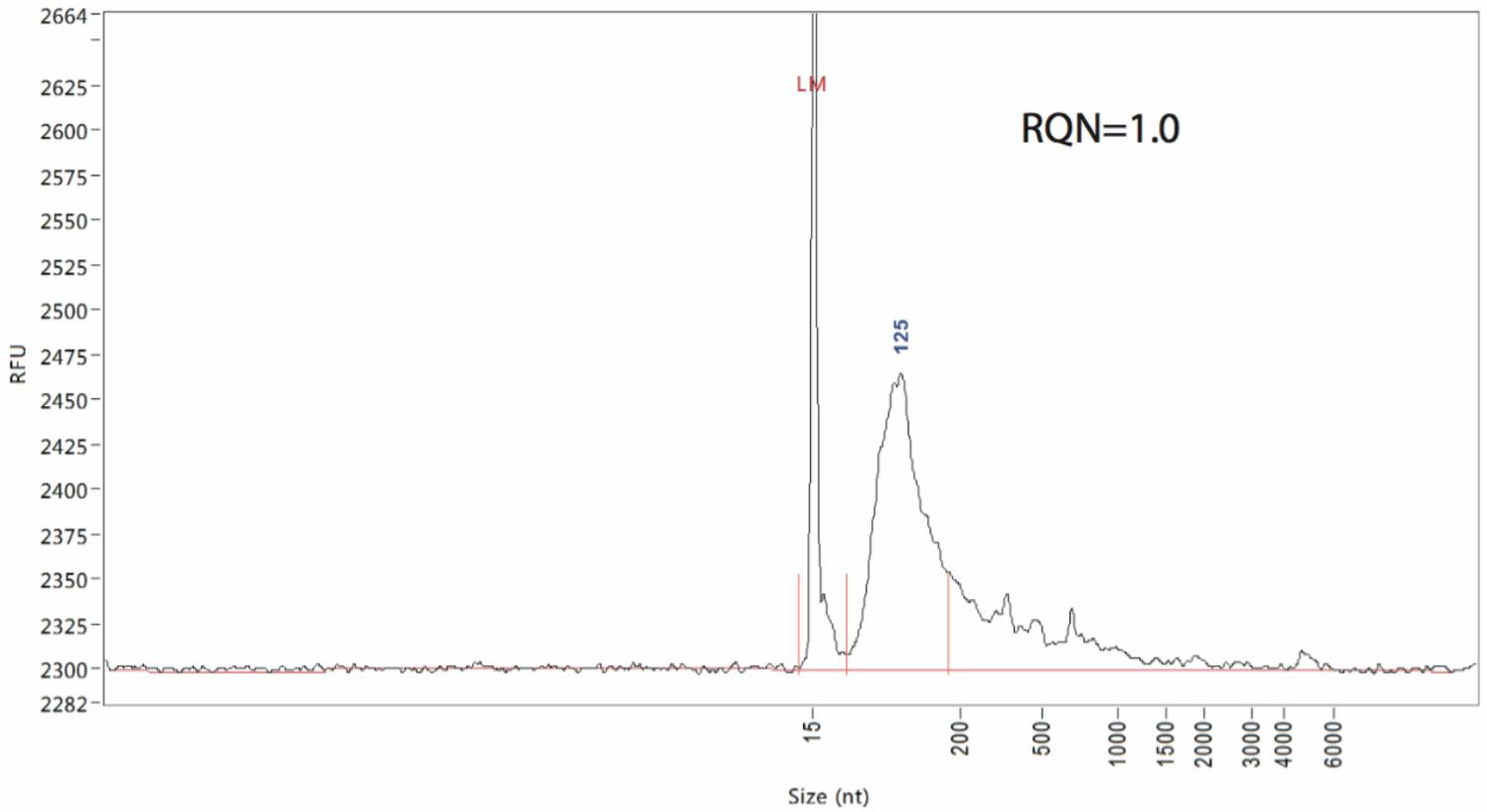
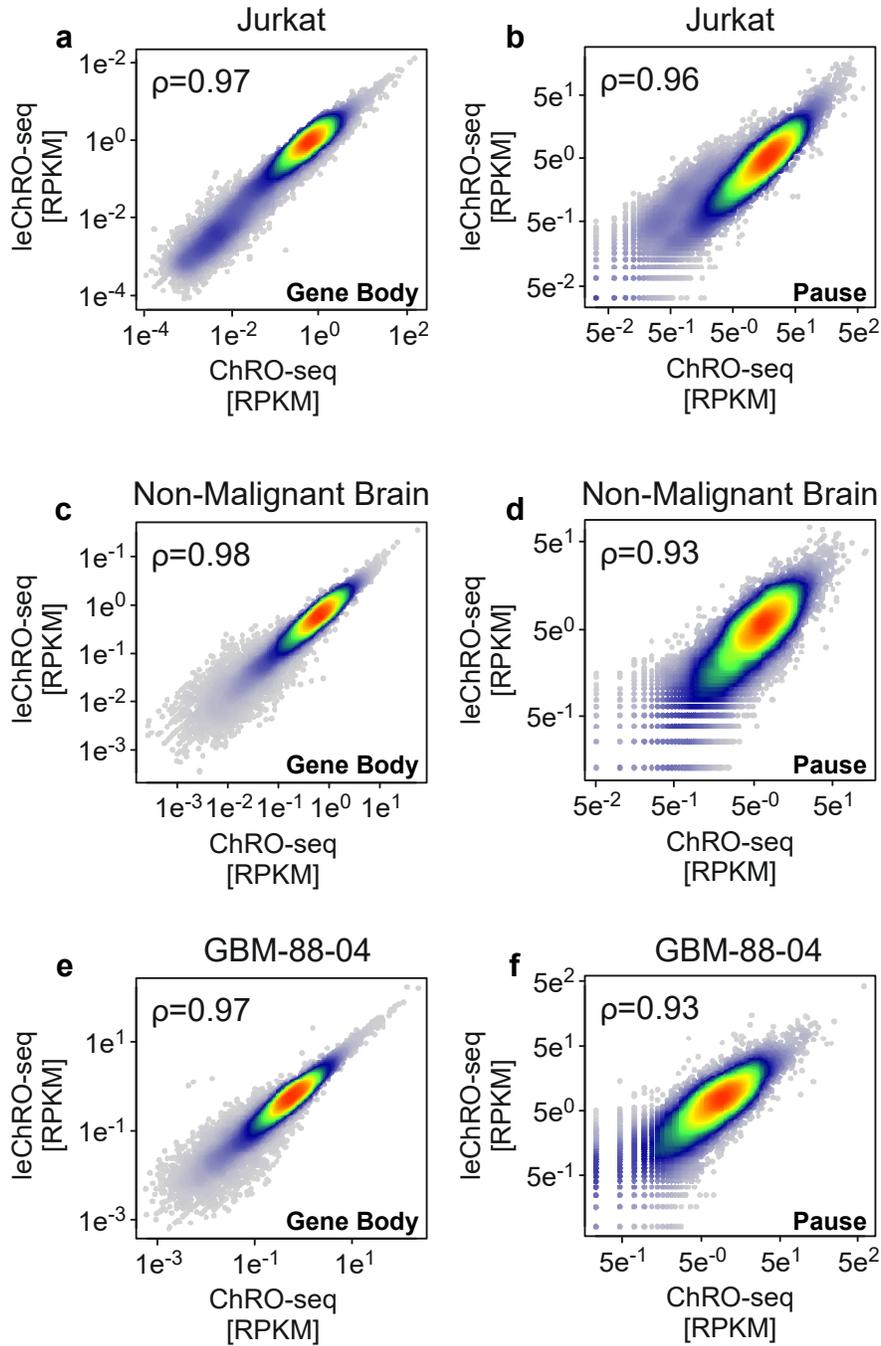


**Fig.S1. ChRO-seq in 11 canine tissue samples.** ChRO-seq libraries were prepared from canine samples isolated from fetal gonadal tissues from female (XX) or male (XY) embryos, fetal liver, or adult mitral valves. Samples were clustered using complete-link clustering based on Spearman's rank correlation.

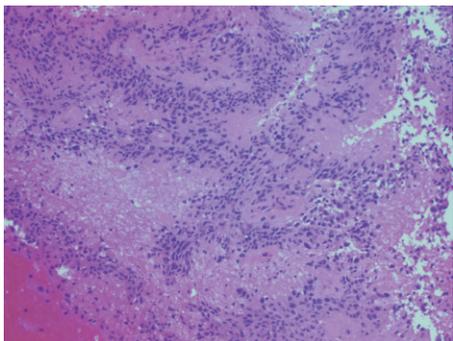


**Fig.S2. Bioanalyzer analysis of RNA isolated from GBM-2.** The plot reported by the Bioanalyzer software shows the size of RNA isolated from GBM-2 in units of nucleotides (nt, X-axis) as a function of the relative fluorescence units (RFU, Y-axis). RNA Quality Number (RQN = 1) shown in the trace denotes extensive RNA degradation. The mode of the distribution of RNA sizes is shown (125 nt).

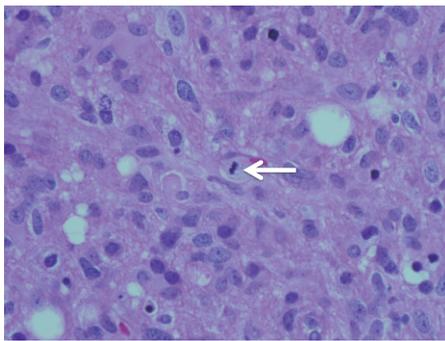


**Fig.S3. Correlation between ChRO-seq and leChRO-seq.** (a-f) Scatterplots show the density of reads mapping in the gene bodies (+1000 to gene end) (a, c, e) or in the promoter proximal pause near the transcription start site (b, d, f) of RefSeq genes. All axes are in units of reads per kilobase per million mapped (RPKM). Spearman's rank correlation ( $\rho$ ) is shown in each plot. The color scale denotes the density of points.

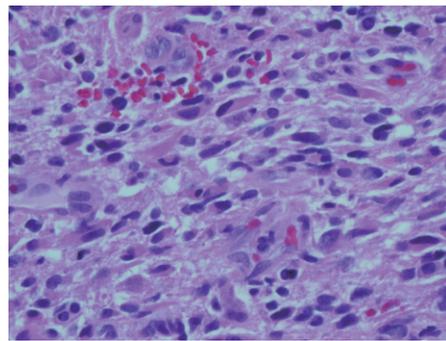
A



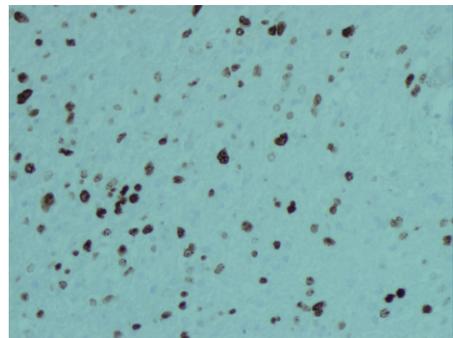
B



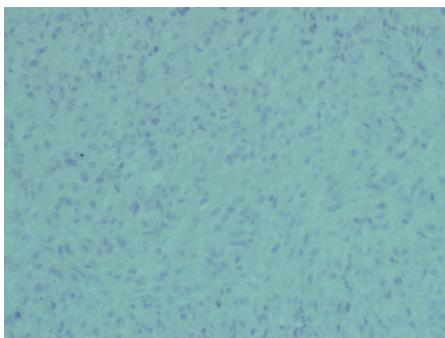
C



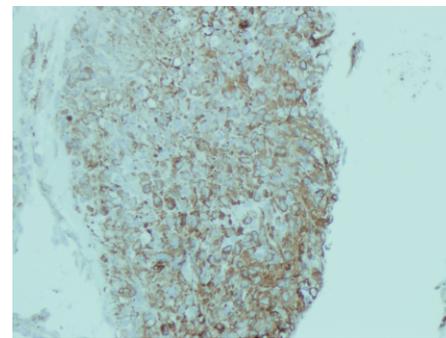
D



E



F

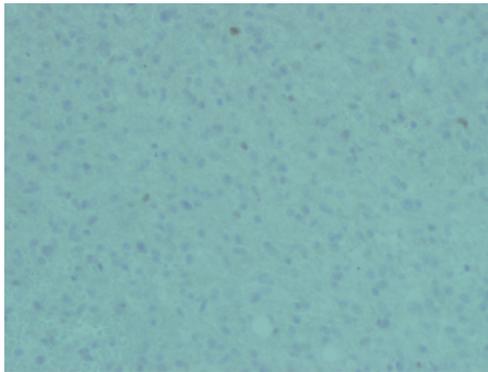
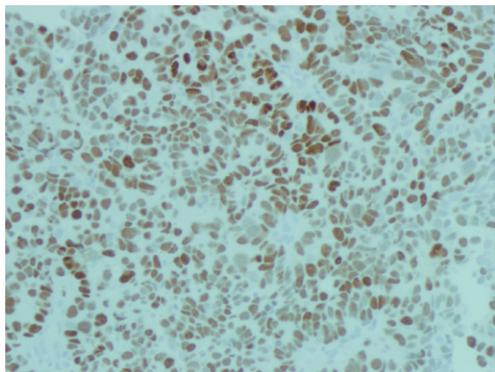


G

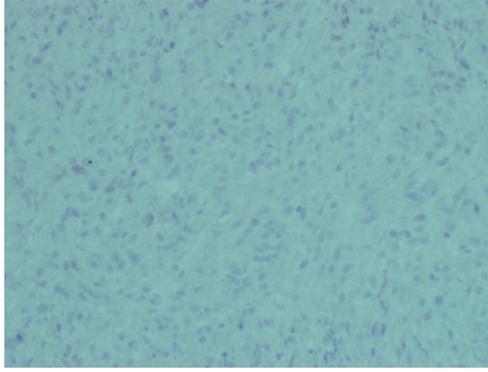
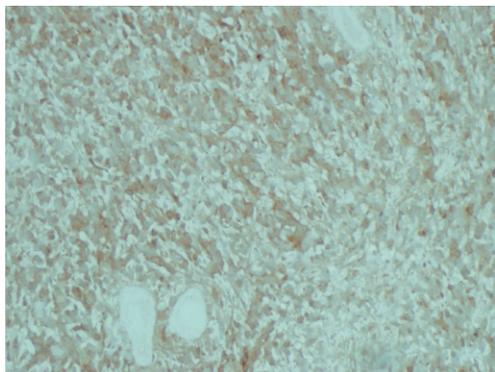
control

tumor

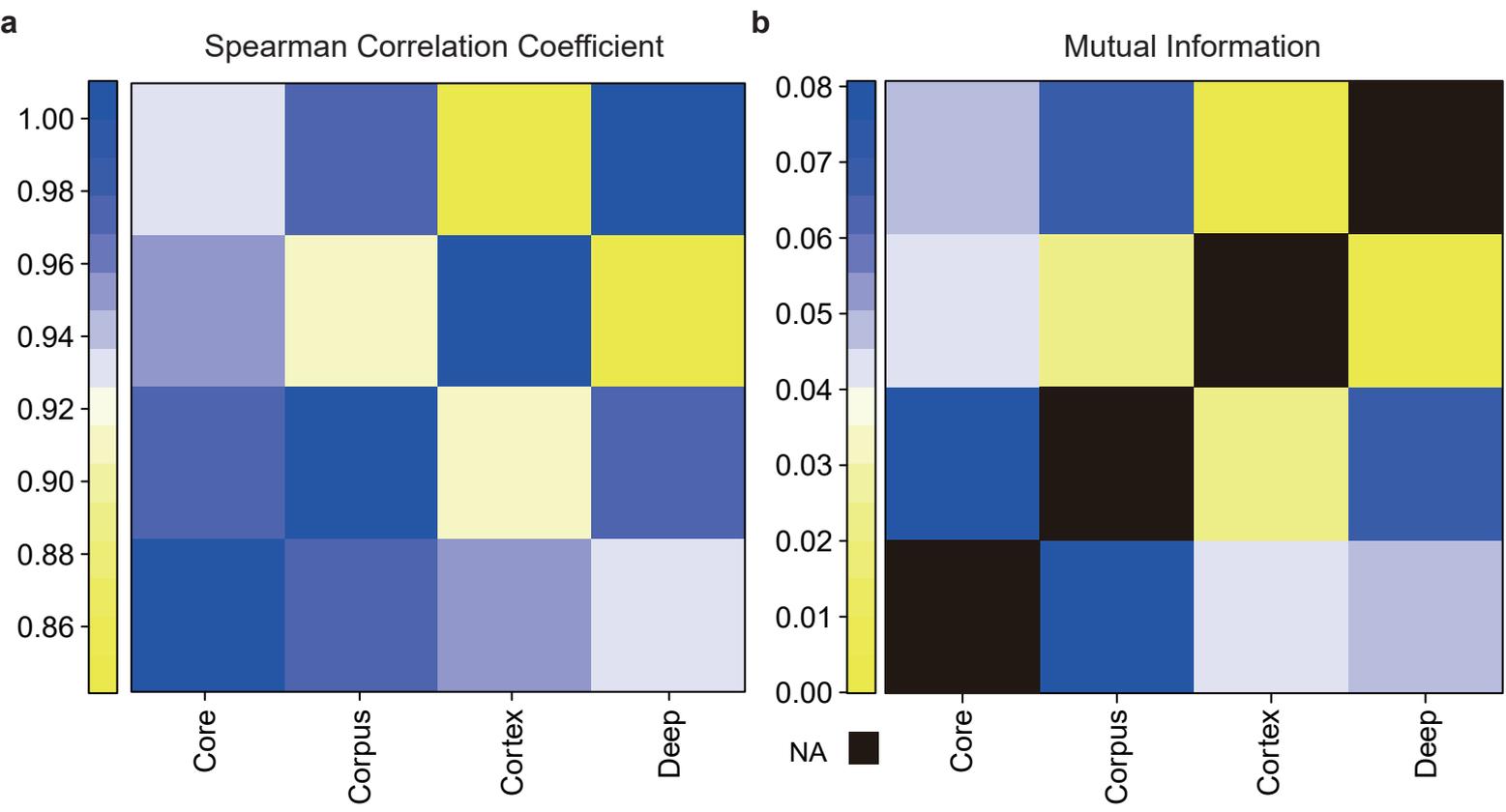
p53



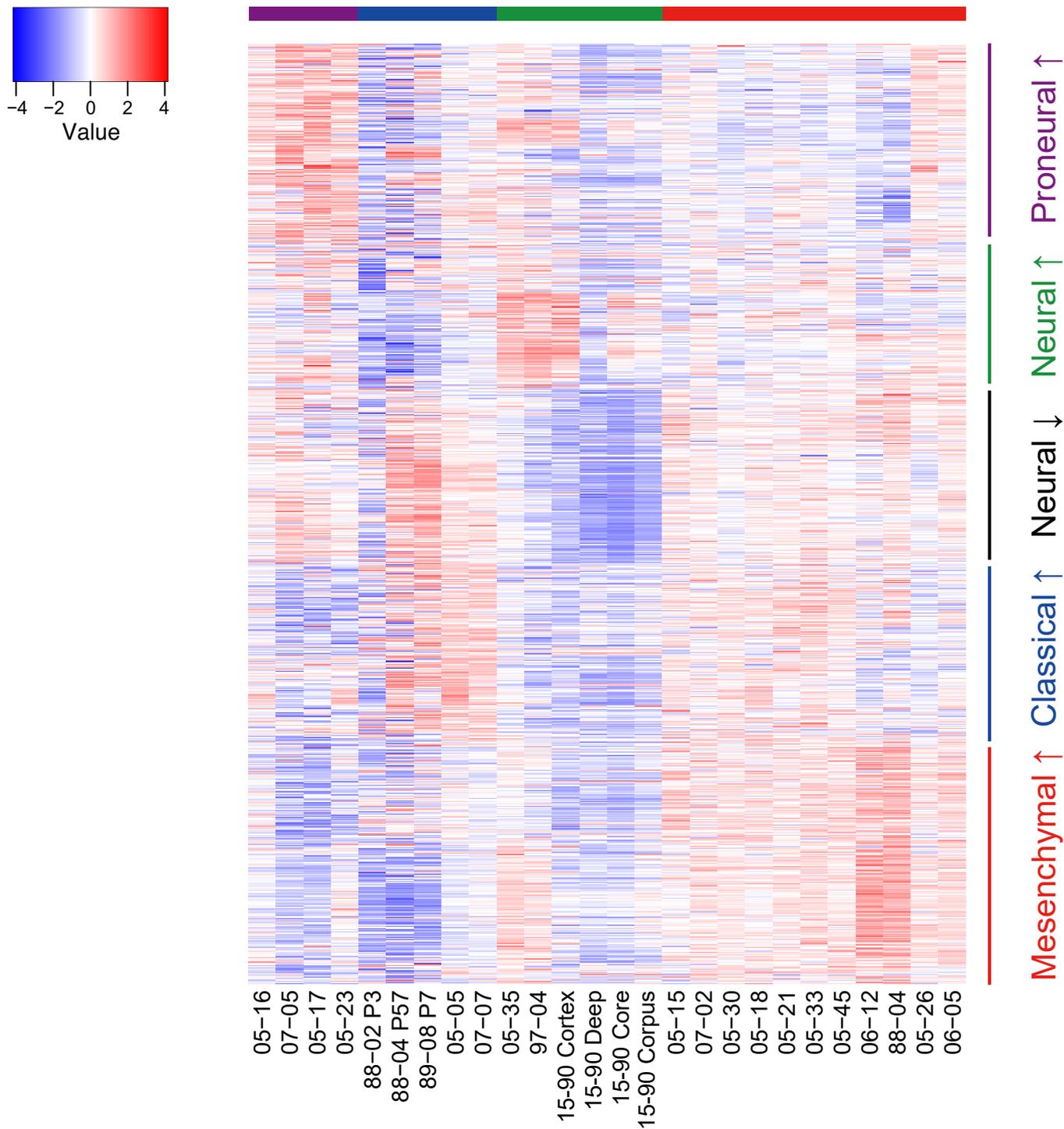
IDH1



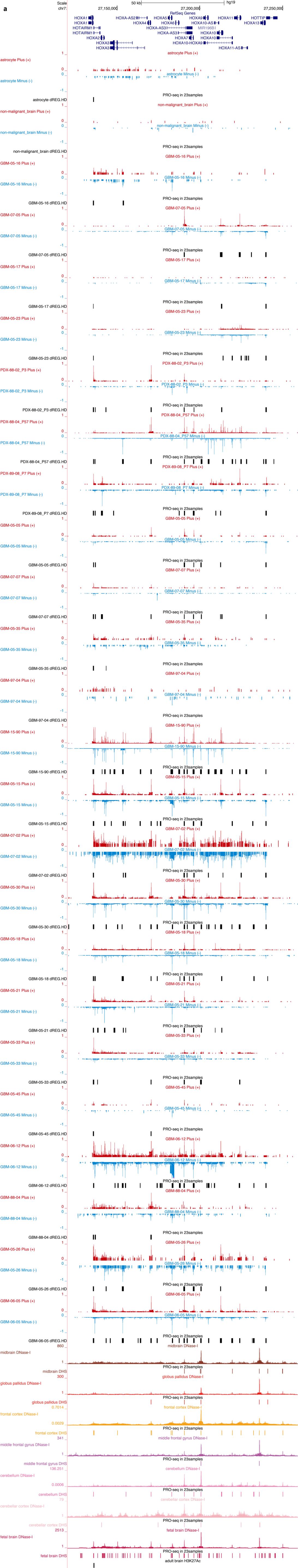
**Fig.S4 Brain biopsies display immunohistochemical markers of high grade glioma in GBM-1.** (A) Pseudopalisading borders with necrotic centers. (B) Actively mitotic cells (arrow). (C) Neovascularization. (D) Active proliferation stained by ki67+. (E) IDH1 staining is negative. (F) GFP is stained as positive. (G) Additional markers of high grade glioma between the tumor include p53<sup>-/-</sup> and IDH<sup>-/-</sup> using an IDH-1 positive glioblastoma as a positive control.

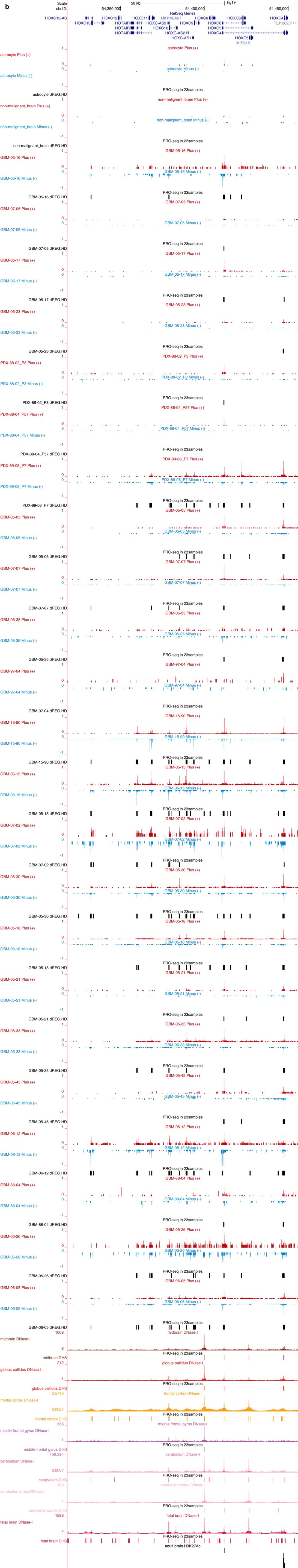


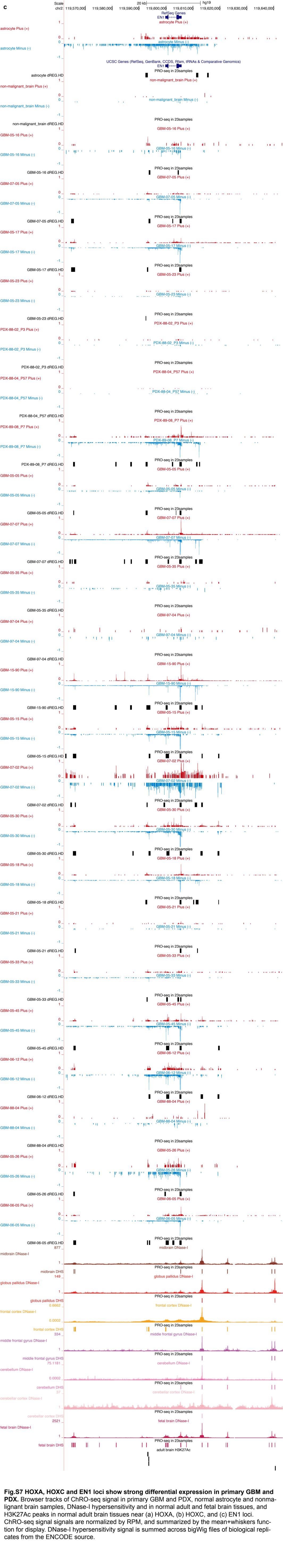
**Fig.S5 Comparison between four regions dissected from GBM-15-90. (a)** Spearman's rank correlation between RPM normalized transcription levels in gene bodies. **(b)** Mutual information between TREs discovered independently in the four tumor regions.



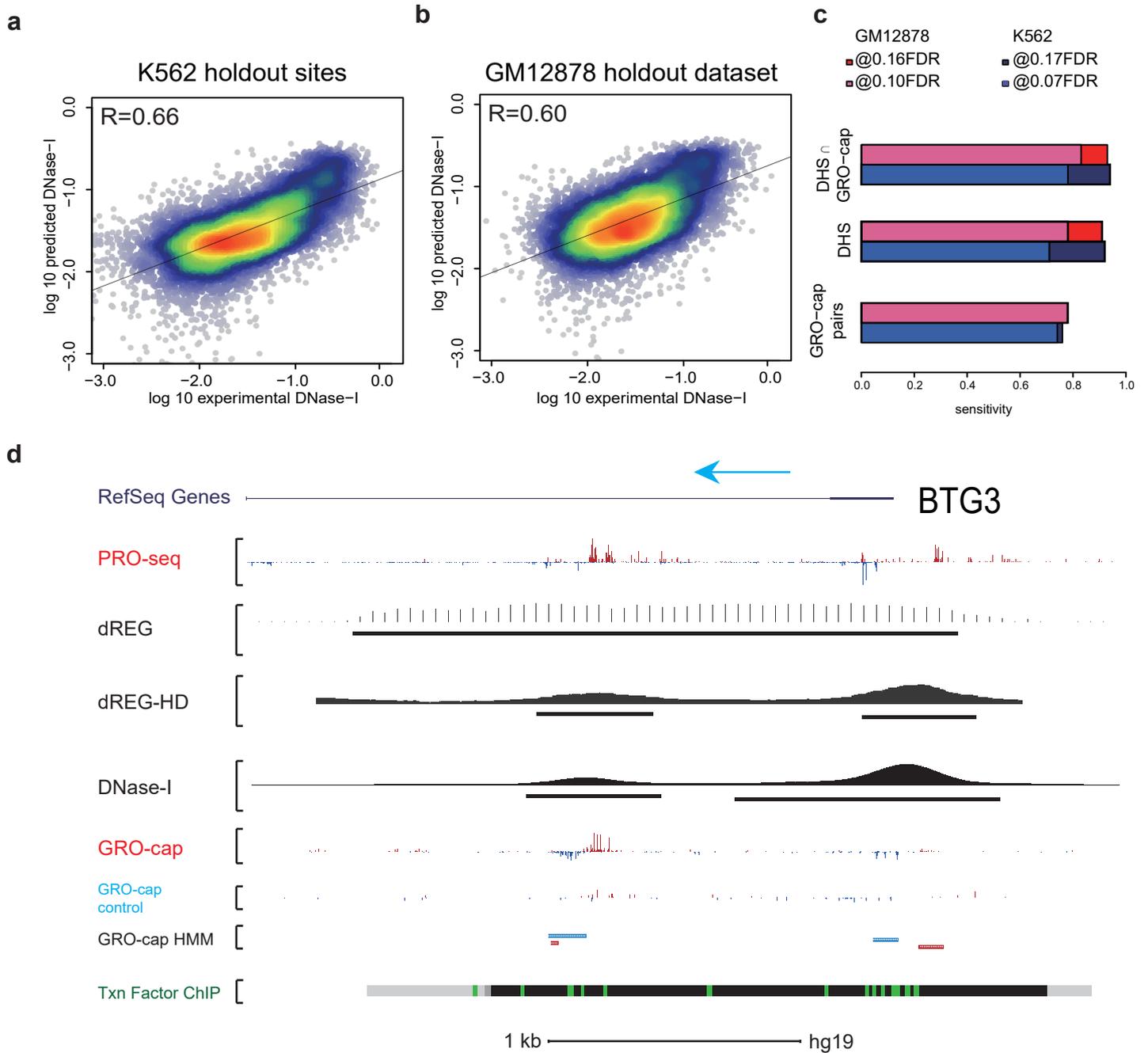
**Fig.S6 Expression of molecular subtype predictor genes in primary GBM / PDX samples.** Heatmap shows the expression of predictor genes of four molecular subtypes of glioblastoma. Red colors indicate higher transcription activity and blue colors indicate lower activity. Samples are ordered based on their subtype.



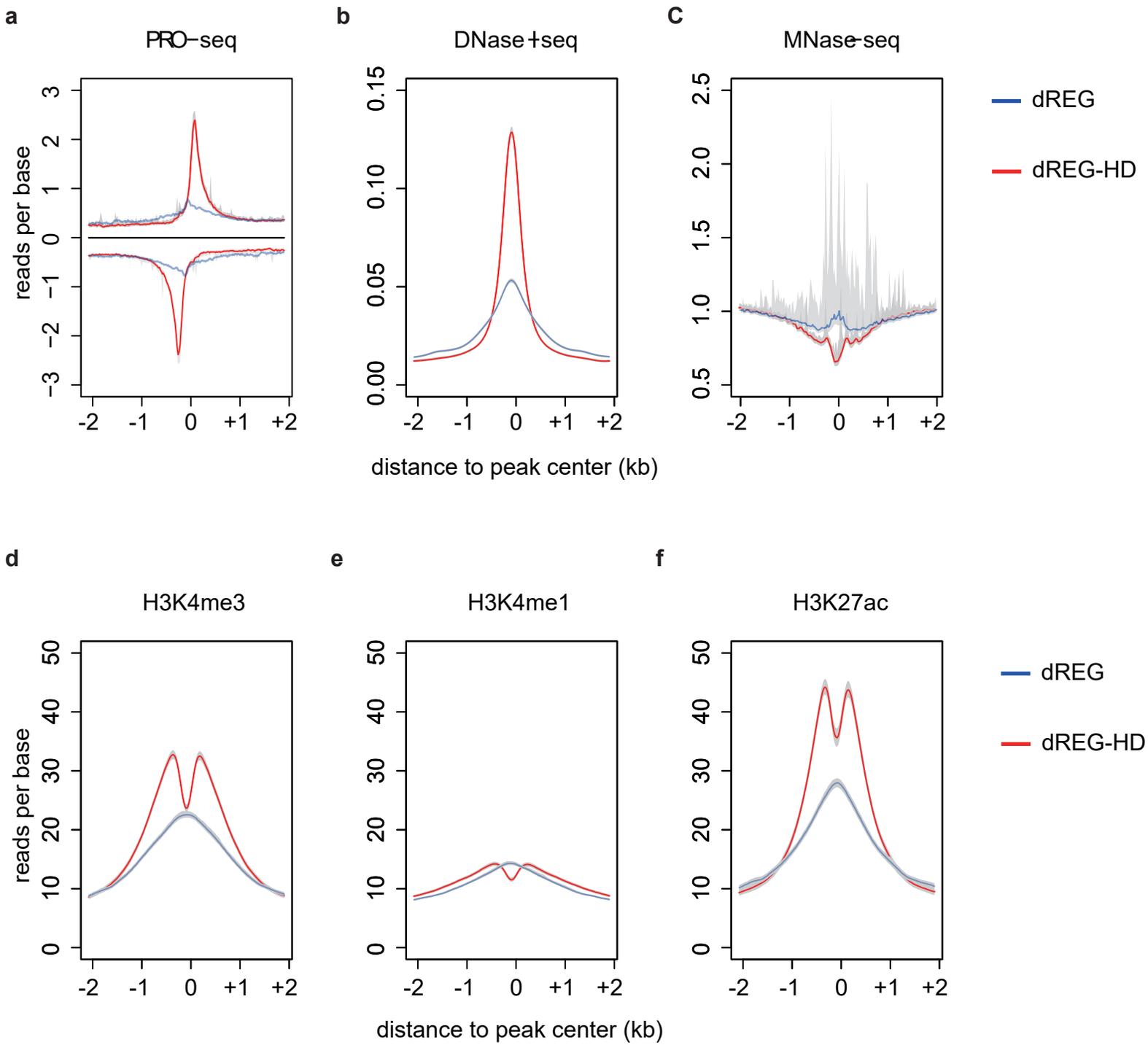




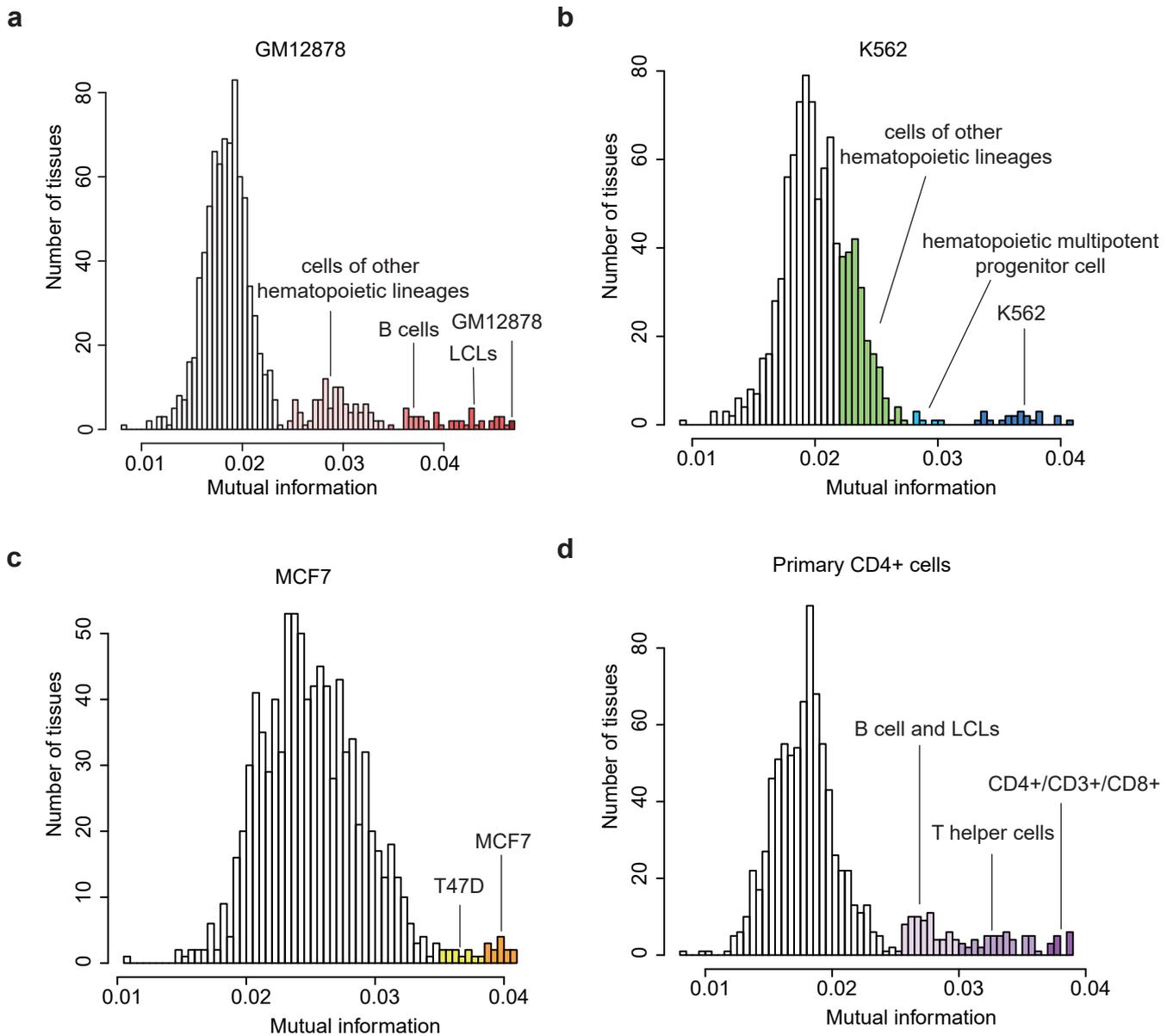
**Fig.S7 HOXA, HOXC and EN1 loci show strong differential expression in primary GBM and PDX.** Browser tracks of ChRO-seq signal in primary GBM and PDX, normal astrocyte and non-malignant brain samples, DNase-I hypersensitivity and in normal adult and fetal brain tissues, and H3K27Ac peaks in normal adult brain tissues near (a) HOXA, (b) HOXC, and (c) EN1 loci. ChRO-seq signal signals are normalized by RPM, and summarized by the mean+whiskers function for display. DNase-I hypersensitivity signal is summed across bigWig files of biological replicates from the ENCODE source.



**Fig.S8 dREG-HD refines TRE predictions by imputing DNase-I hypersensitivity.** (a and b) Scatter density plot between predicted and experimental DNase-I hypersensitivity signals in K562 holdout sites that were not used during training (a) and a complete holdout dataset in GM12878 (b). DNase-I hypersensitivity signals are summed for non-overlapping 80bp intervals and plotted on log<sub>10</sub> scale. (c) Sensitivity of dREG-HD to detect DNase-I hypersensitivity sites (DHSs) that intersect dREG regions, paired GRO-cap HMM peaks, and the intersection of DHSs and GRO-cap pairs (See methods). Prediction in K562 and GM12878 are colored in blue and red respectively. The sensitivity analyzed under 'relaxed' dREG-HD setting was colored in dark red/blue, and those under 'stringent' setting were colored in light red/blue. The expected false discovery rate of the 'relaxed' and 'stringent' settings are indicated above the barplot. (d) Browser track of a region near the transcription start site of BTG3 in K562 cells. From top to bottom tracks represent: 1) RefSeq genes showing the transcription start site of BTG3; 2) PRO-seq colored in red for forward strand and blue for the reverse strand; 3) dREG scores and the peak; 4) dREG-HD scores and peaks; 5) DNase-I hypersensitivity signal and peaks; 5) GRO-cap reads. 6) The no-TAP control experiment matched to GRO-cap signal; 6) Transcription start sites identified using the GRO-cap signal; 7) Potential transcription factor binding detected by ChIP-seq. Peak calls are colored in gray and black and the best match to a transcription factor binding motif is colored in green.

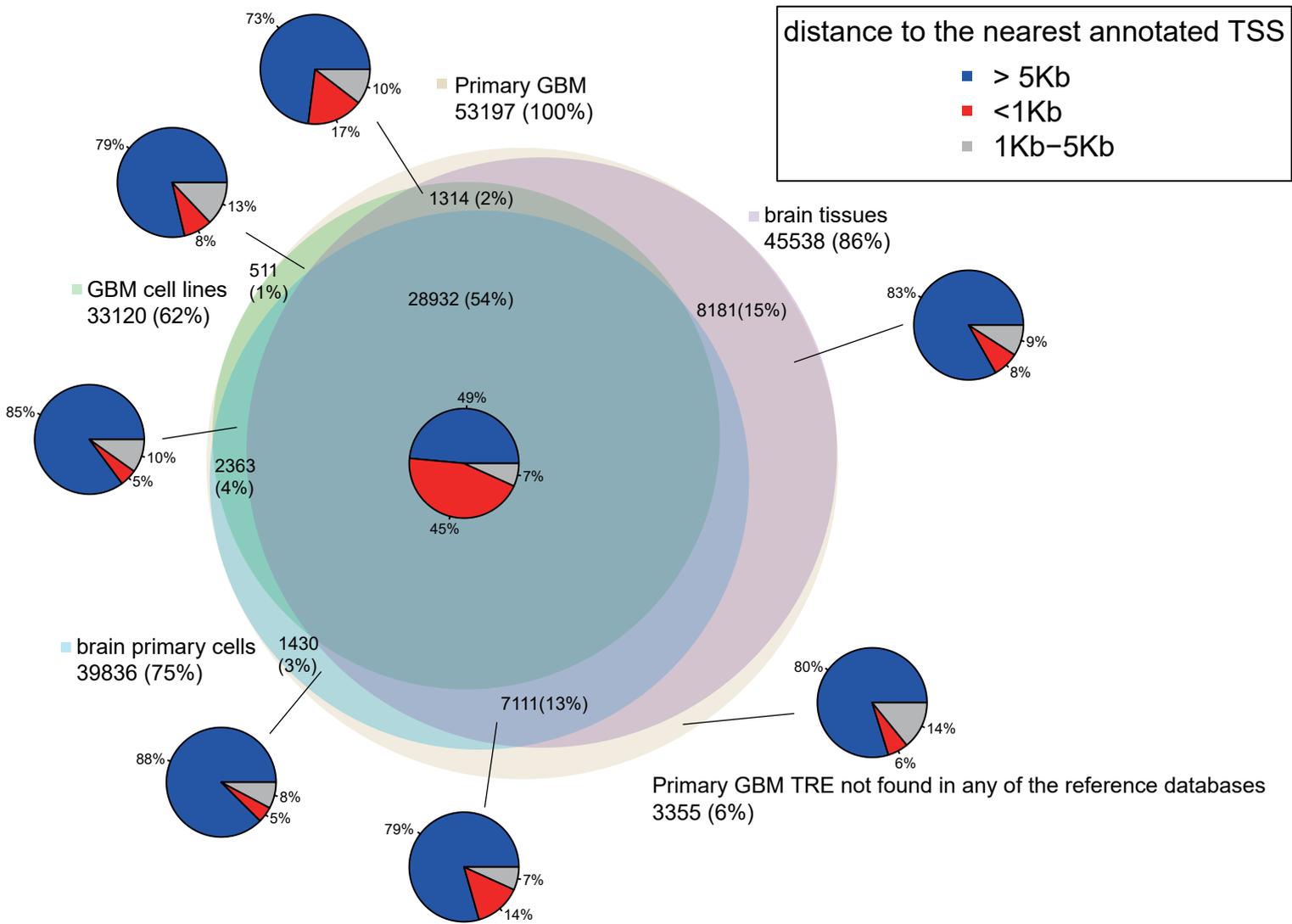


**Fig.S19 Metaplots for PRO-seq, chromosomal accessibility, and histone modifications that marks active TREs.** Signals of the indicated mark over dREG and dREG-HD regions are shown in blue and red, respectively. Shadows marks the 25 and 75 percentiles of 1000 samples of 10% of the data (see methods).

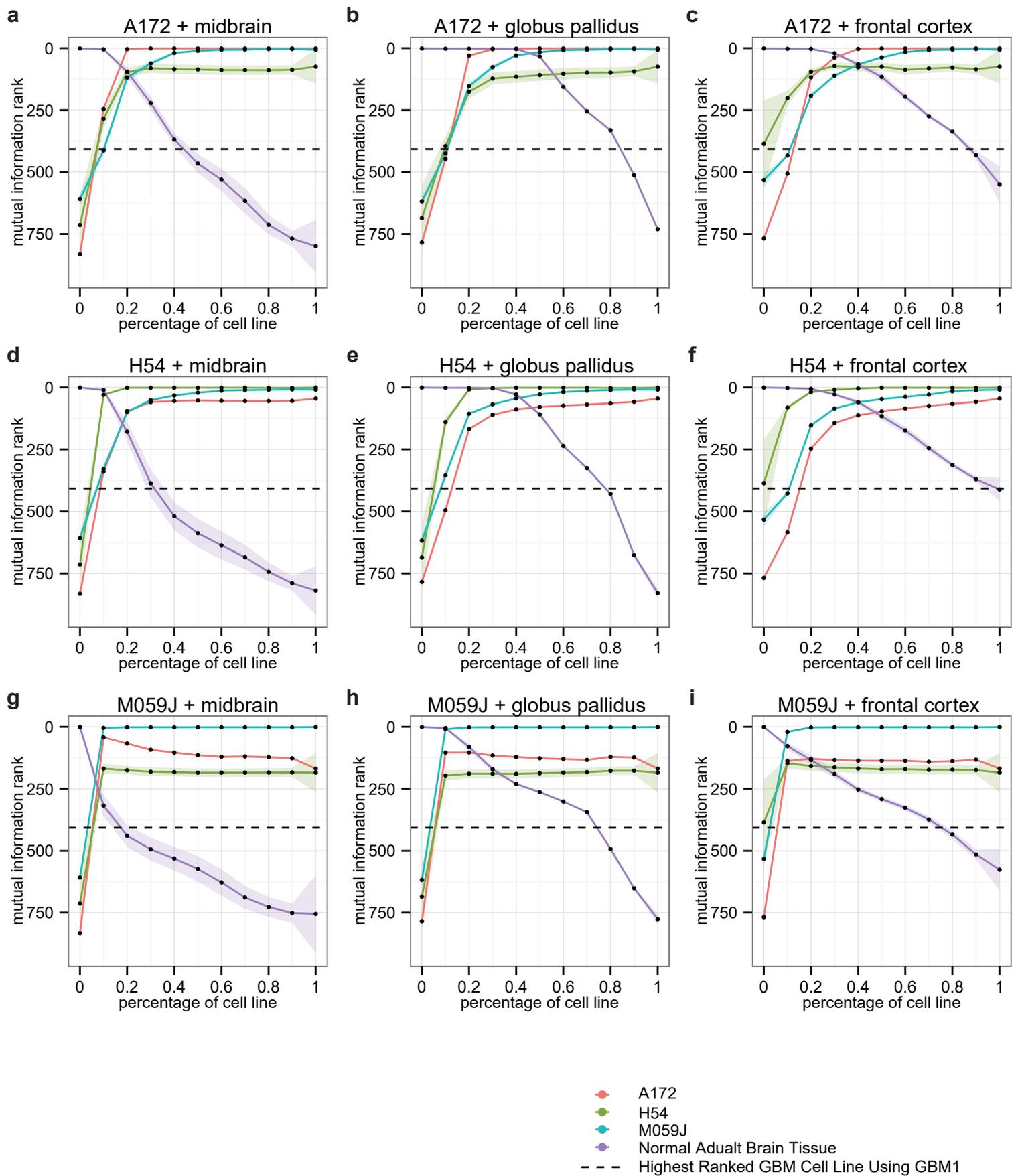


**Fig.S10 Mutual information is an accurate similarity measure for TREs.** Mutual information (y-axis) between dREG-HD sites identified using PRO-seq or GRO-seq data in the indicated sample (**a**:K562, **b**:GM12878, **c**:MCF-7, **d**:human primary CD4+ T-cells) and DNase-I hypersensitive sites from reference samples. Samples are plotted by rank order (x-axis, from the lowest to the highest mutual information). Only 30-40 samples with the highest mutual information are included in the plot. In all cases, mutual information selects the sample that was most similar in the reference DHS data, including those of the same or similar cell types, are highlighted.

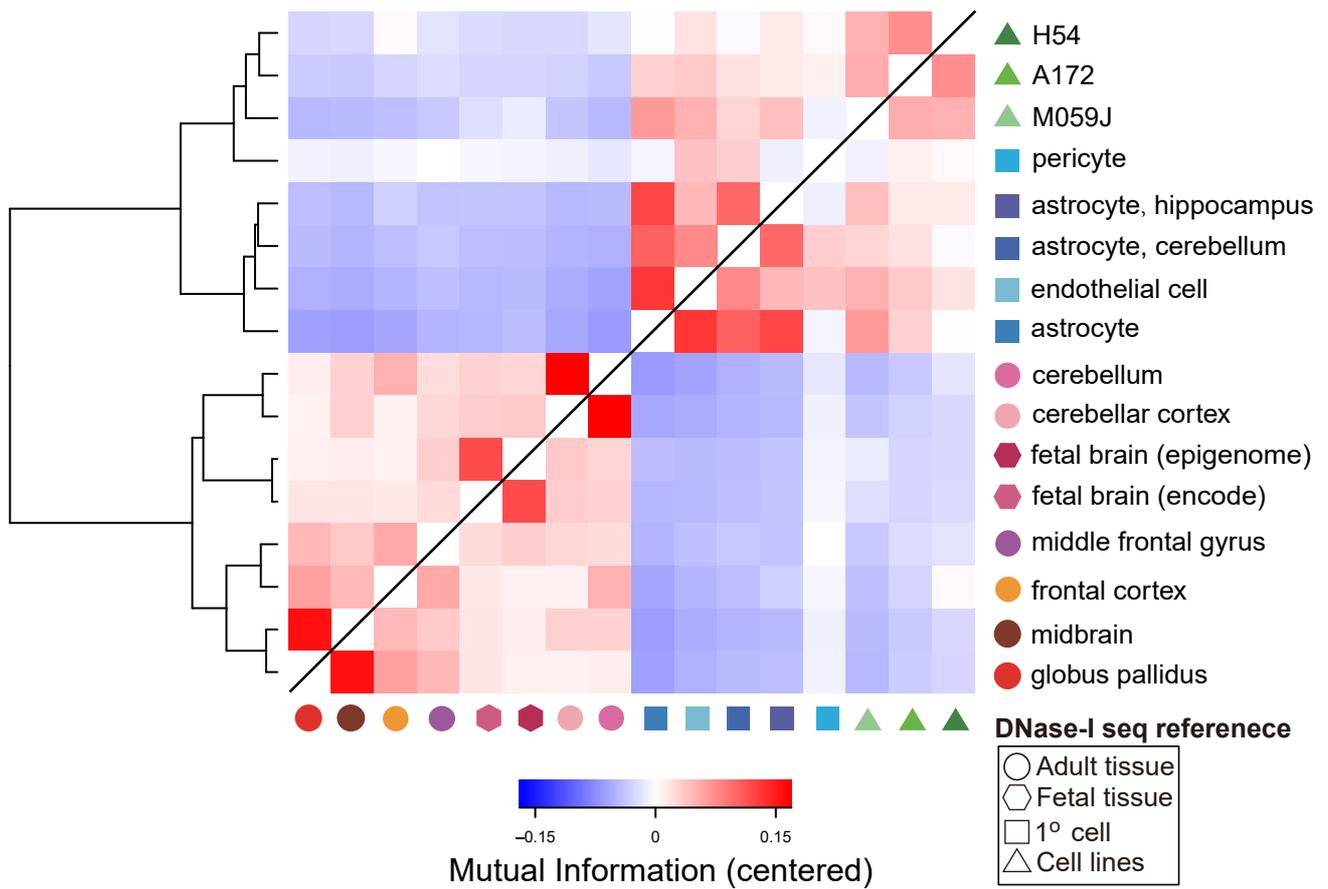




**Fig.S12 Venn diagram showing similarity in TREs between primary GBMs, normal brain tissue, and primary brain cells grown in tissue culture.** Venn diagram denotes the overlap between TREs found in GBM-15-90 and normal brain (pink), GBM cell line models (green), or primary brain cells that were dissociated from normal brain tissue and grown in culture for a limited number of passages (teal). For each overlap, the number and fraction of TREs is shown. Pie charts denote the fraction of TREs that are >5kb from the nearest annotated transcription start site (blue), <1kb (red) or between 1kb-5kb (gray).

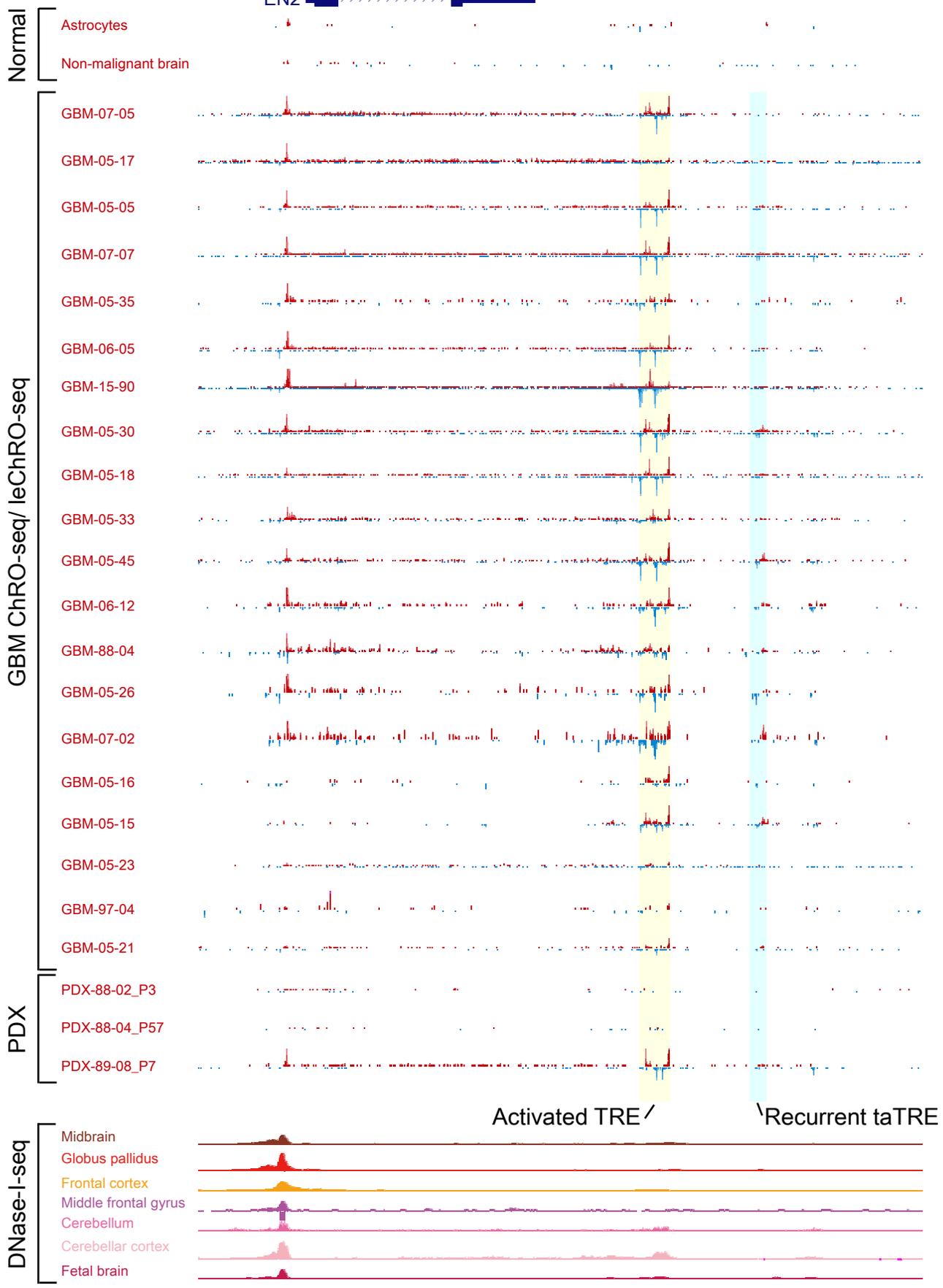


**Fig.S13 Mutual information in simulated mixtures containing normal brain tissue and GBM cell lines.** (a-i) Each panel shows a complex mixture of the normal brain tissue and GBM cell line indicated above the plot. The x-axis indicates the proportion of the cell line in the simulated data. The y-axis shows the rank of mutual information (from highest to the lowest) for three GBM cell lines and normal brain tissue. In cases where biological replicates are available the line denotes the mean mutual information among replicates and shadows mark the range. The dashed line shows the highest ranked GBM cell line in GBM-15-90 (H54).



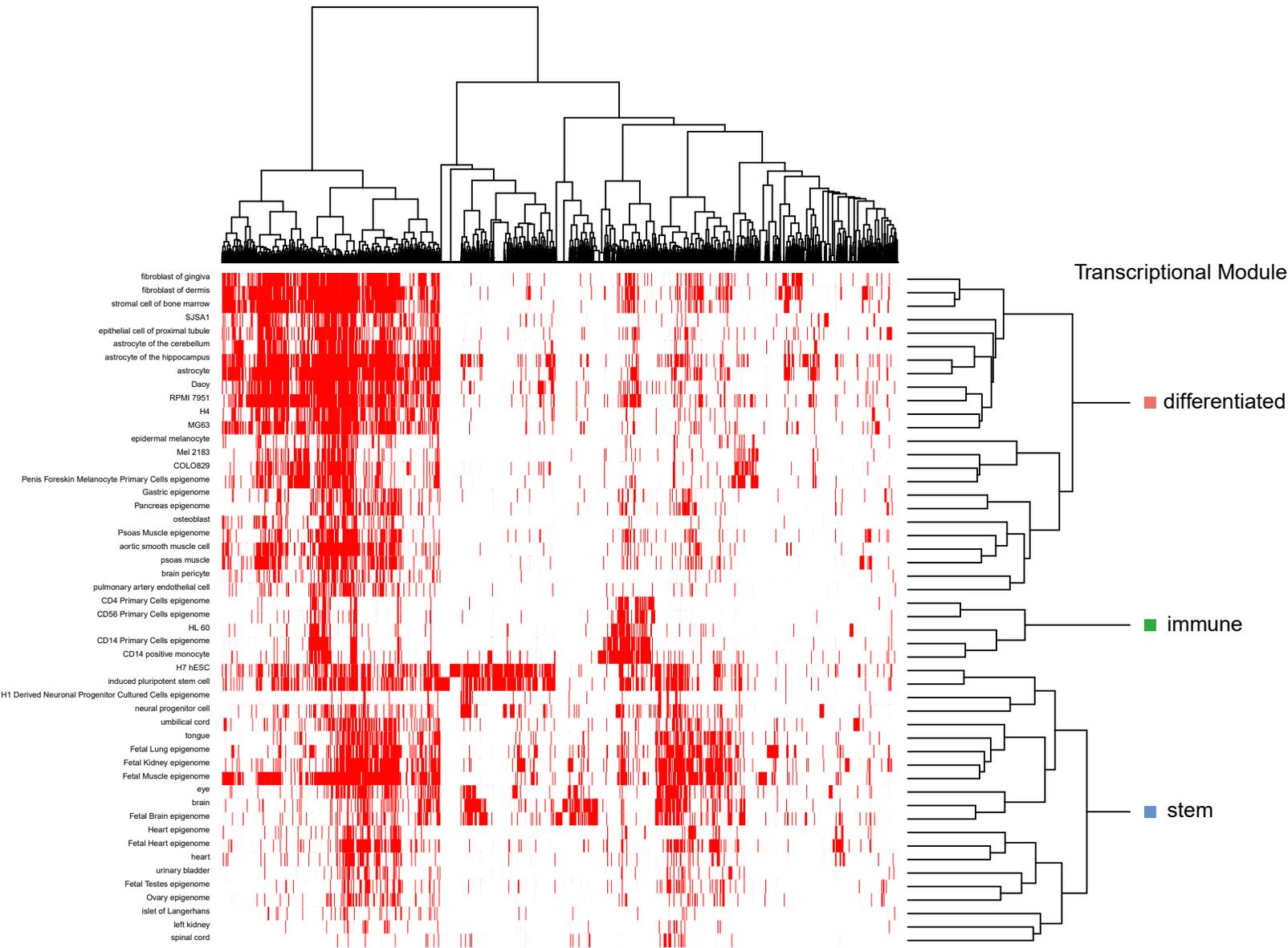
**Fig.S14** Pairwise mutual information among TREs from reference DHSs centered by the mean of each sample.

10 kb



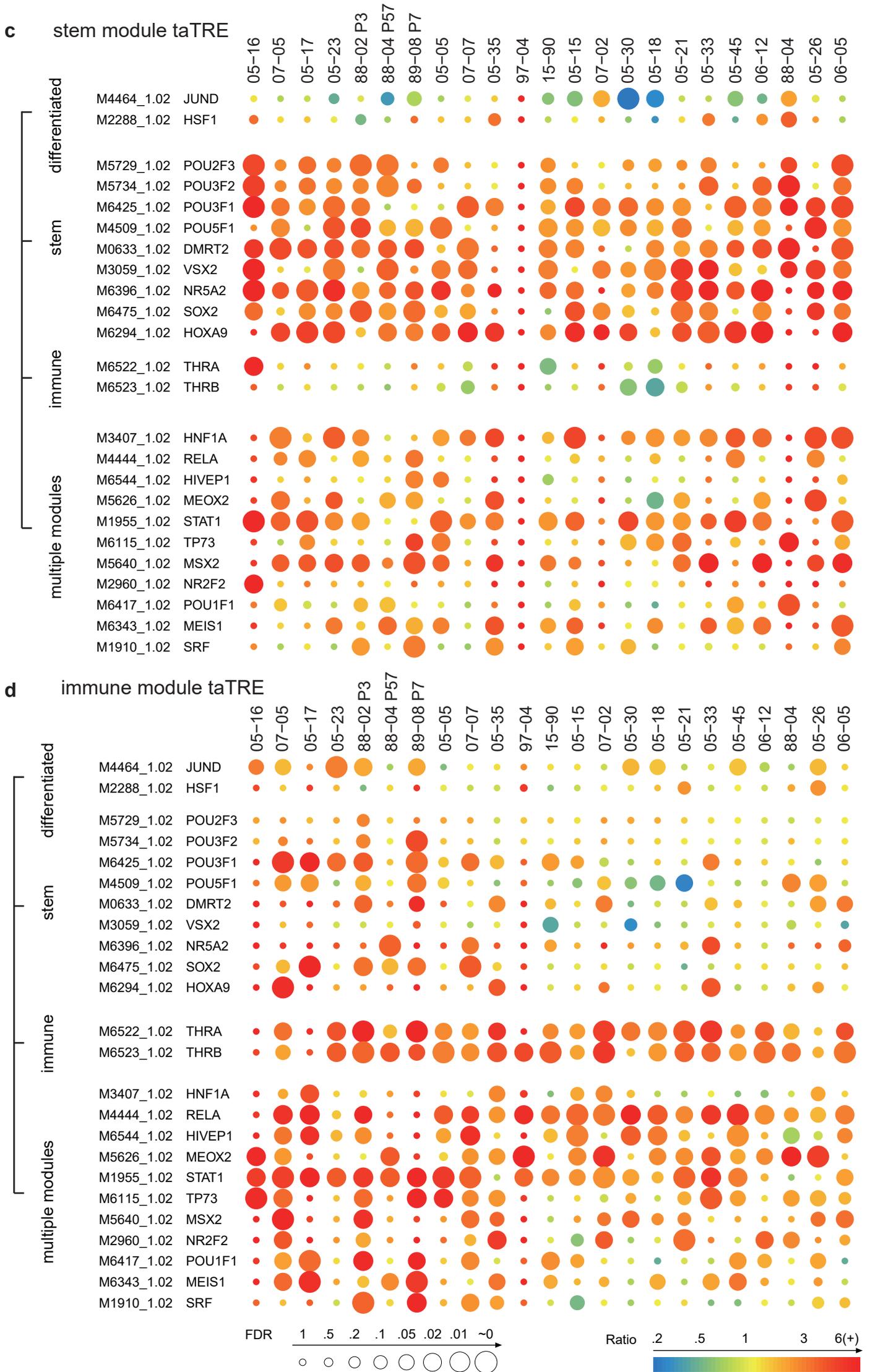
Activated TRE / \ Recurrent taTRE

Adult brain H3K27Ac

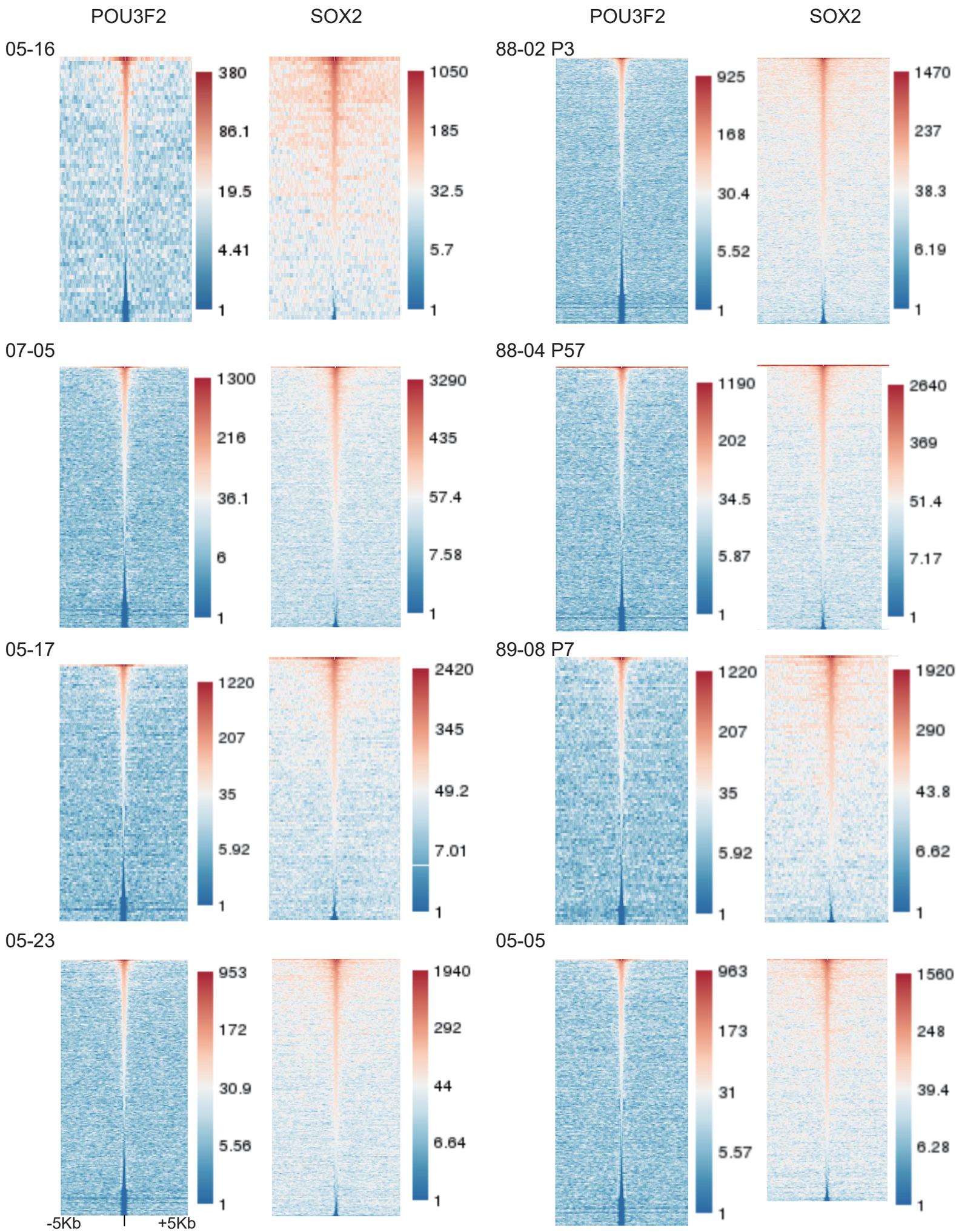


**Fig.S16 Clustering of taTREs-enriched reference samples.** Clustering of reference samples enriched for taTREs based on the activation of TRE. Activate TREs are marked in red; inactive ones are in white. Row dendrograms are cut down to three trees, each corresponding to the indicated transcriptional regulatory program (i.e., stem- or fetal-like, immune, and differentiated).





**Fig.S17 Transcription factor binding motifs enriched in TREs in the indicated regulatory program compared with normal brain.** Motif enrichment is analyzed on all taTREs (a), and regulatory program-specific taTREs (b, c, and d) for each GBM sample. Motifs are ordered by the four categories in Fig.4c. The false discovery rate is represented by the radius of the circle and enrichment (red) or depletion (blue) are represented by the color.



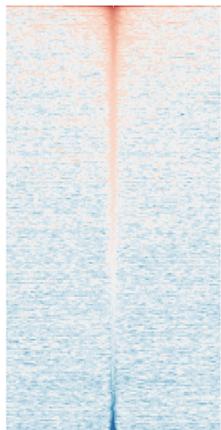
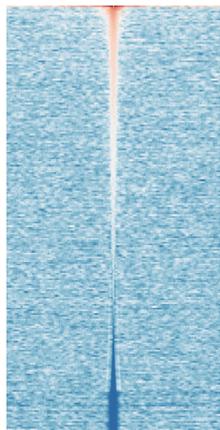
POU3F2

SOX2

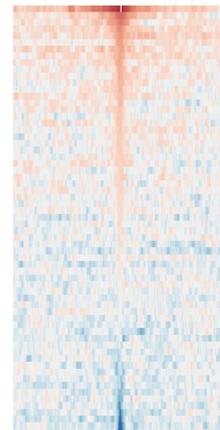
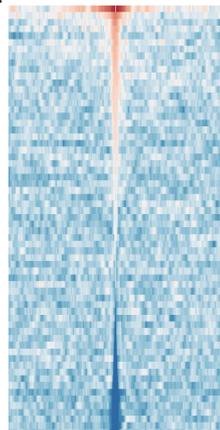
POU3F2

SOX2

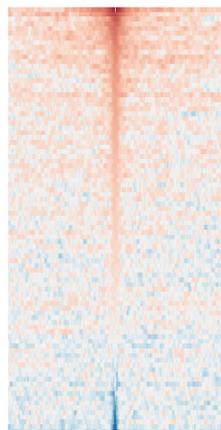
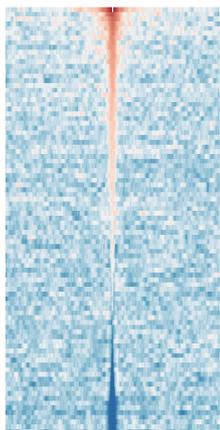
05-21



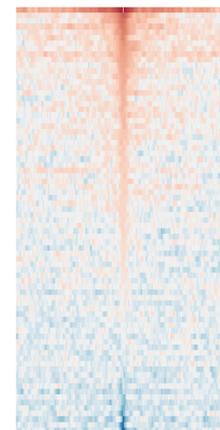
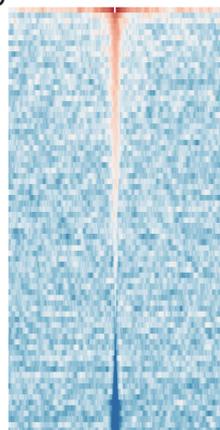
88-04



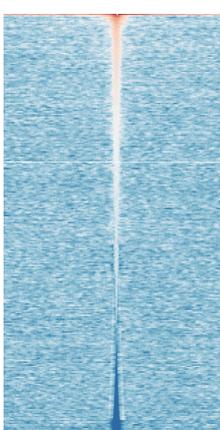
05-33



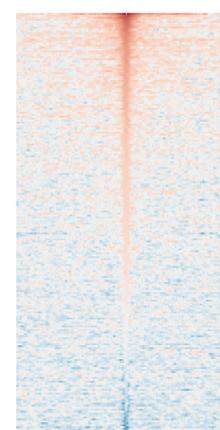
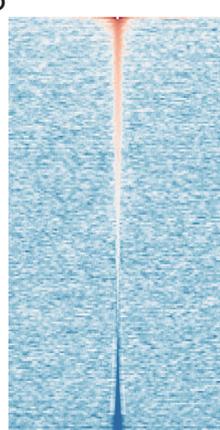
05-26



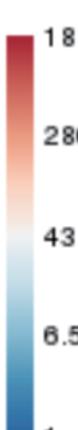
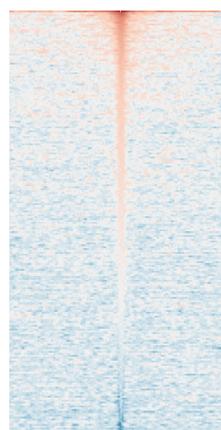
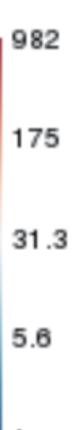
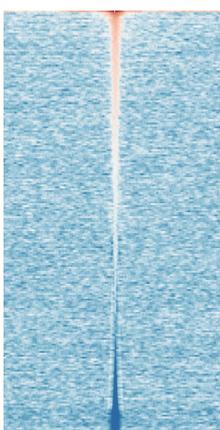
05-45



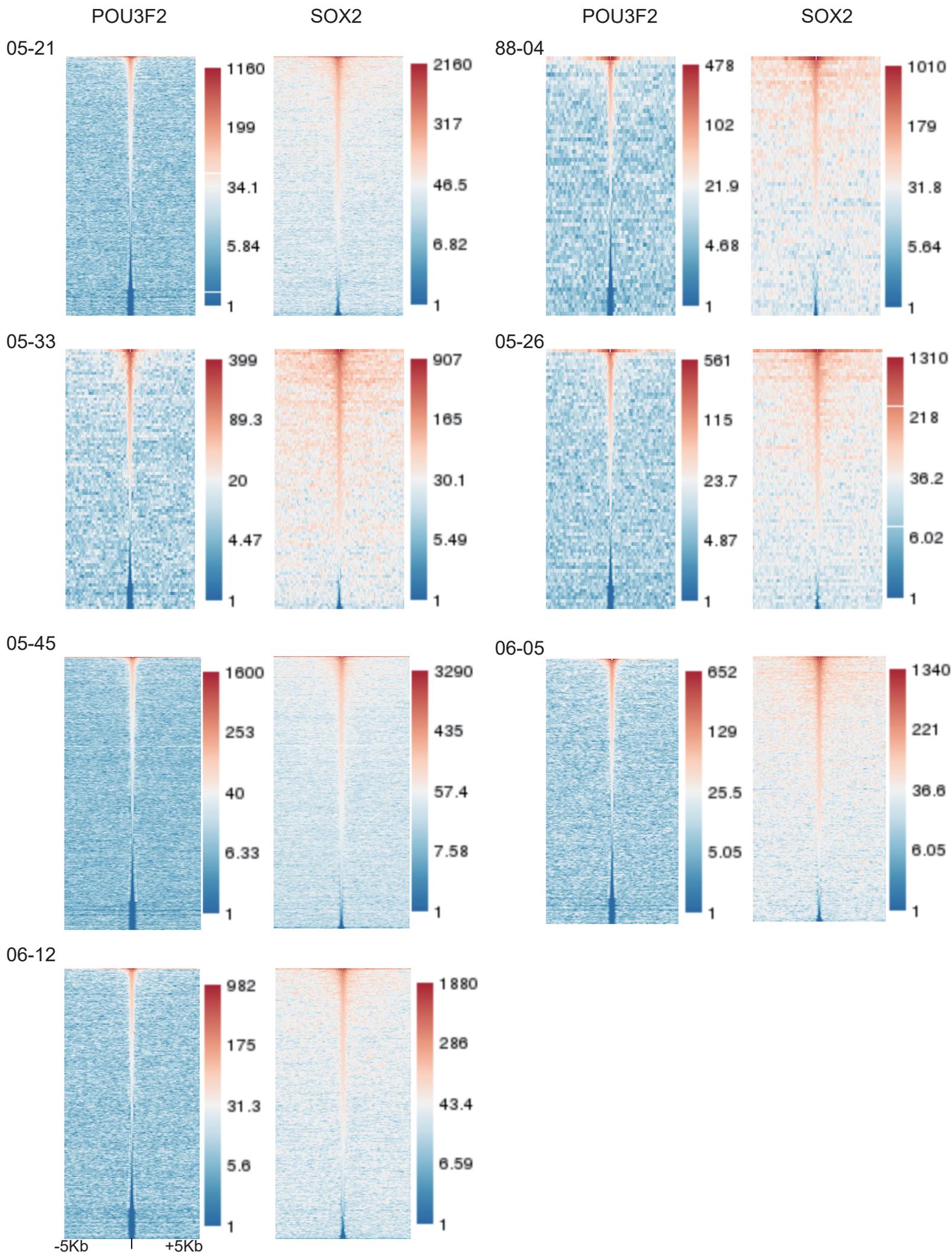
06-05



06-12



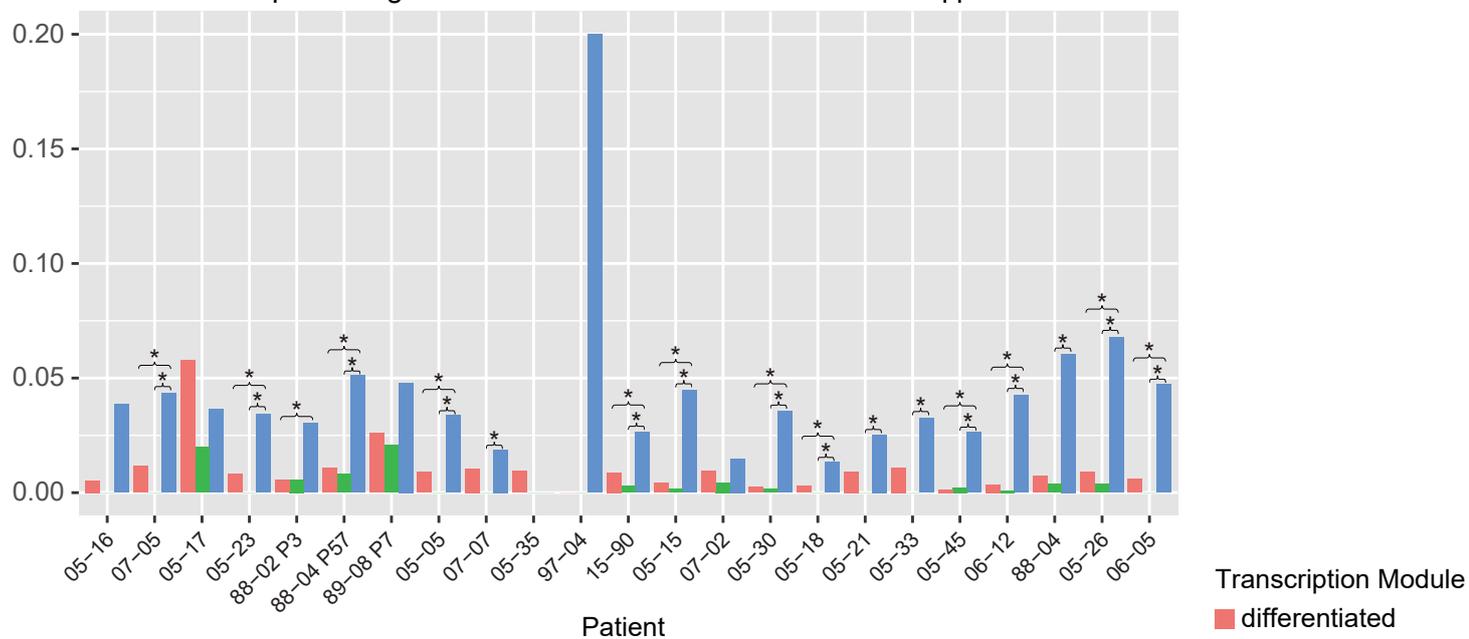
-5Kb | +5Kb



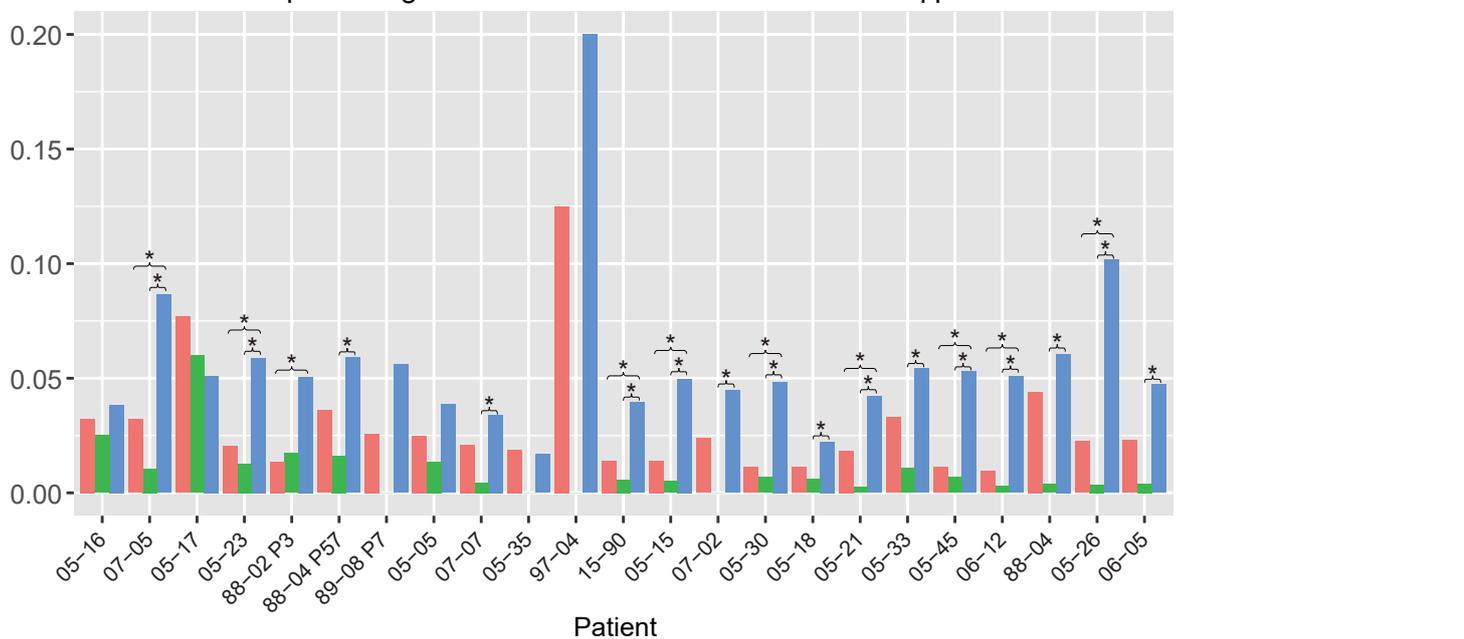
**Fig.S18 taTREs show enrichment of POU3F2 and SOX2 binding in tumor propagating cells.** Heat-maps show ChIP-seq signals for POU3F2 (green) and SOX2 (red) in tumor propagating cells  $\pm 5$ kb surrounding the center of taTREs. Data was from (26). Rows were ordered by the sum of ChIP-seq signals. Plots are made using the R pheatmap package (49).

**a**

percentage of intersections with POU3F2 ChIP-seq peak

**b**

percentage of intersections with SOX2 ChIP-seq peak



**Fig.S19 Stem program taTREs enriched for POU3F2 and SOX2 ChIP-seq peaks.** The height of bars shows the percentage of POU3F2 (a) and SOX2 (b) ChIP-seq peaks found intersected with taTRE in each of the primary GBM / PDX samples. taTREs from differentiated and stem programs are colored in red and green respectively. Primary GBM / PDX samples in which ChIP-seq peaks were enriched in stem program taTREs are marked by asterisk ( $p < 0.05$ ).