD19-1), in a *TC3-PBX1*/CD19 comparison (ALL-169 vs. CD19-1), and in a dic(9;20)/CD19 comparison (ALL-199 vs. CD19-1). Green, positive values; red, negative values; center lines, median; boxes, interquartile range (IQR); whiskers, 1.5 × IRQ. Differential behavior is similar in all comparisons with some noticeable differences especially over bivalent promoters (PromBiv). Enhancer and promoter annotations, obtained from Ernst and Kellis[65], were chosen to be among the ones with the highest median observed methylation discordance (Supplementary Table 15).

Supplementary Fig. 13 | Distributions of analysis regions with methylation discordance in four cytogenetic subtypes of ALL. Percentage of analysis regions with significant methylation discordance, measured by the Jensen-Shannon distance, within selected genomic features when comparing *ETV6-RUNX1*, hyperdiploid, TCF3-PBX1, and dicentric chromosome (9;20) ALL with CD19. Although differential behavior is similar across cytogenetic subtypes, there are noticeable differences between subtypes and samples within genomic features. Enhancer and promoter annotations, obtained from Ernst and Kellis[65], were chosen to be among the ones with the highest median observed methylation discordance (Supplementary Table 15).

Supplementary Fig. 14 | UHRF1 silencing in the Reh cell line and abrogation of clonogenicity. **a**, Western blot for UHRF1 and β-actin control in Reh cells following lentriviral CRISPR/Cas9 disruption of *UHRF1* (UHRF1-KO) in non-targeting (NT) control shows silencing of *UHRF1* in UHRF1-KO Reh cells. **b**, A representative light-micrograph of a methylcellulose assay of NT and UHRF1-KO Reh cells demonstrates complete abrogation of clonogenicity in UHRF1-KO Reh cells. **c**, Number of colonies formed in multiple methylcellulose assays of NT and UHRF1-KO Reh cells.

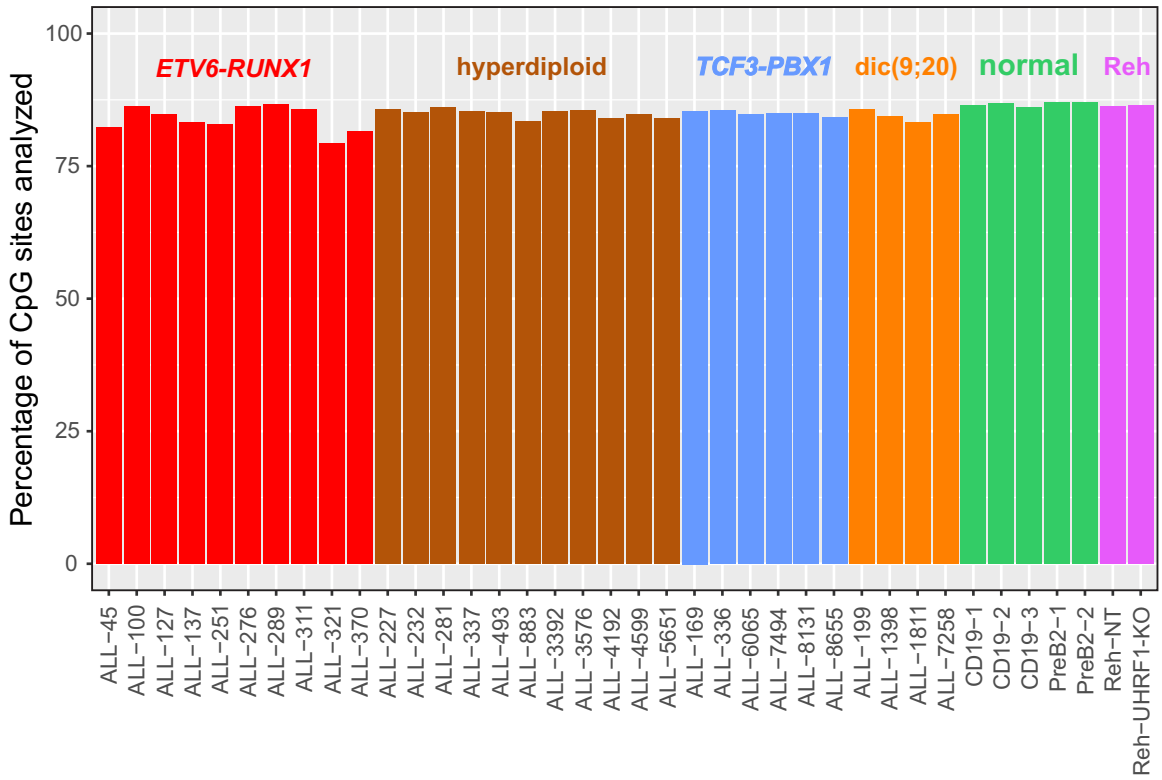
Supplementary Fig. 15 | Functional epigenetic modules identified in *ETV6-RUNX1* ALL. Significant functional epigenetic modules (FEMs) identified in *ETV6-RUNX1* ALL include a *TCF3* FEM with 89 genes, in (**a**), a *PLXNB1* FEM with 10 genes, in (**b**), a *LYN* FEM with 74 genes, in (**c**), an

MME FEM with 21 genes, in (**d**), and a *SLC9A3R2* FEM with 15 genes, in (**e**). Colors indicate differential methylation (blue core) and differential gene expression (red ring).

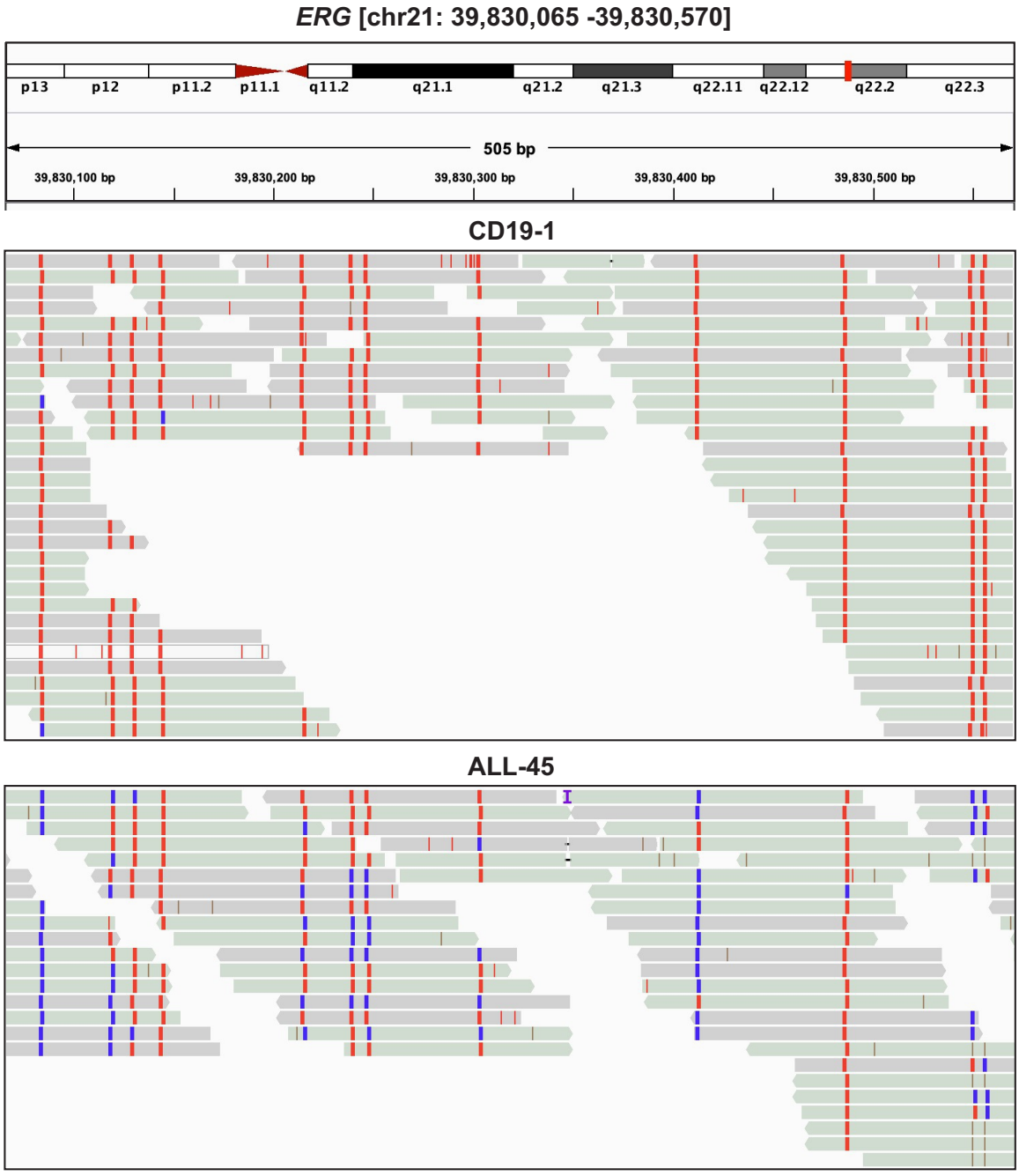
Supplementary Fig. 16 | Flow cytometry gating. Flow cytometry gating strategy based on forward scatter (FSC), side scatter (SSC), live/dead viability stain, CD19+, and CD10+ for the indicated ALL samples.

Supplementary Fig. 17 | Quality assessment of Bismark methylation extractor. Representative M-bias plots for read 1 and read 2, associated with the ALL-127 and CD19-1 samples, generated by the Bismark methylation extractor using the 'mbias_only' flag.

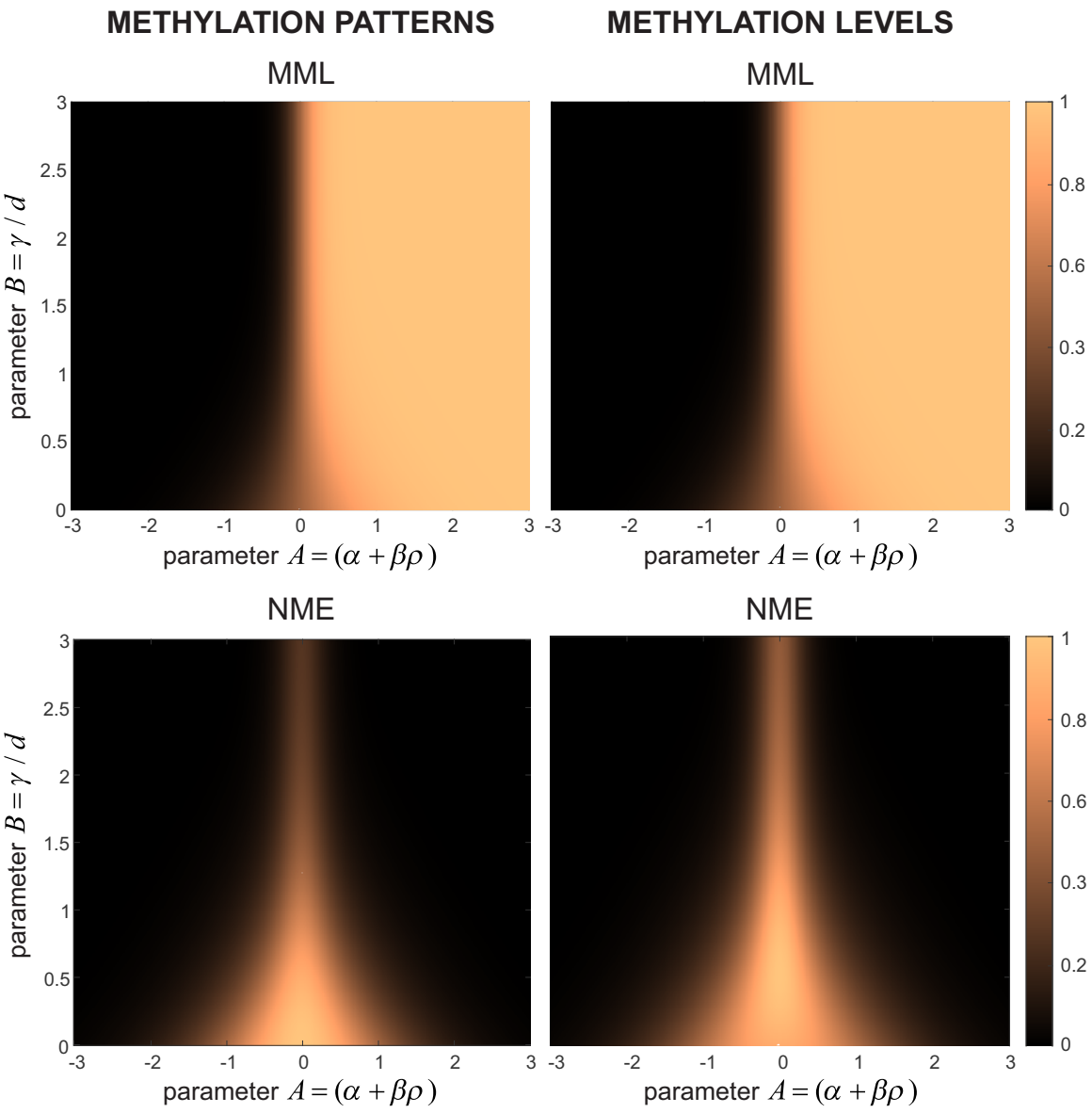
Supplementary Fig. 18 | Computation of gene variability. Local polynomial regression (LOESS) curves associated with gene variability computation, fitted to pairs of $\log_2$-scaled weighted variances and means in four *ETV6-RUNX1* and CD19 samples. The variability level of a gene in a sample is computed as the vertical distance (i.e., residual) of the gene (dot) from the corresponding LOESS curve.



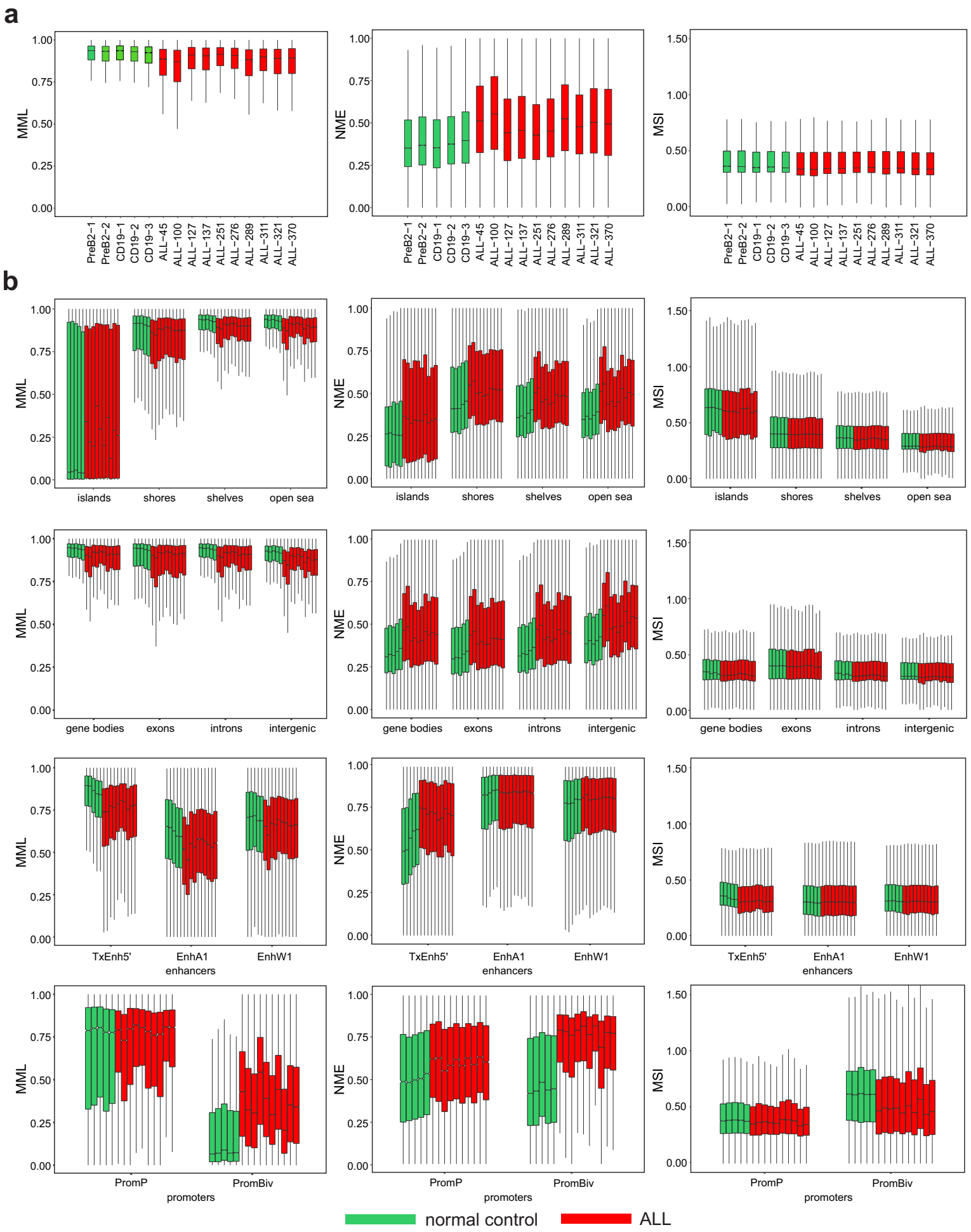
SUPPLEMENTARY FIGURE 1



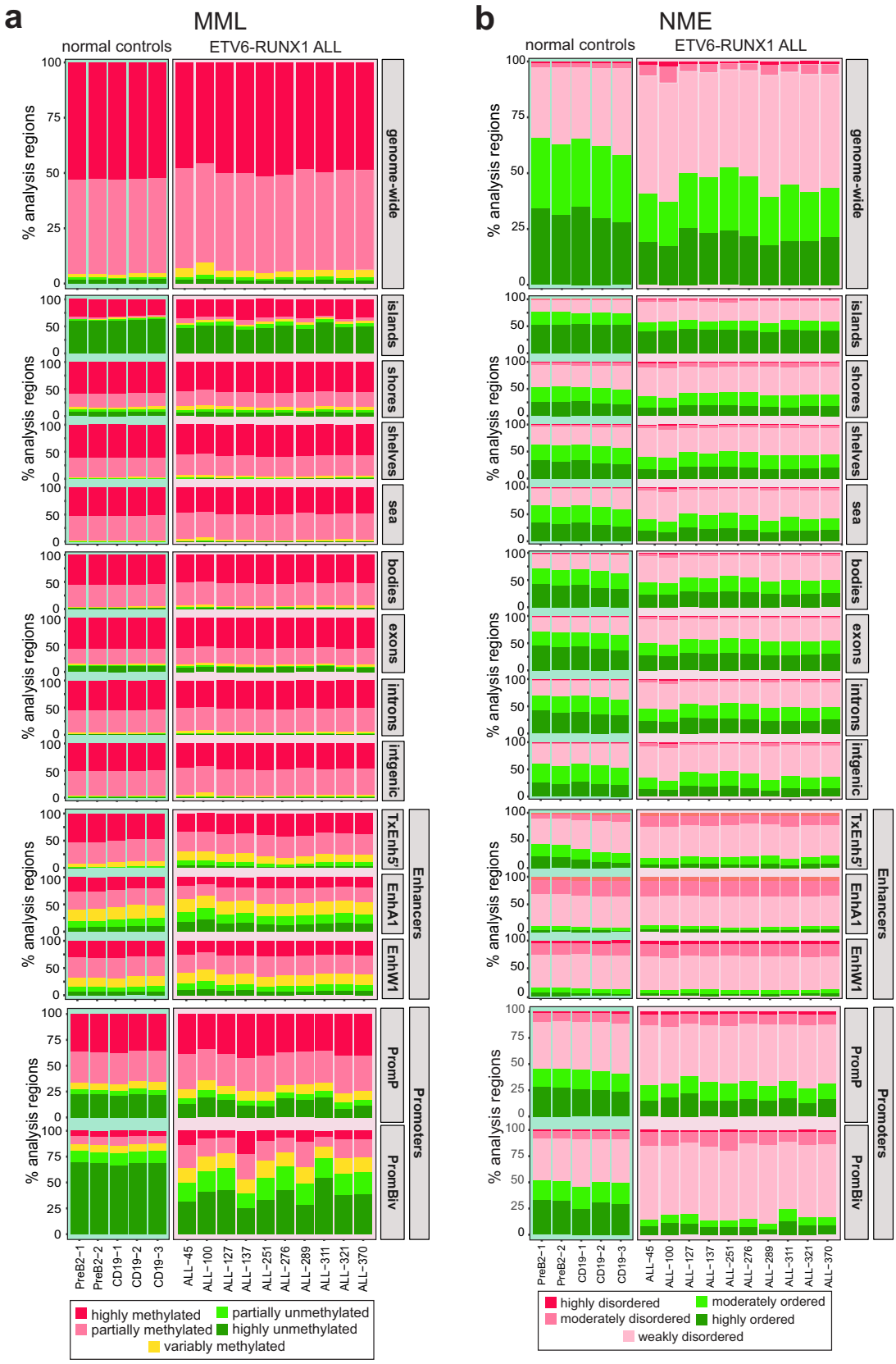
SUPPLEMENTARY FIGURE 2



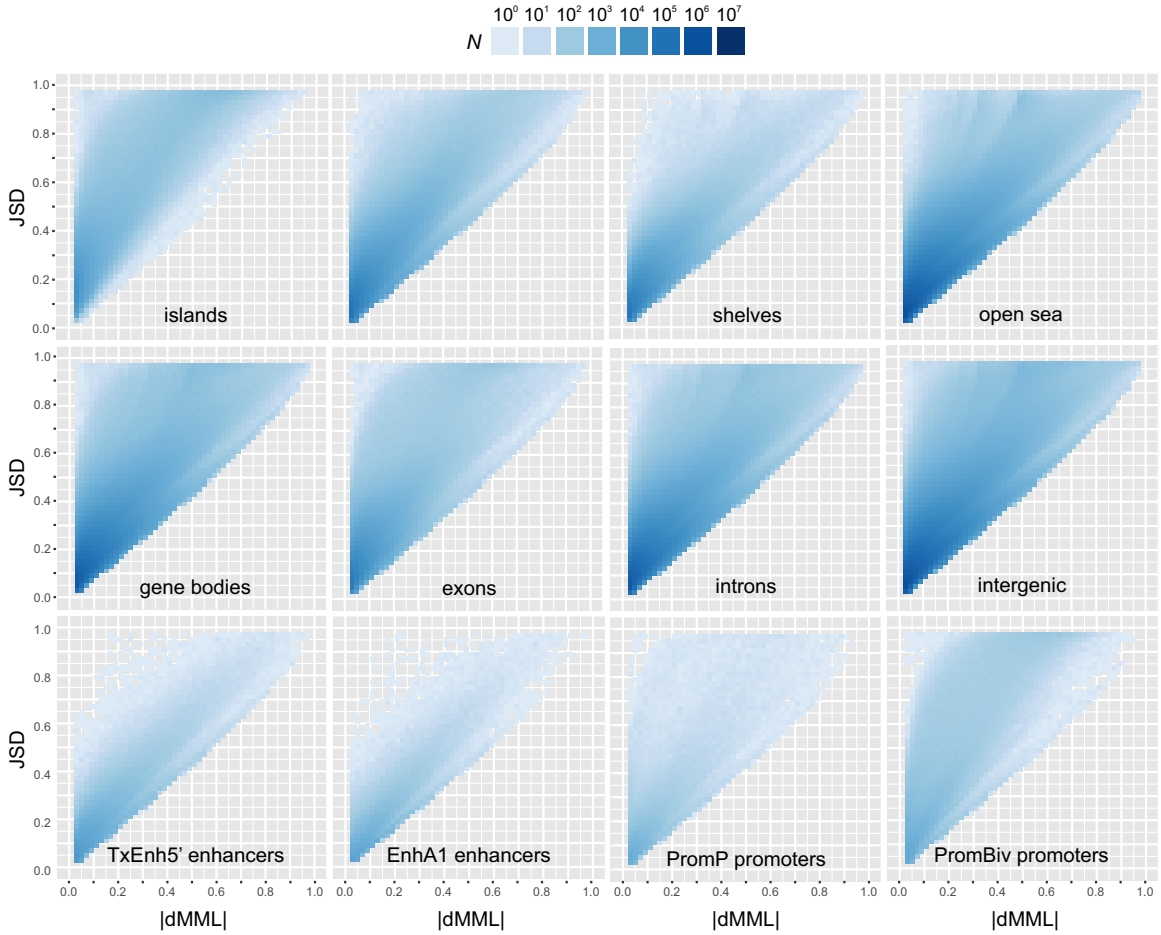
SUPPLEMENTARY FIGURE 3



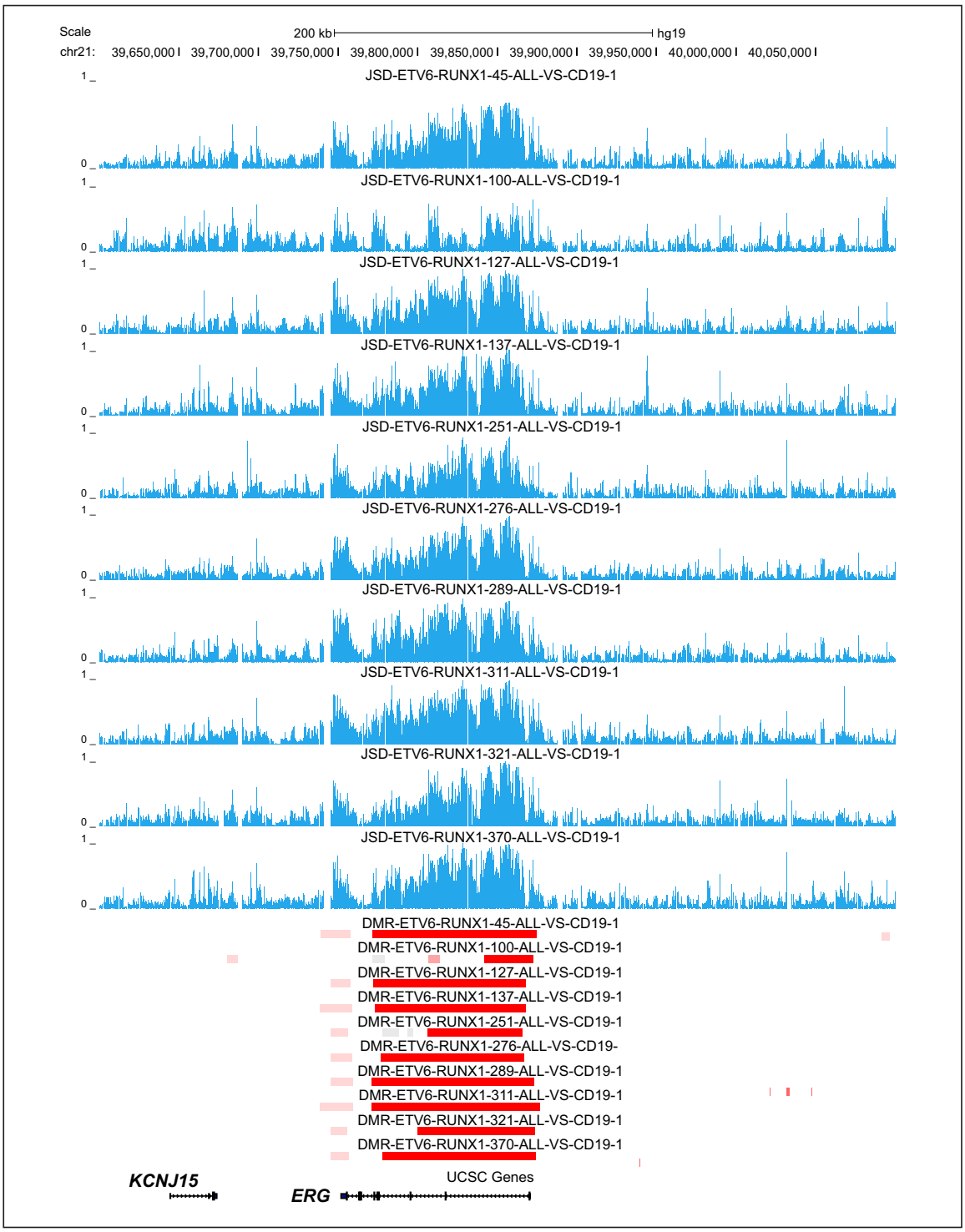
SUPPLEMENTARY FIGURE 4



SUPPLEMENTARY FIGURE 5

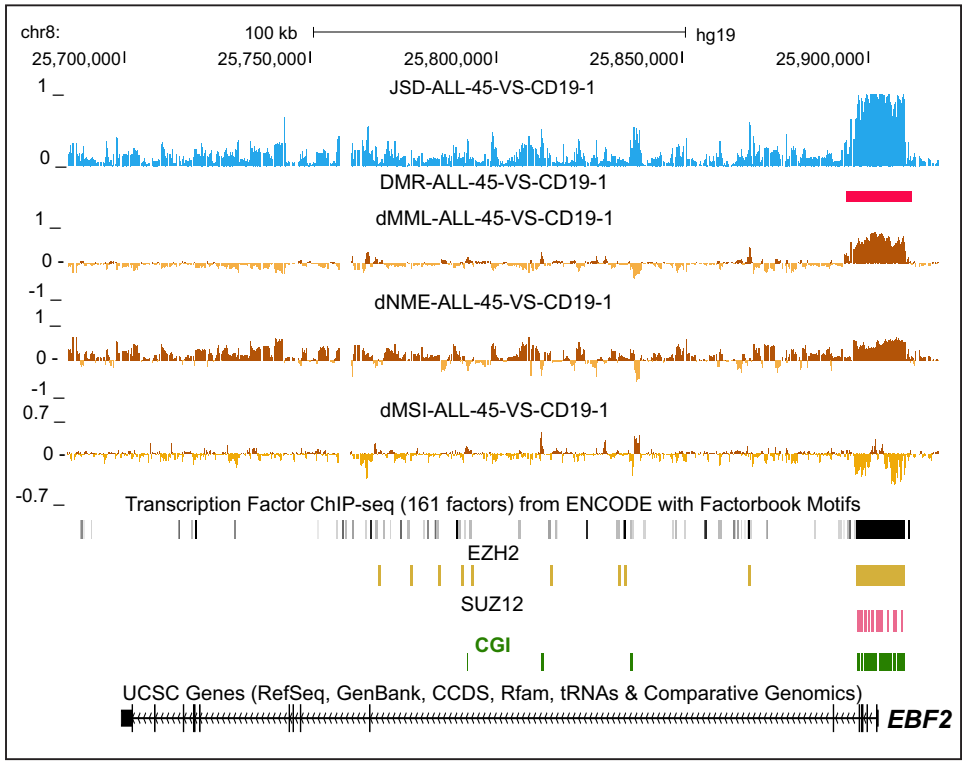


SUPPLEMENTARY FIGURE 6

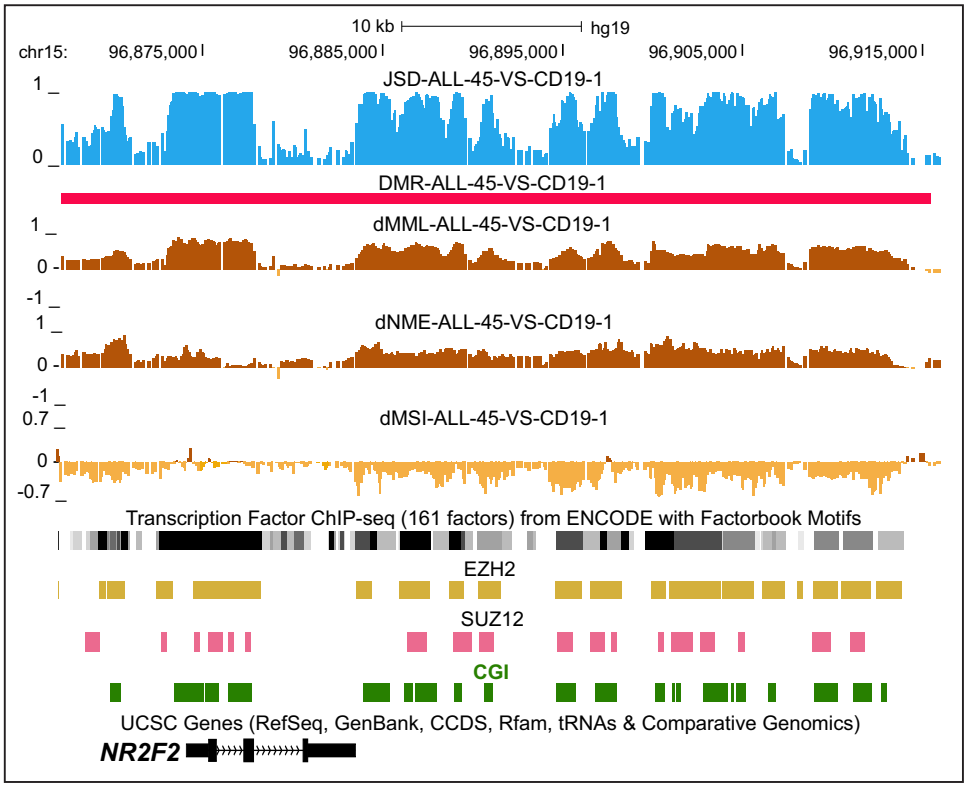


SUPPLEMENTARY FIGURE 7

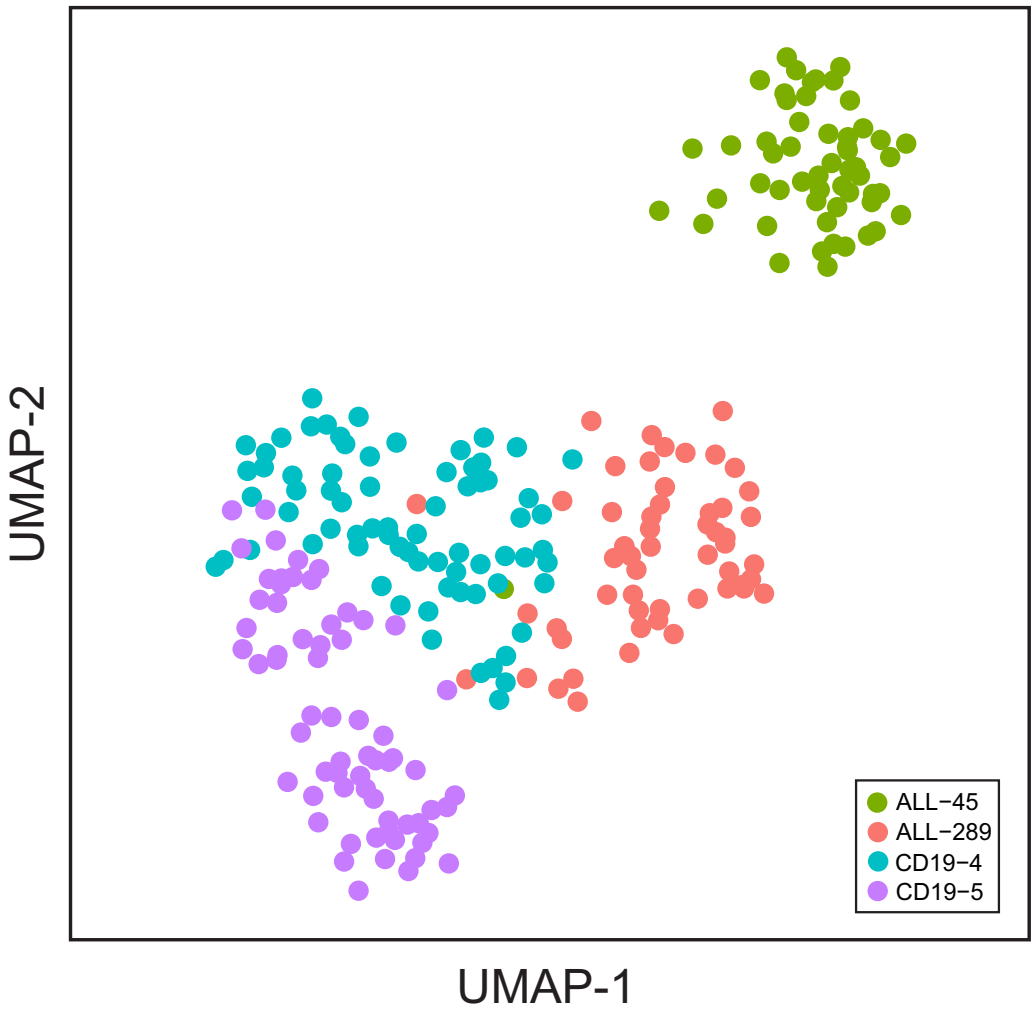
a



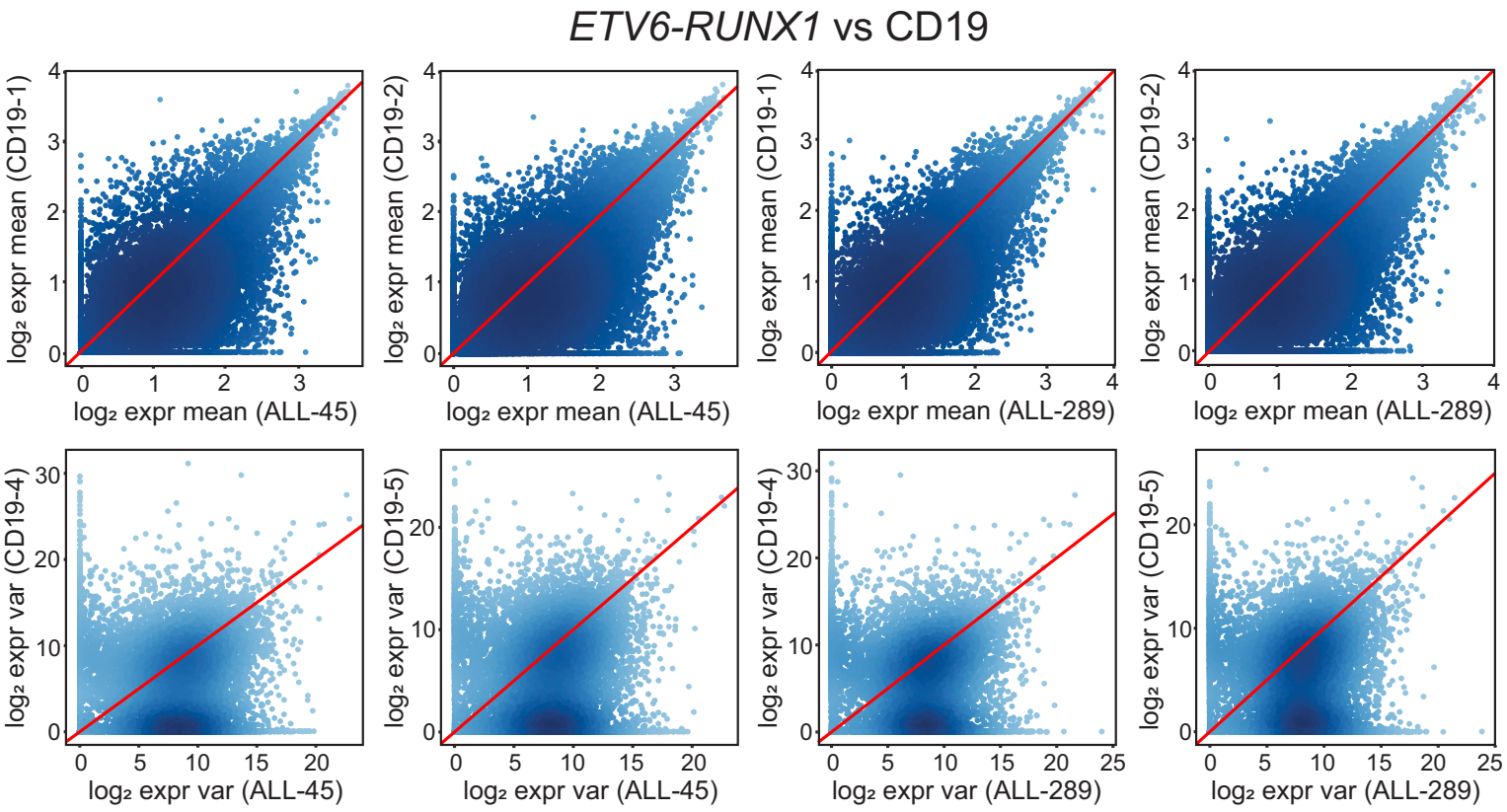
b



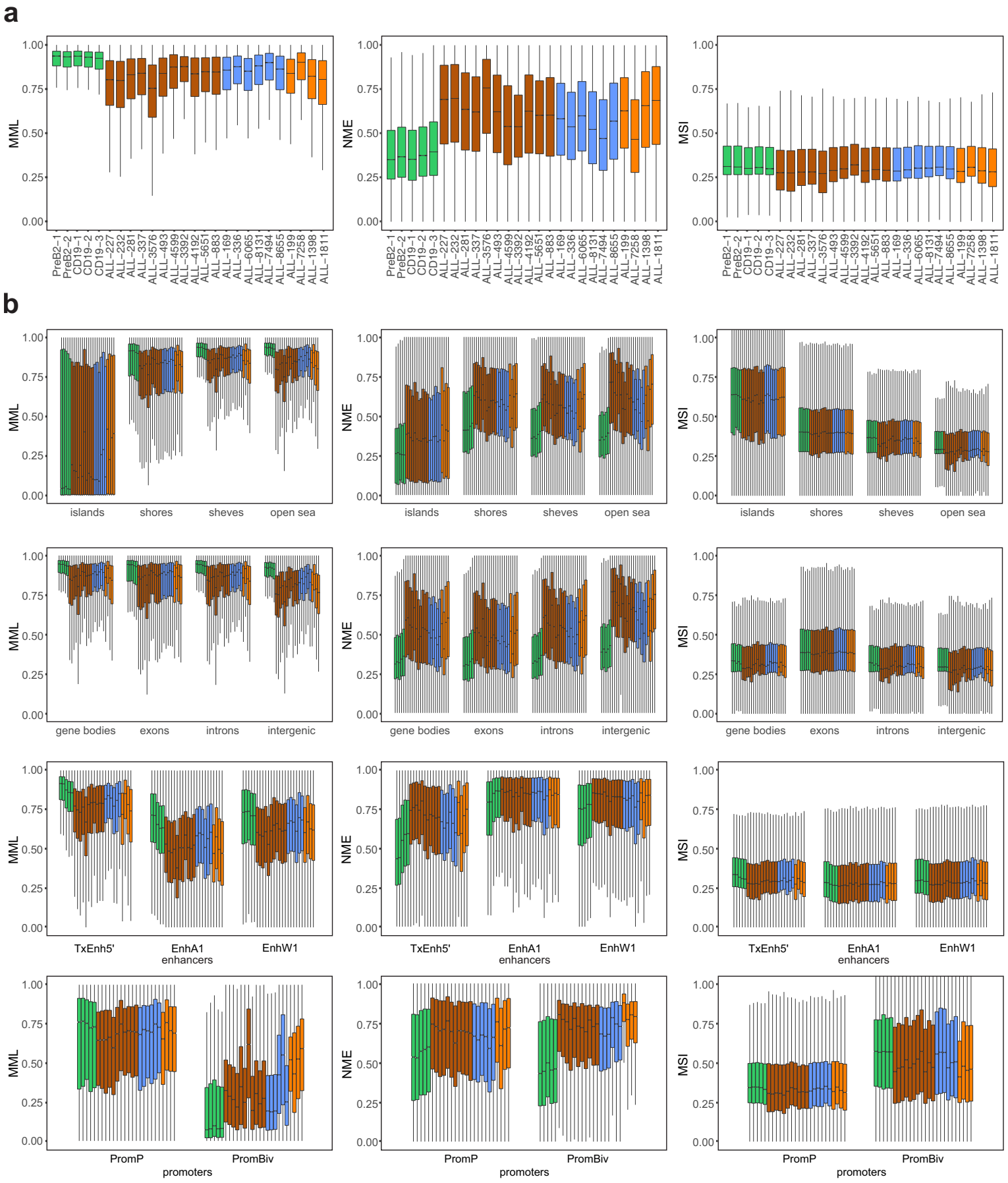
SUPPLEMENTARY FIGURE 8



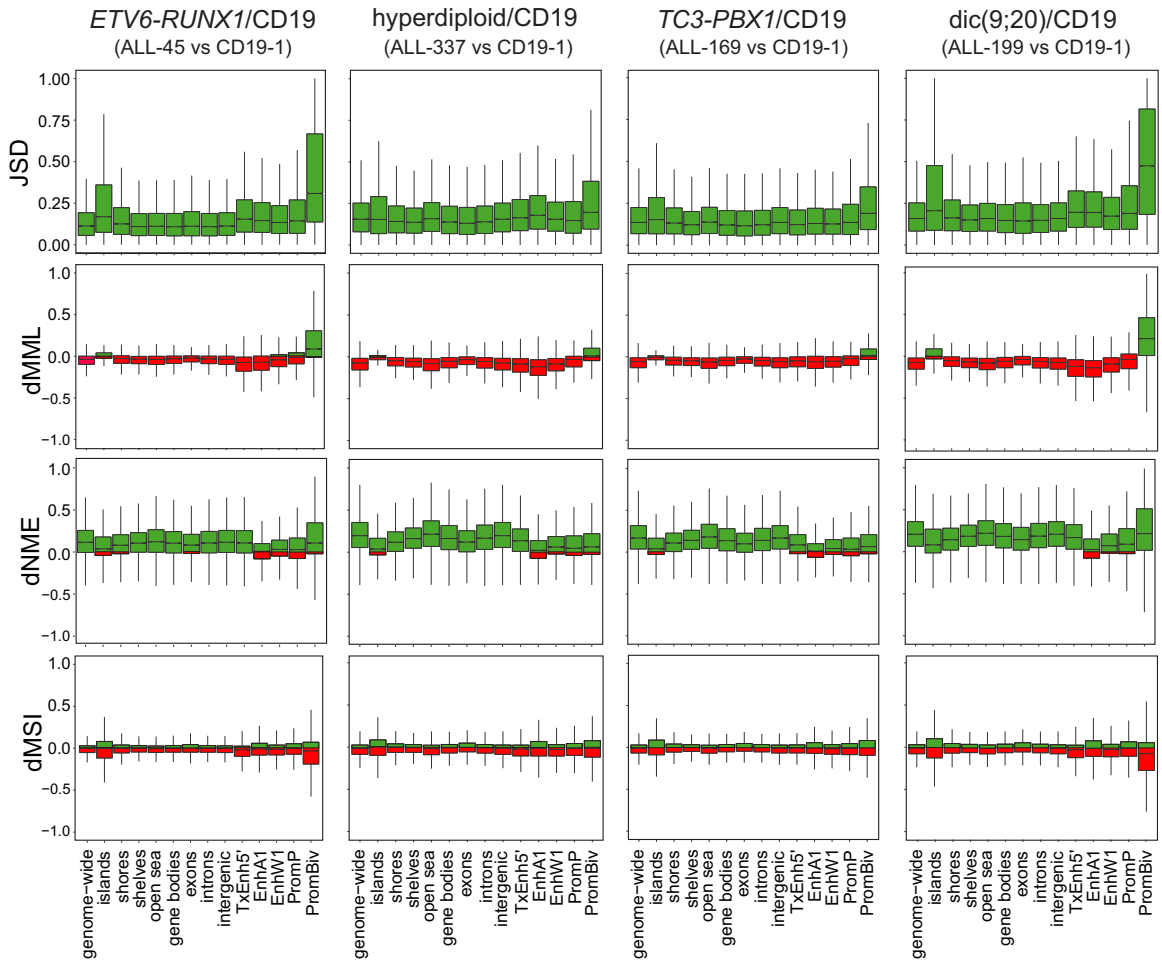
SUPPLEMENTARY FIGURE 9



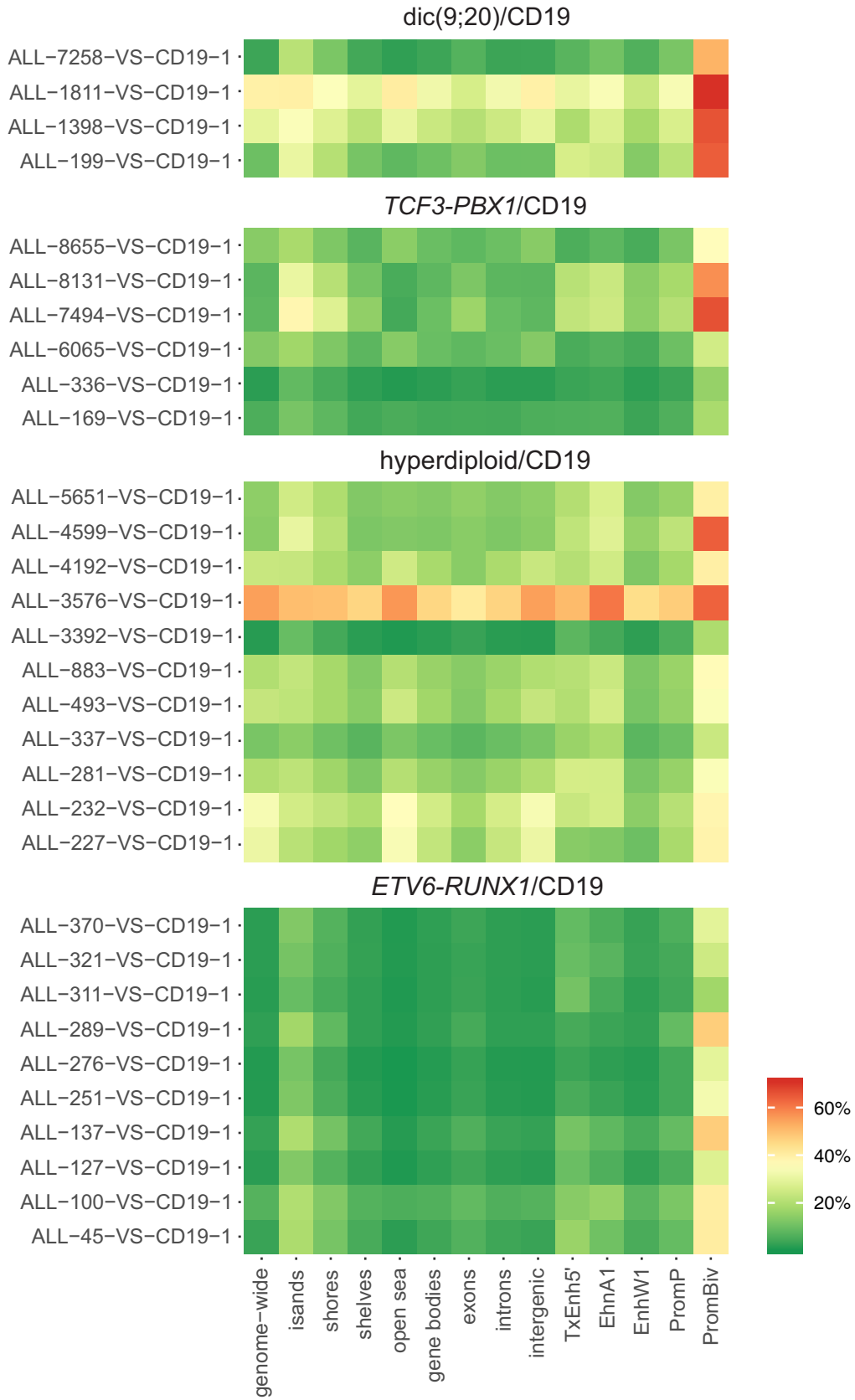
SUPPLEMENTARY FIGURE 10



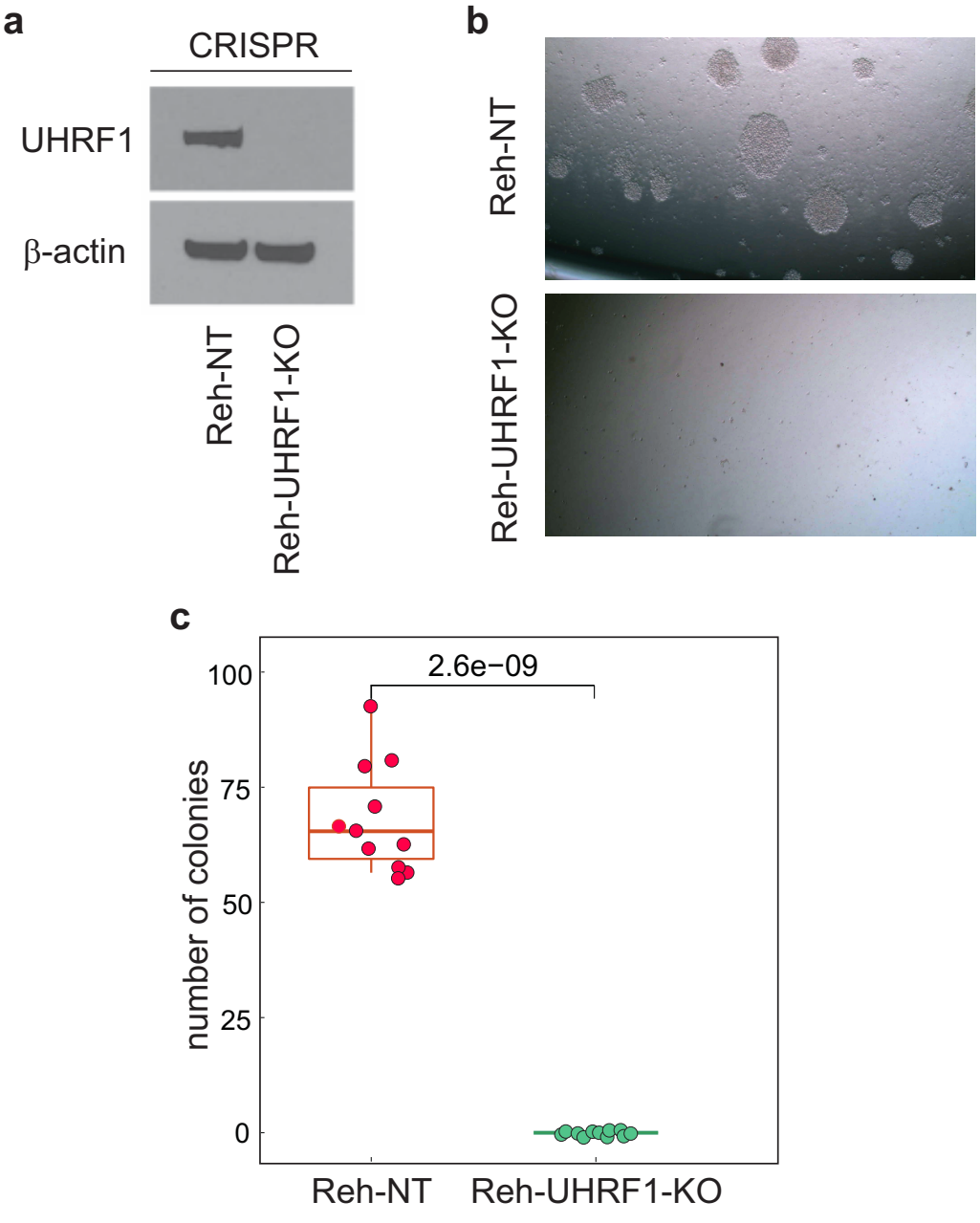
SUPPLEMENTARY FIGURE 11



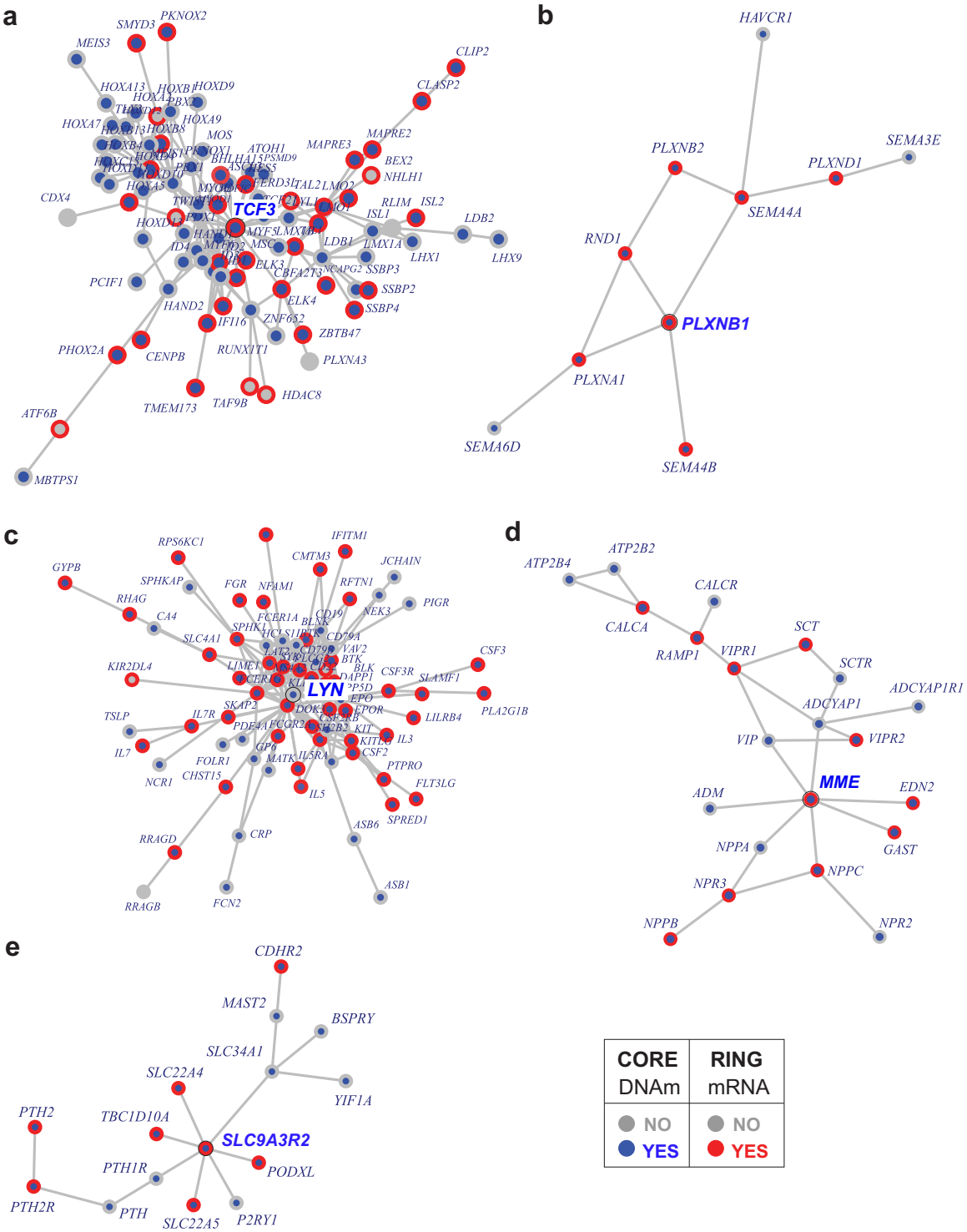
SUPPLEMENTARY FIGURE 12



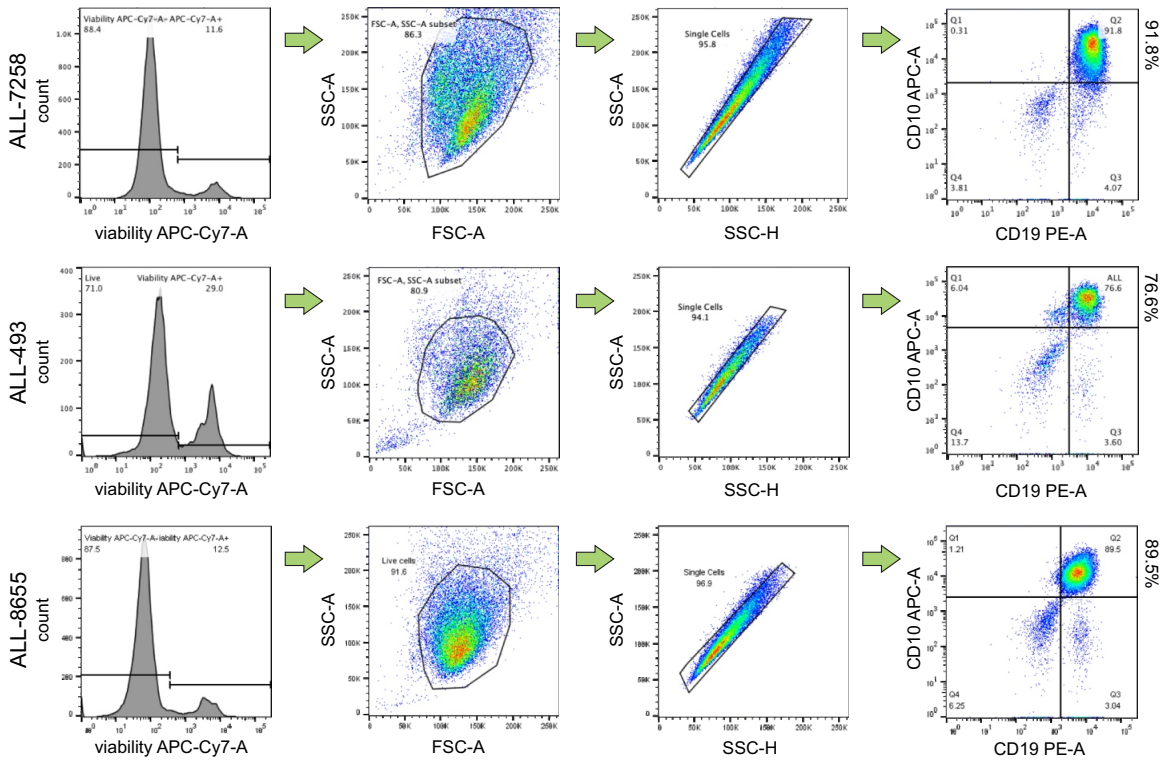
SUPPLEMENTARY FIGURE 13



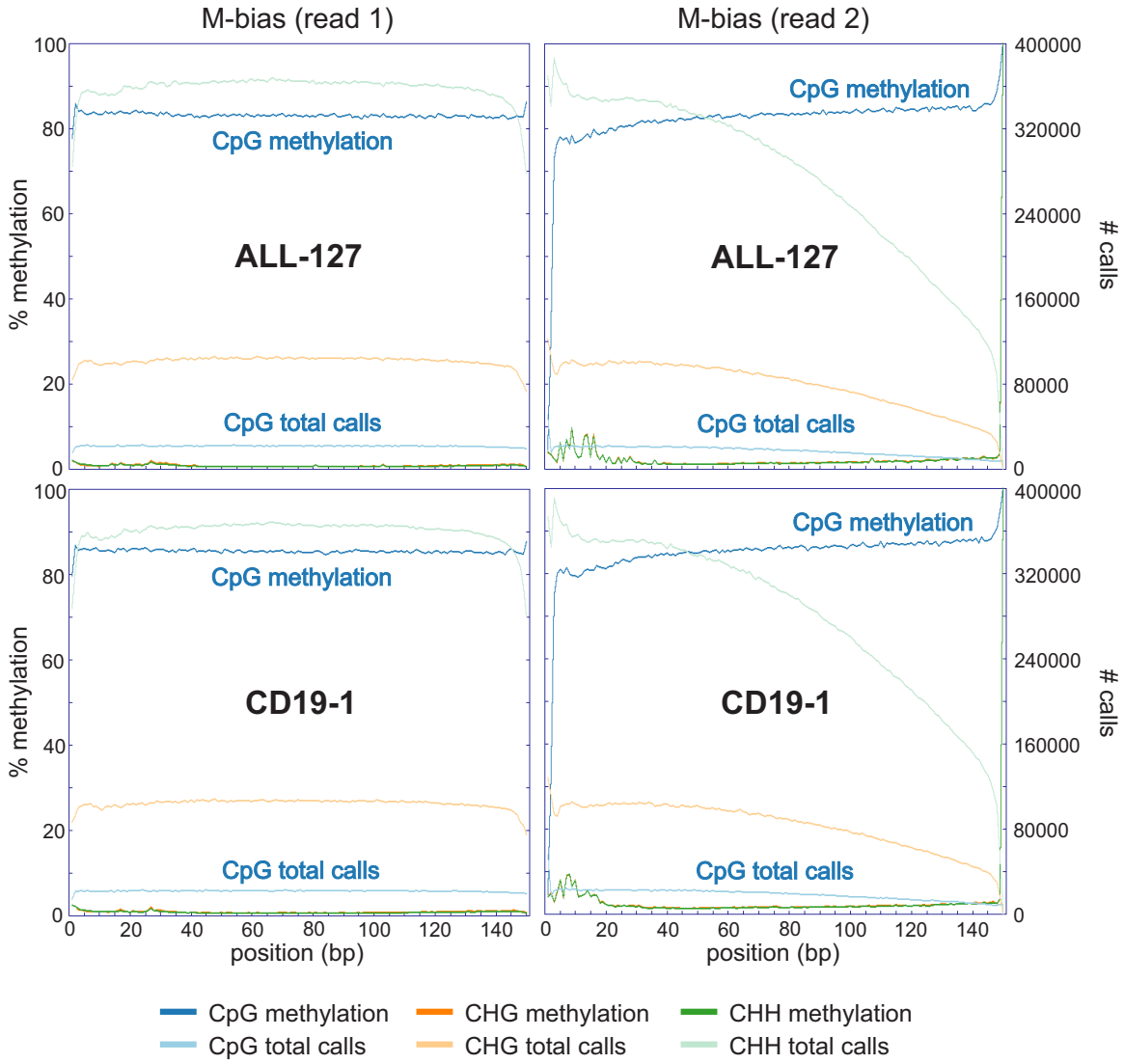
SUPPLEMENTARY FIGURE 14



SUPPLEMENTARY FIGURE 15

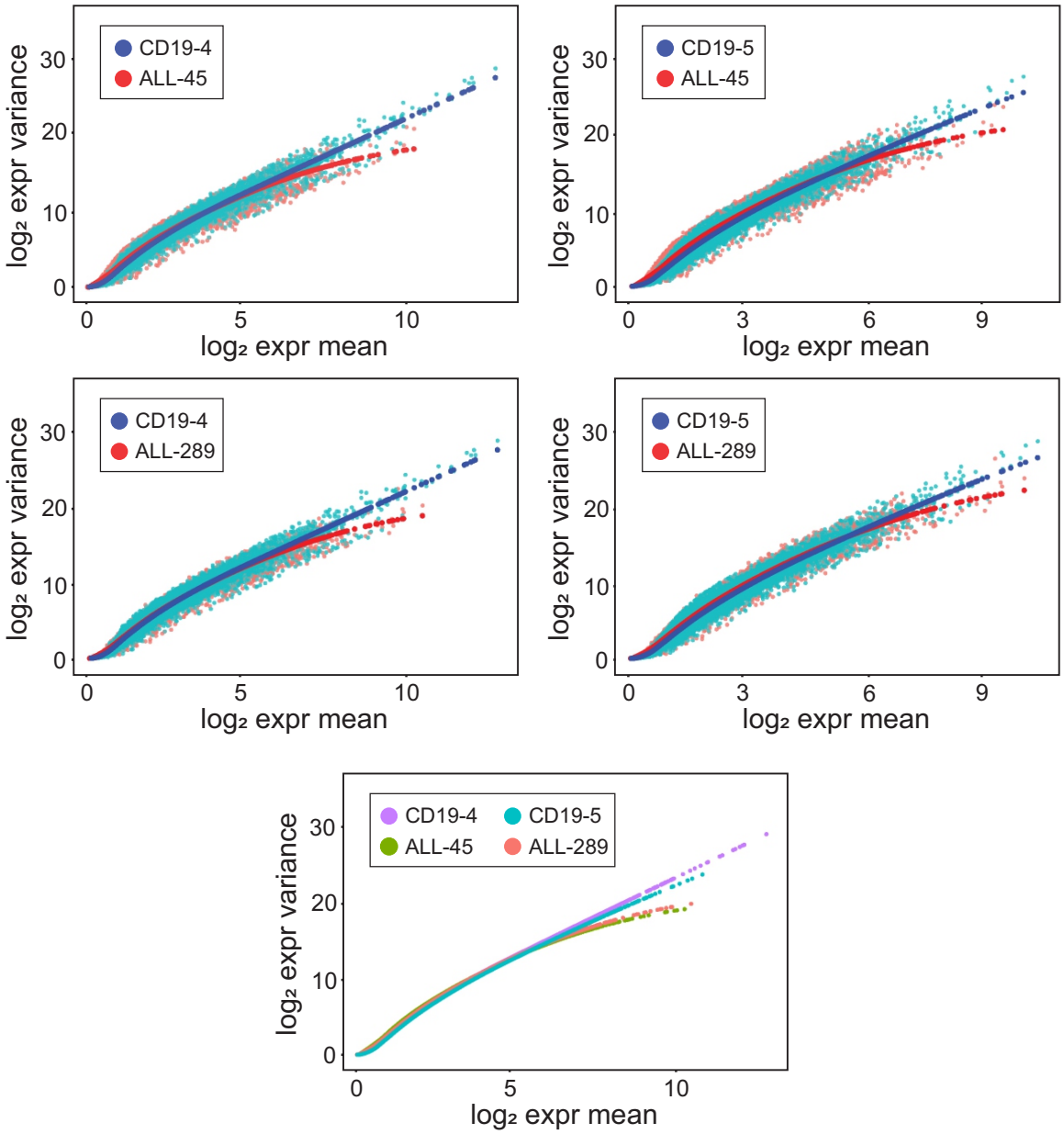


SUPPLEMENTARY FIGURE 16



SUPPLEMENTARY FIGURE 17

ETV6-RUNX1



SUPPLEMENTARY FIGURE 18

Supplementary Tables

Supplementary Table 1. WGBS samples, statistics, and clinical features.

Supplementary Table 2. Ranked list of genes using the average mutual information (MI) between the methylation level within their promoters and the phenotype in the ALL-45 vs CD19-1 comparison, computed as the square of the magnitude of the Jensen-Shannon distance (JSD), and associated gene set enrichment analysis (GSEA) results.

Supplementary Table 3. Genes (411) with significant (Q-value ≤ 0.05) Jensen-Shannon distance (JSD) magnitudes within their promoters in the ALL-45 vs CD19-1 comparison but not with absolute differential mean methylation levels (dMMLs), and associated GSEA results.

Supplementary Table 4. DMR-based odds ratio enrichment analysis in ETV6-RUNX1 ALL

Supplementry Table 5. Bulk RNA-seq results of eight ETV6-RUNX1 ALL and three normal CD19+ cell samples, including (a) gene expression values are given in log2(1+TPM) (TPM: transcripts per million) and are quantile normalized across samples and (b) Bulk RNA-seq differential analysis results. This table provides values for the moderated t-statistic and associated P-values and Q-values when performing differential expression analysis between the eight ETV6-RUNX1ALL and the three normal CD19+ samples. To account for multiple testing, Q-values were computed using the Benjamini-Hochberg procedure.

Supplementary Table 6. Single-cell RNA-seq data associated with two ETV6-RUNX1ALL and two normal CD19+ cell samples.

Supplementary Table 7. Differential mean expression behavior of DMGs and non-DMGs in the ALL-45 vs CD19-1 comparison.

Supplementary Table 8. Differential mean expression and differential variability behavior of bivalent genes and non-bivalent genes in the ALL/CD19 comparison.

Supplementary Table 9. DMR odds ratio enrichment analysis in all four cytogenetic subtypes of ALL.

Supplementary Table 10. Differentially methylated genes (DMGs) common to four cytogenetic subtypes of ALL in given samples. Column B provides

Supplementary Table 11. Ranked lists of genes using the average mutual information (MI) between the methylation level within their gene bodies and the phenotype in the ETV6-RUNX1/CD19-1 comparisons, computed as the square of the magnitude of the Jensen-Shannon distance (JSD).

Supplementary Table 12. Ranked list of genes using the average mutual information (MI) between the methylation level within their promoters and the phenotype in the Reh-UHRF1-KO vs Reh-NT comparison computed as the square of the magnitude of the Jensen-Shanon distance (JSD). The table also shows the average mean methylation level (MML) within the promoter of each gene as well as the average normalized methylation entropy (NME). Only genes for which the mutual information can be computed are listed.

Supplementary Table 13. Components of a UHRF1/translocation gene regulatory network, their differential expression, degree, and connectivity.

Supplementary Table 14. Gene regulatory network modules, also known as functional epigenetic modules (FEMs), characterized by significant discordance in mean gene expression and DNA methylation. Only FEMs comprised of at least 10 genes and FDR Q-values no higher than 0.05 are reported.

Supplementary Table 15. Median Jensen-Shannon distance (JSD) of comparisons between all ETV6-RUNX1and control samples within ChromHMM annotations.