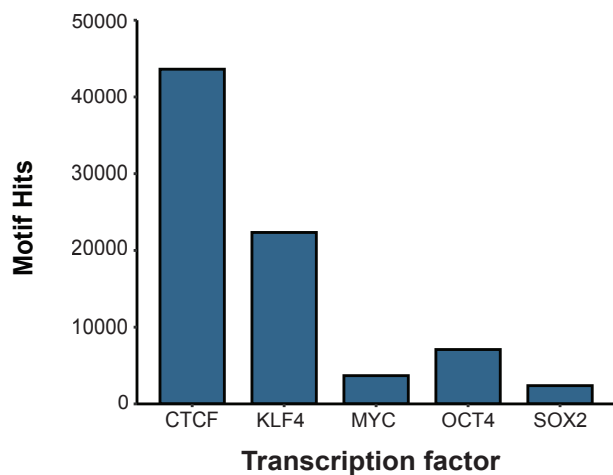
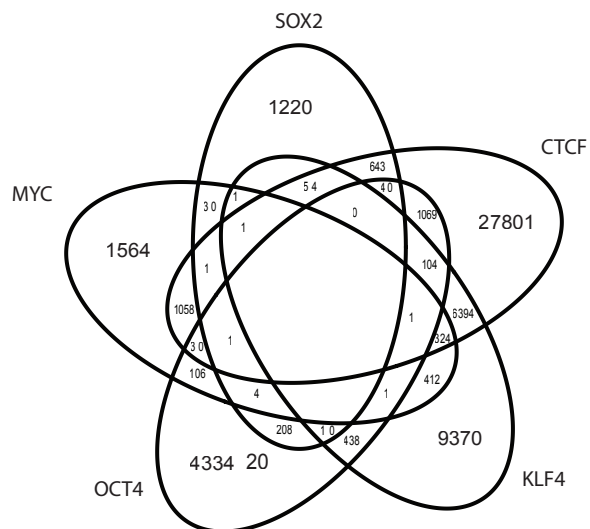
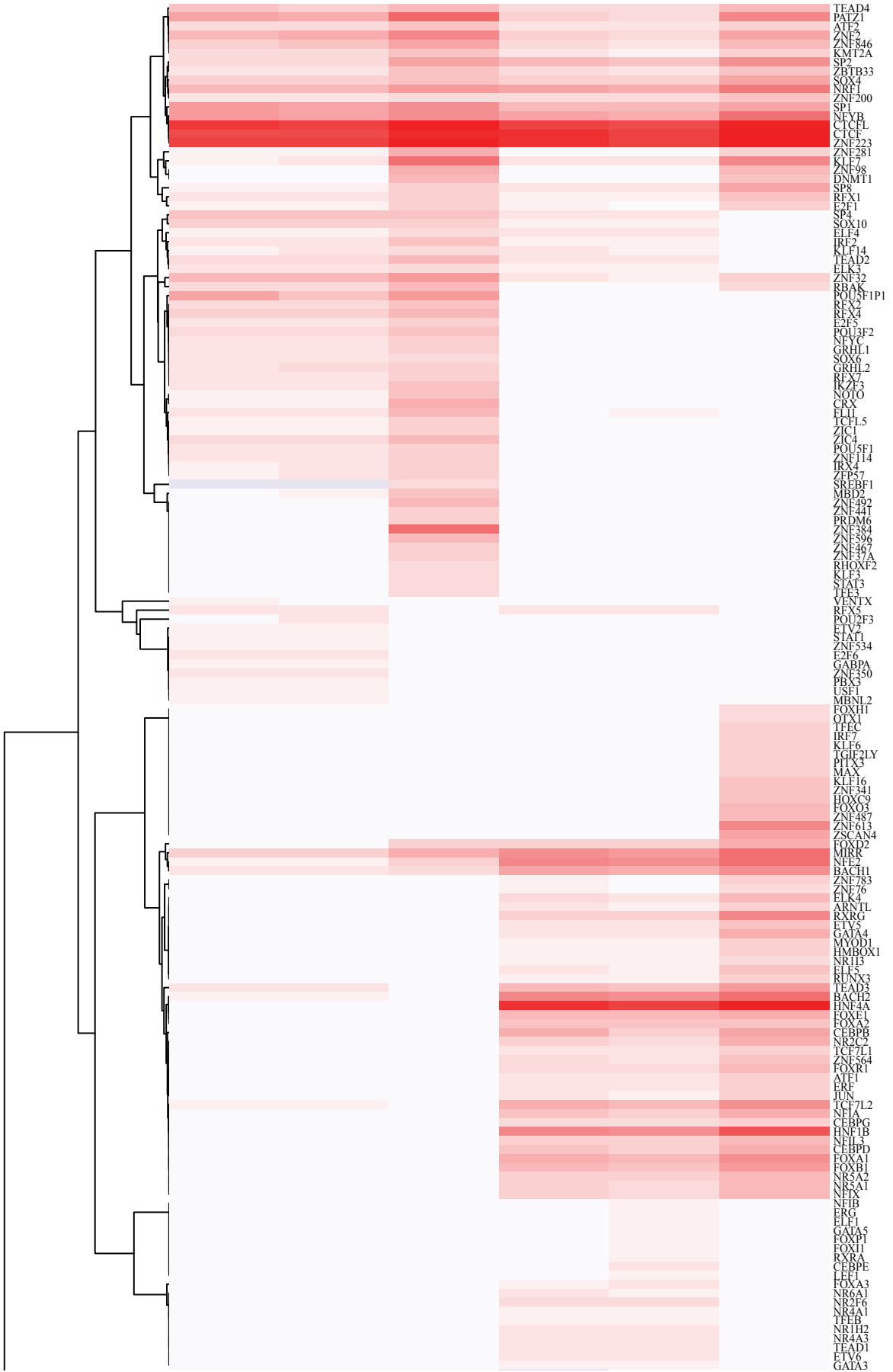
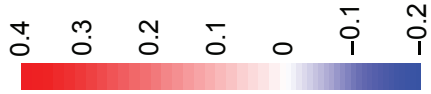


**Figure S1: AUROC/AUPRC values for selected classifiers in this study.** a) Three ML models (LR, Basset and Is-gkm) were trained (except Basset) and tested on various DHS lengths as input. Positives were DHS sites obtained from the HepG2 cell line. Negatives were random DNA regions of equal length as the positive and matched for GC and repeat element content. b) AUROC/AUPRC values of the LR model trained on DHS sites (200 bp) from ESC-H1 and HepG2 cell types using ~ top 150 features (for each cell type) obtained after feature selection via three different methods (ElasticNet, L1 and RFE).

**a****b**

**Figure S2: Motif hits in ESC-H1 DHS sites.** a) Number of motif hits for ESC-H1 regulators (OCT4, Sox2, KLF4, and MYC) and CTCF, within ESC-H1 DHS sites. b) Venn diagram indicating overlap of ESC-H1 DHS sites containing motif hits for ESC-H1 regulators (OCT4, Sox2, KLF4, and MYC) and CTCF.

LR coefficients



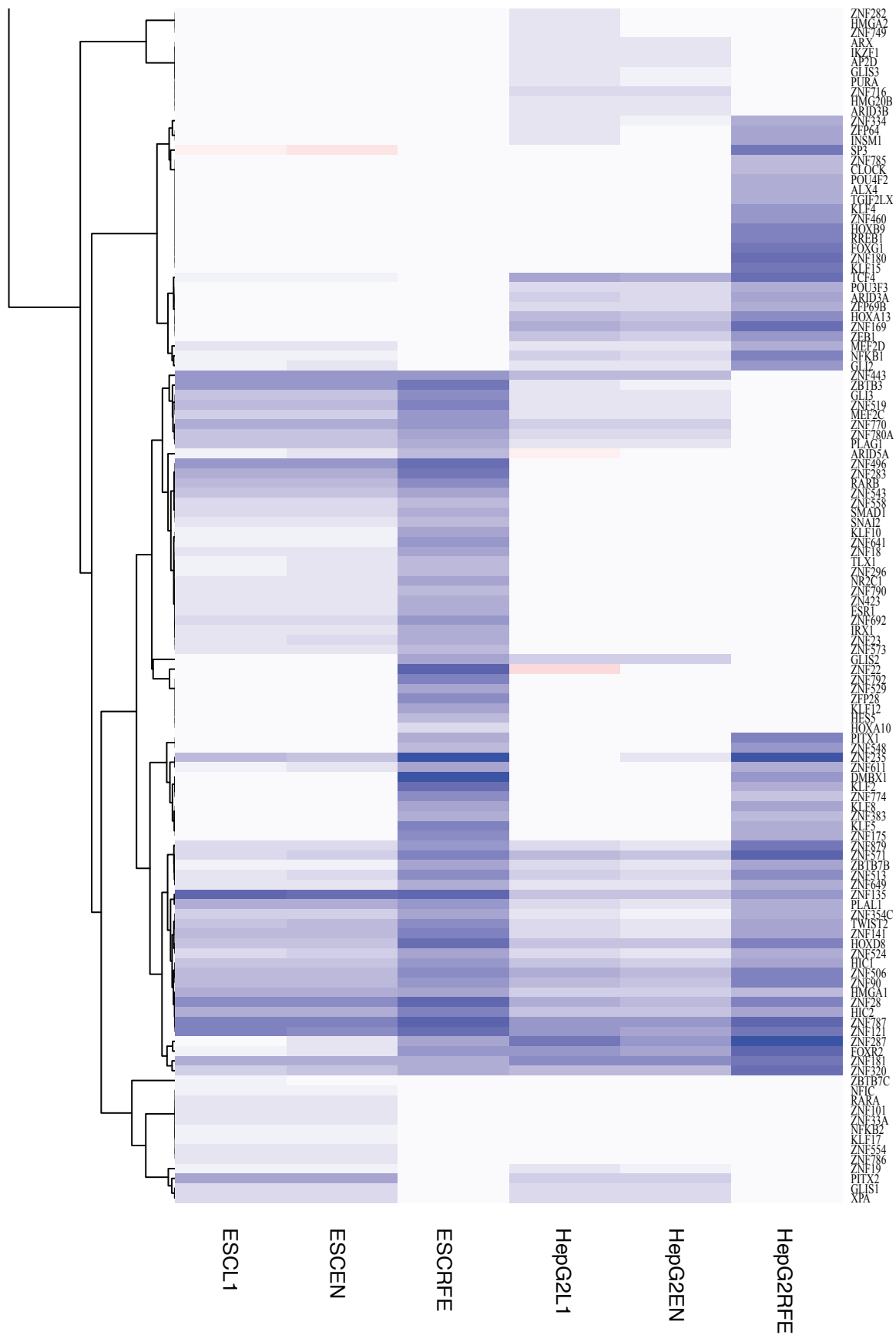
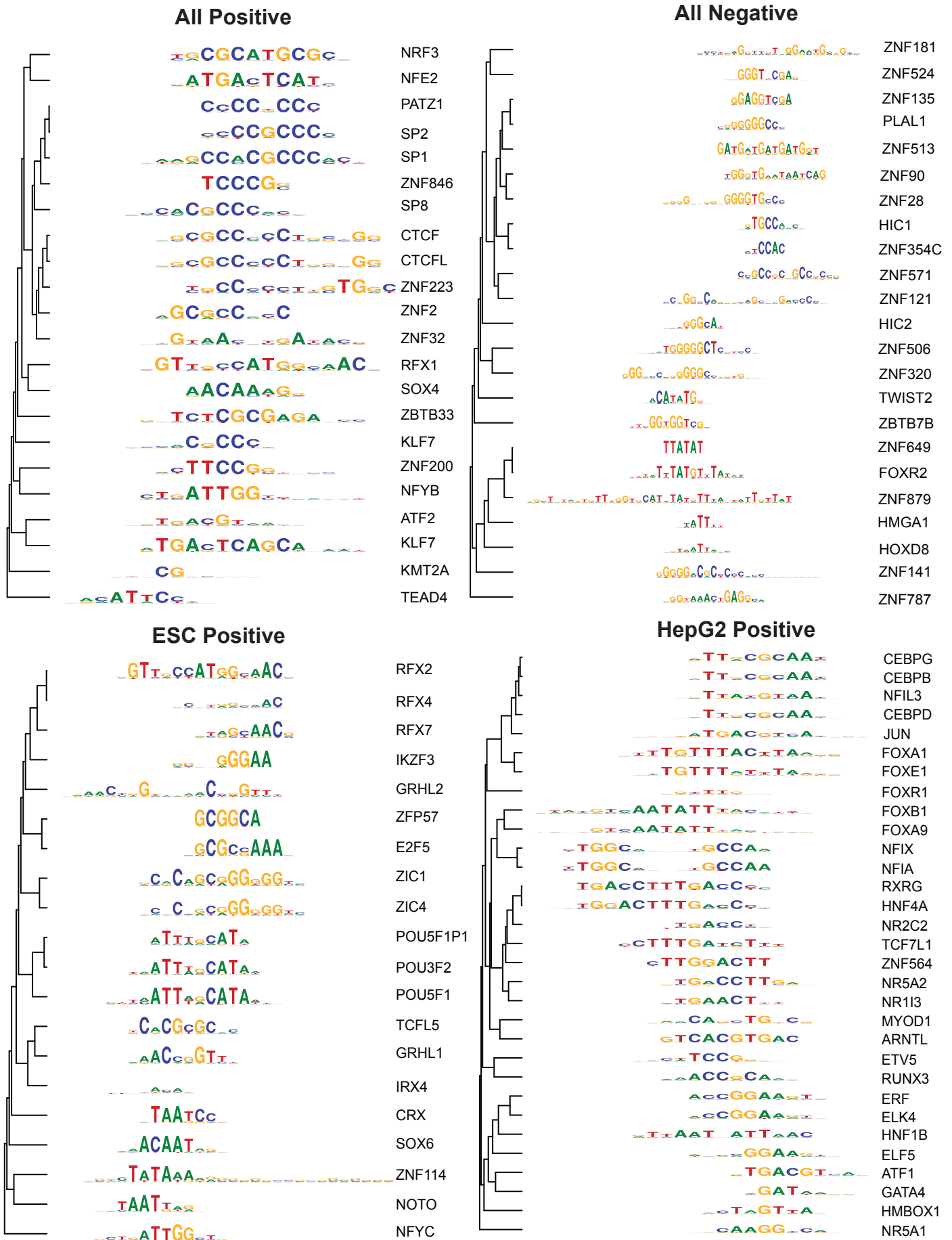


Figure S3: Heatmap (in Figure 6) with all TF names



**Figure S4: Motifs of features shown in pullouts in Figure 6.** Motifs were clustered and aligned using tree layout command in MotifStack R package [68].