Supplementary Table 1

Cell type specific genes and regulatory regions selected for each cell type

Gene	Enhancer Region	Gene Region	Cell Type
GATA1	chrX:48782929-48783129	chrX:48785536-48787536	K562
HNF4A	chr20:44370692-44370892	chr20:44355699-44432845	HepG2
CD19	chr16:28930777-28930977	chr16:28931971-28939342	GM12878

Supplementary Table 2

Genes tested for reactivation for each cell type

GM12878	GATA1
	HNF4A
K562	HNF4A
	CD19
HepG2	CD19
	GATA1

Supplementary Table 3

ENCODE IDs of samples used for the training of the MPRA predictor model

ENCFF996ECA
ENCFF018AMJ
ENCFF345ASG
ENCFF970OLE
ENCFF318XMJ
ENCFF821XQZ
ENCFF358MBK
ENCFF379XWL
ENCFF774CHX
ENCFF138DJM
ENCFF277DDE
ENCFF334EKU
ENCFF857FQR
ENCFF259NMG

ENCFF477LDL
ENCFF484JFE
ENCFF227KRF
ENCFF102ZVT
ENCFF418GRL
ENCFF333BAD
ENCFF307HBZ
ENCFF771HPB
ENCFF359KJL
ENCFF035HKU
ENCFF759PPO
ENCFF705AES
ENCFF256WKS
ENCFF352JAC
ENCFF147SMK
ENCFF311DJW
ENCFF350IJA
ENCFF815ORW
ENCFF402GOL
ENCFF865LNO
ENCFF755GRH
ENCFF440YVF
ENCFF703OIL
ENCFF927USI
ENCFF476FXK
ENCFF742ENC
ENCFF112HAT
ENCFF792IHA
ENCFF267VJ
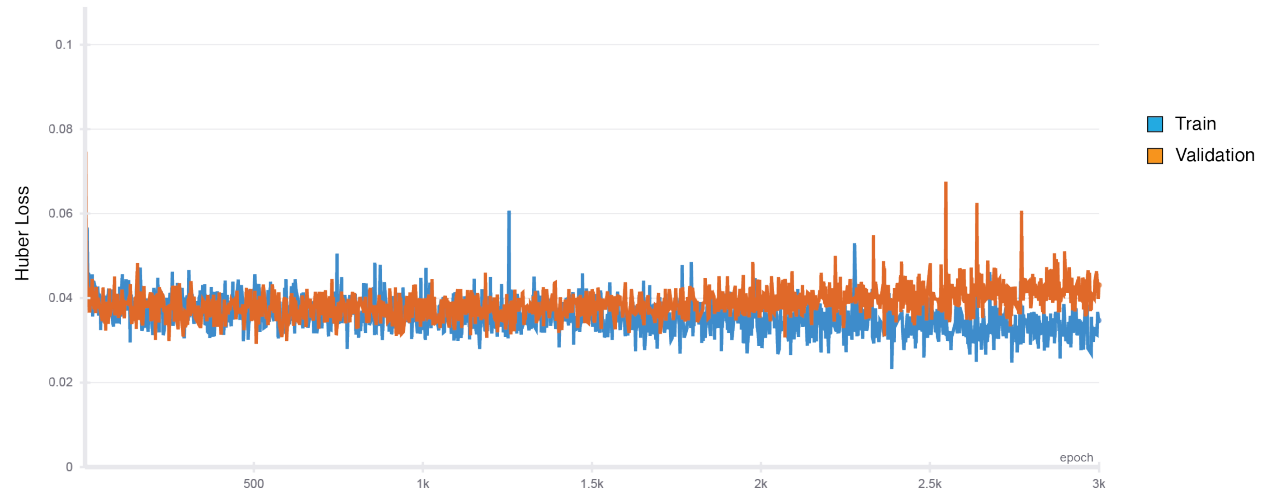
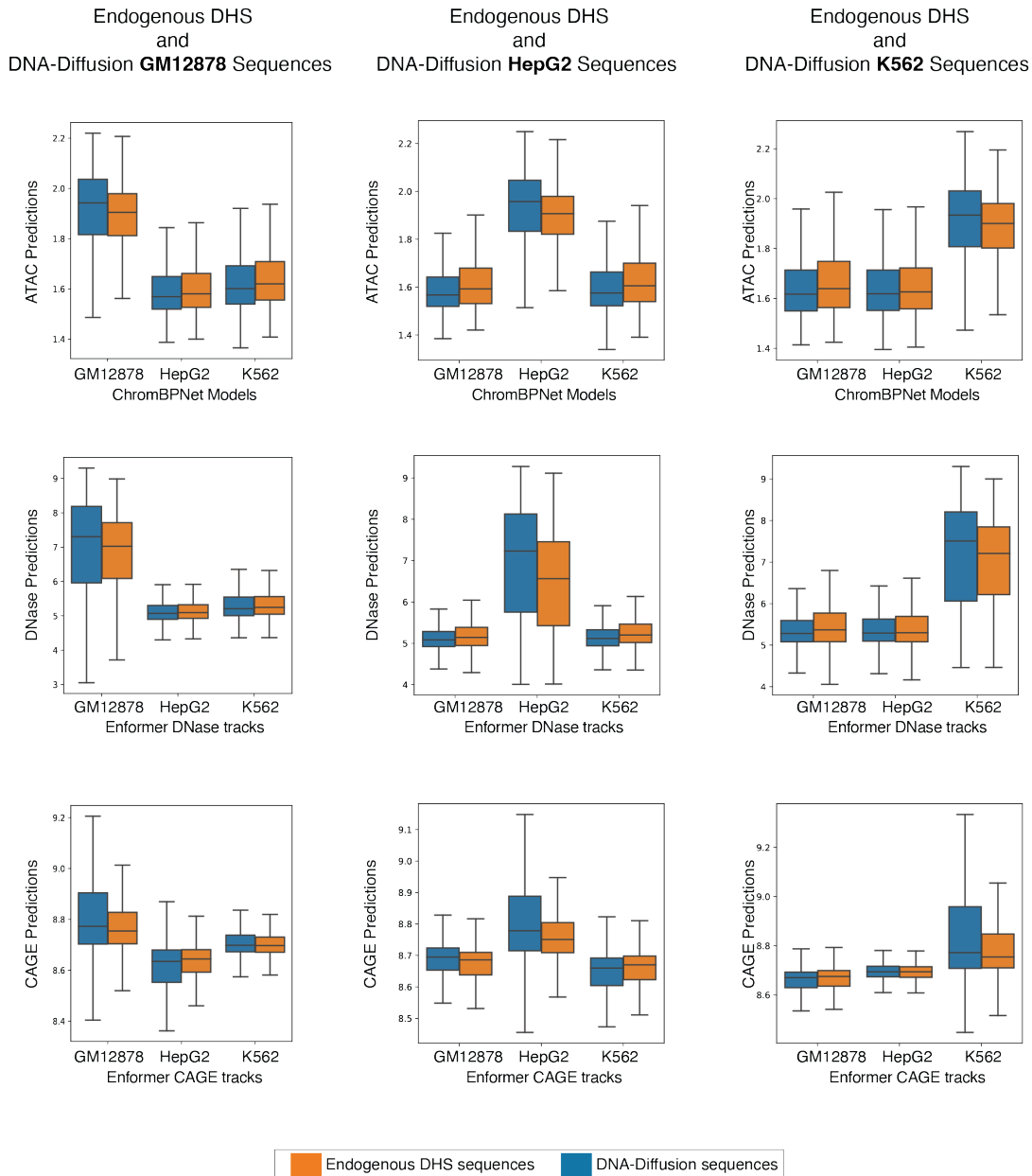**Supplementary Figure 1. DNA-Diffusion training loss plot.**
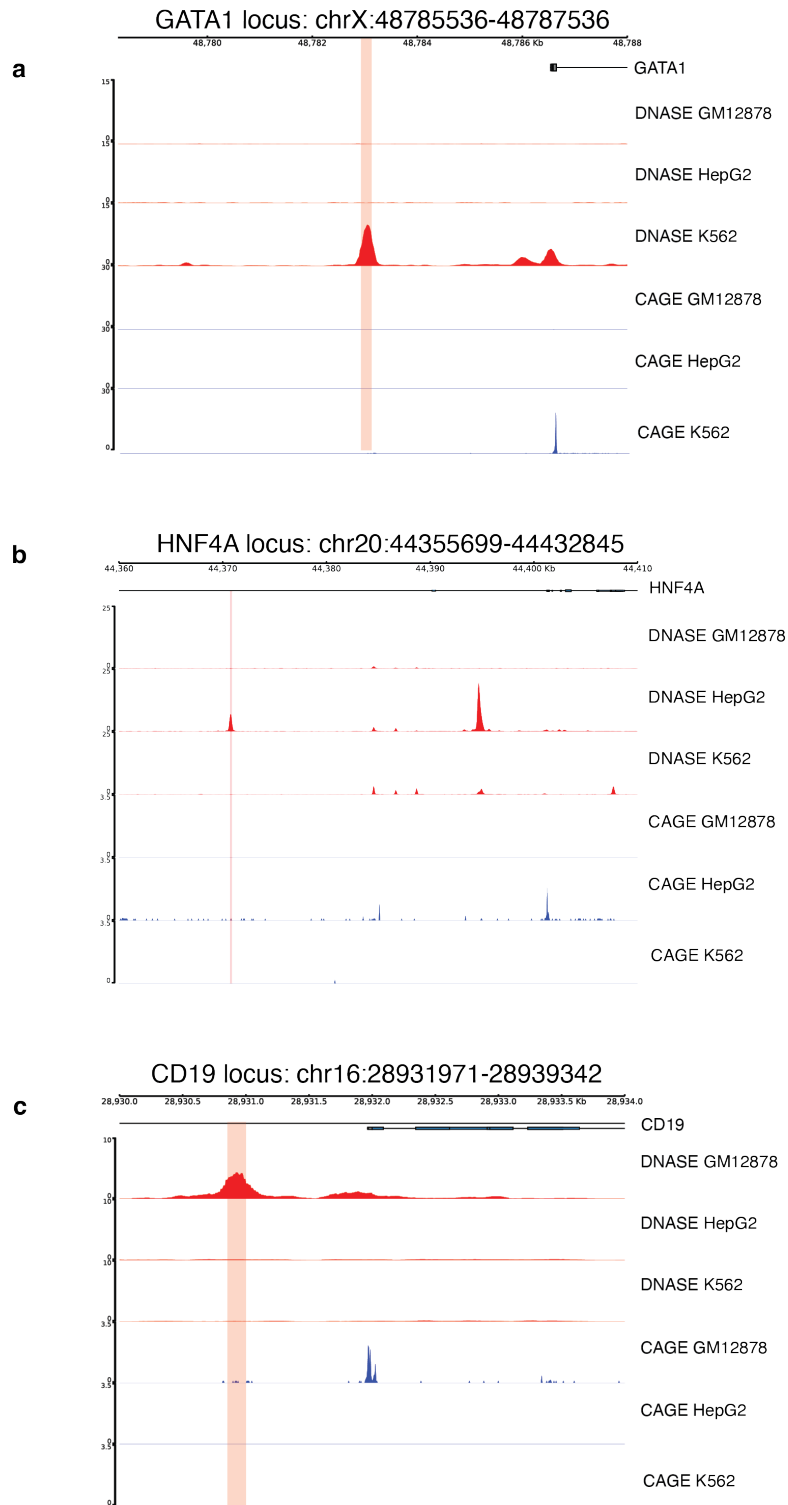
Supplementary Figure 2. Endogenous DNase and CAGE Tracks. Chromatin accessibility (ENCODE DNase, red track) and gene expression (FANTOM CAGE, blue track) profiles are shown for three key gene loci. Each panel displays a window in the genomic region potentially controlling a specific gene (GATA1, HNF4A, CD19).
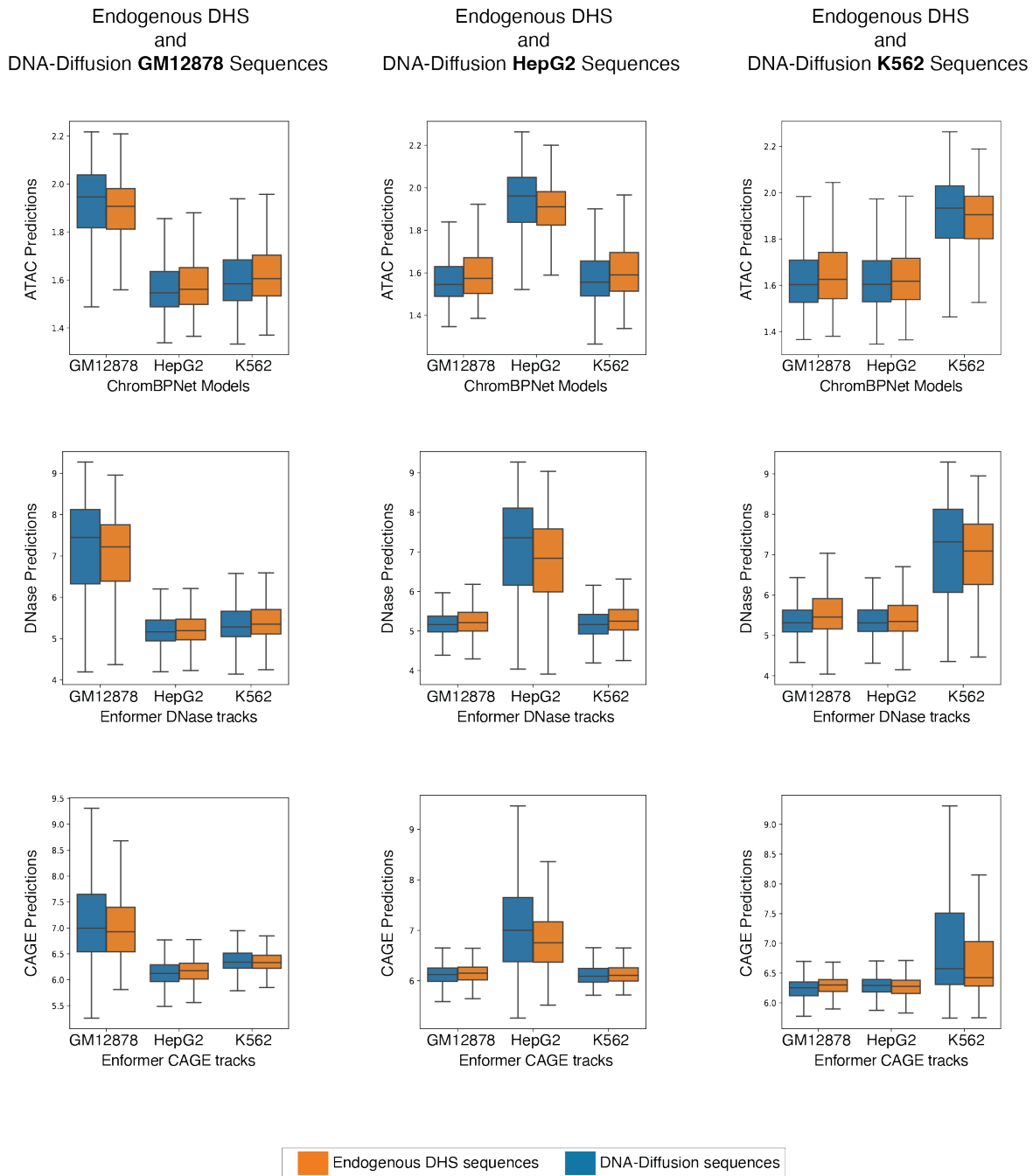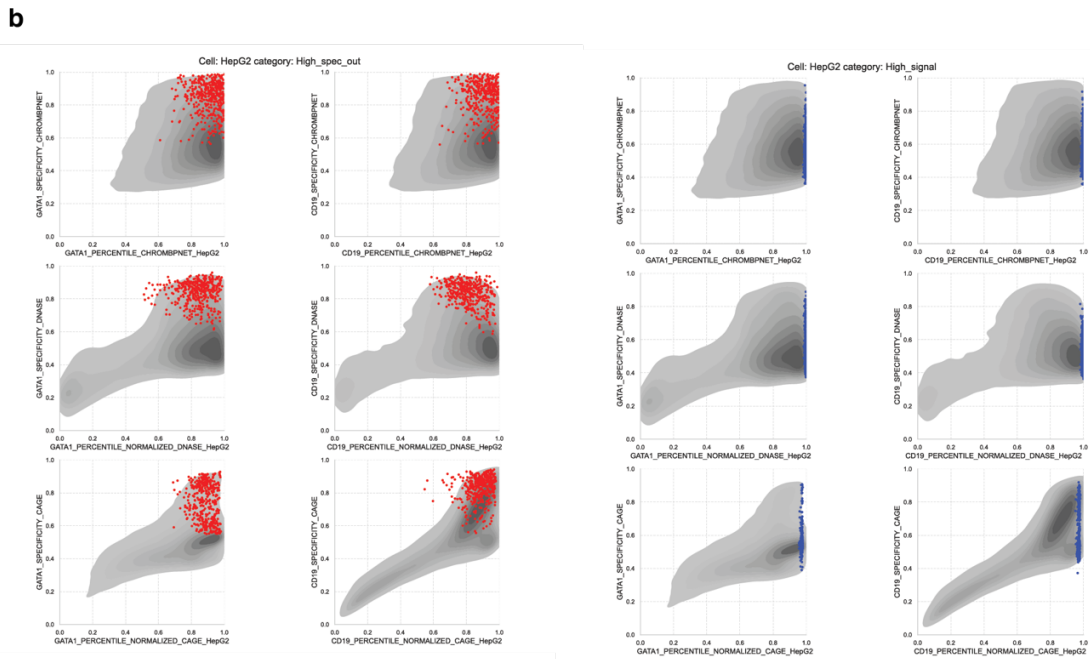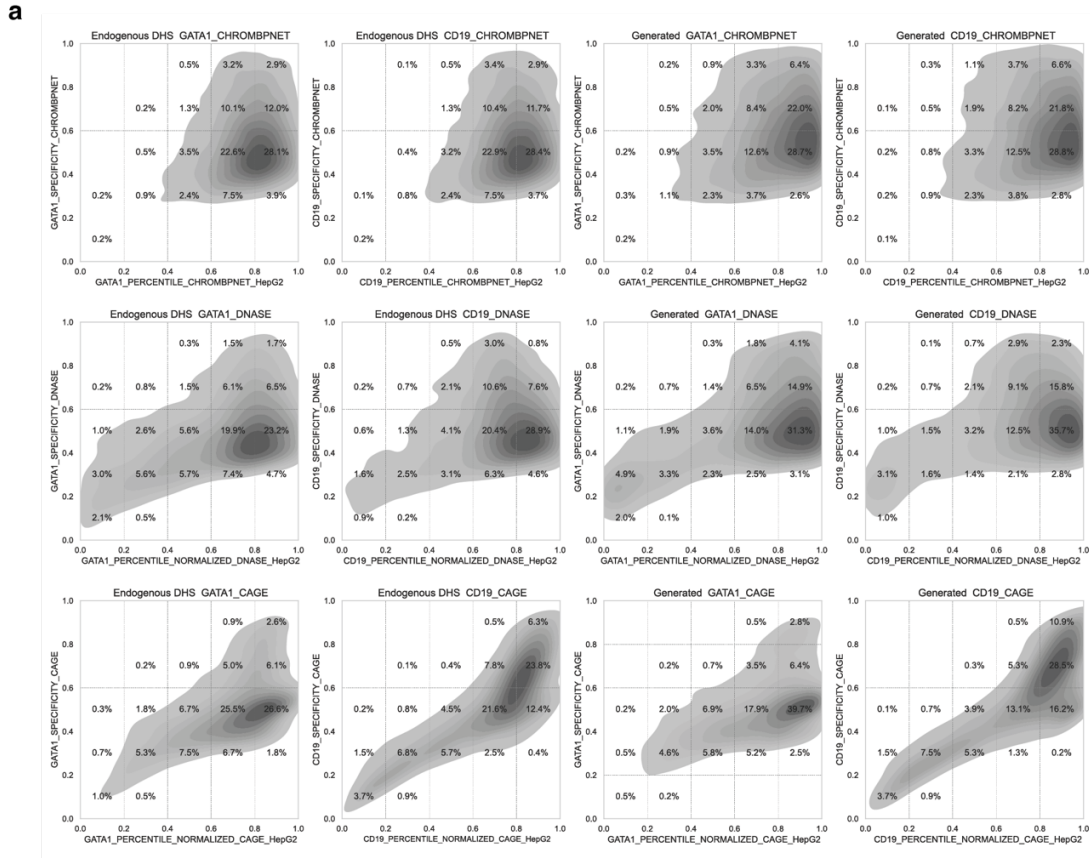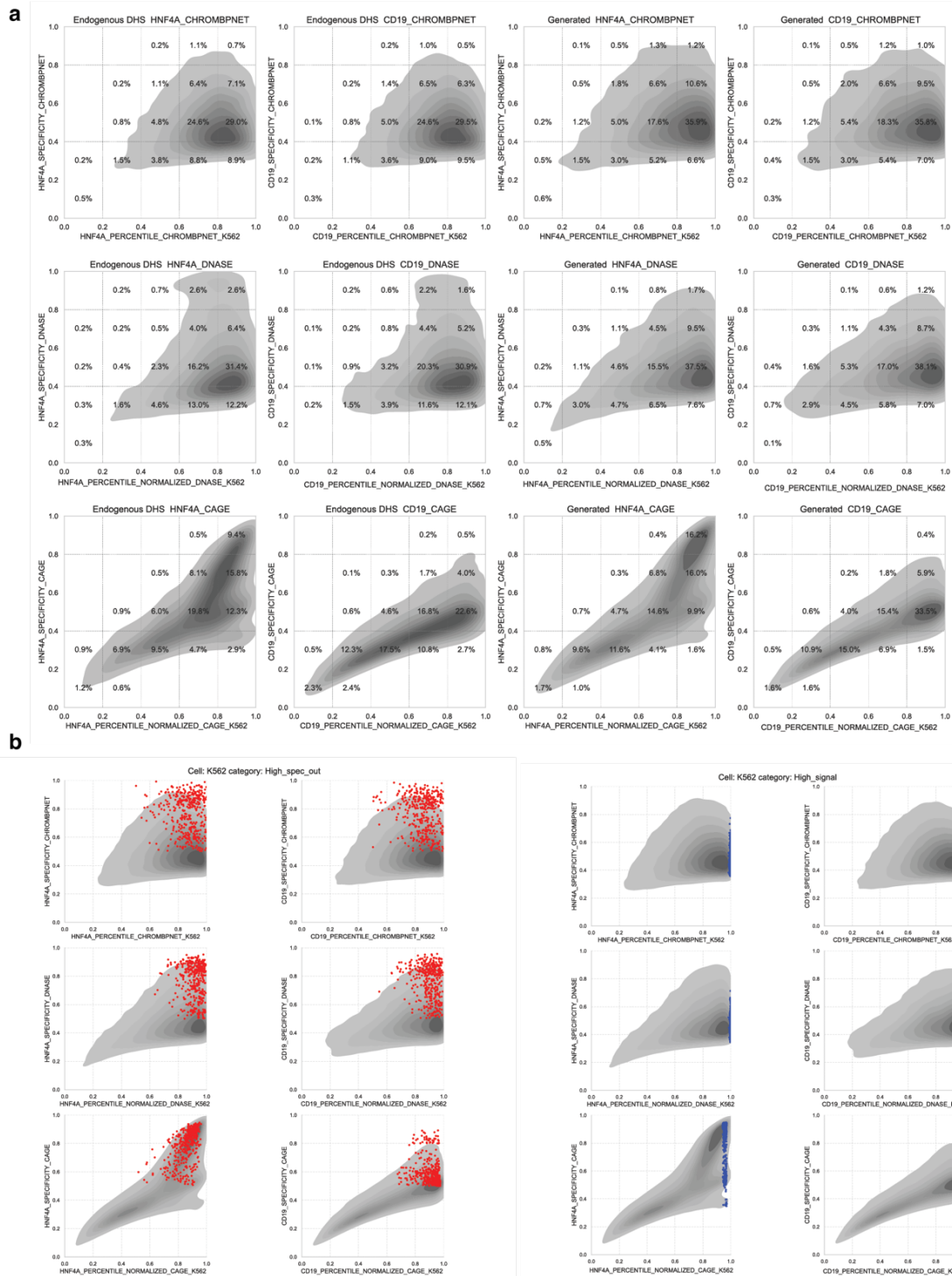
Supplementary Figure 3. Endogenous train and DNA-Diffusion predictions across all three downstream oracles within the HNF4A gene locus. a) Boxplots showing the predicted chromatin accessibility/gene expression activity (ChromBPNet ATAC, Enformer DNase, Enformer CAGE) upon replacement within the HNF4A locus with endogenous DHS train and DNA-Diffusion sequences specific for each cell line (GM12878, HepG2, K562).

Supplementary Figure 4. Endogenous train and DNA-Diffusion predictions across all three downstream oracles within the CD19 gene locus. a) Boxplots showing the predicted chromatin accessibility/gene expression activity (ChromBPNet ATAC, Enformer DNase, Enformer CAGE) upon replacement within the CD19 locus with endogenous DHS train and DNA-Diffusion sequences specific for each cell line (GM12878, HepG2, K562).

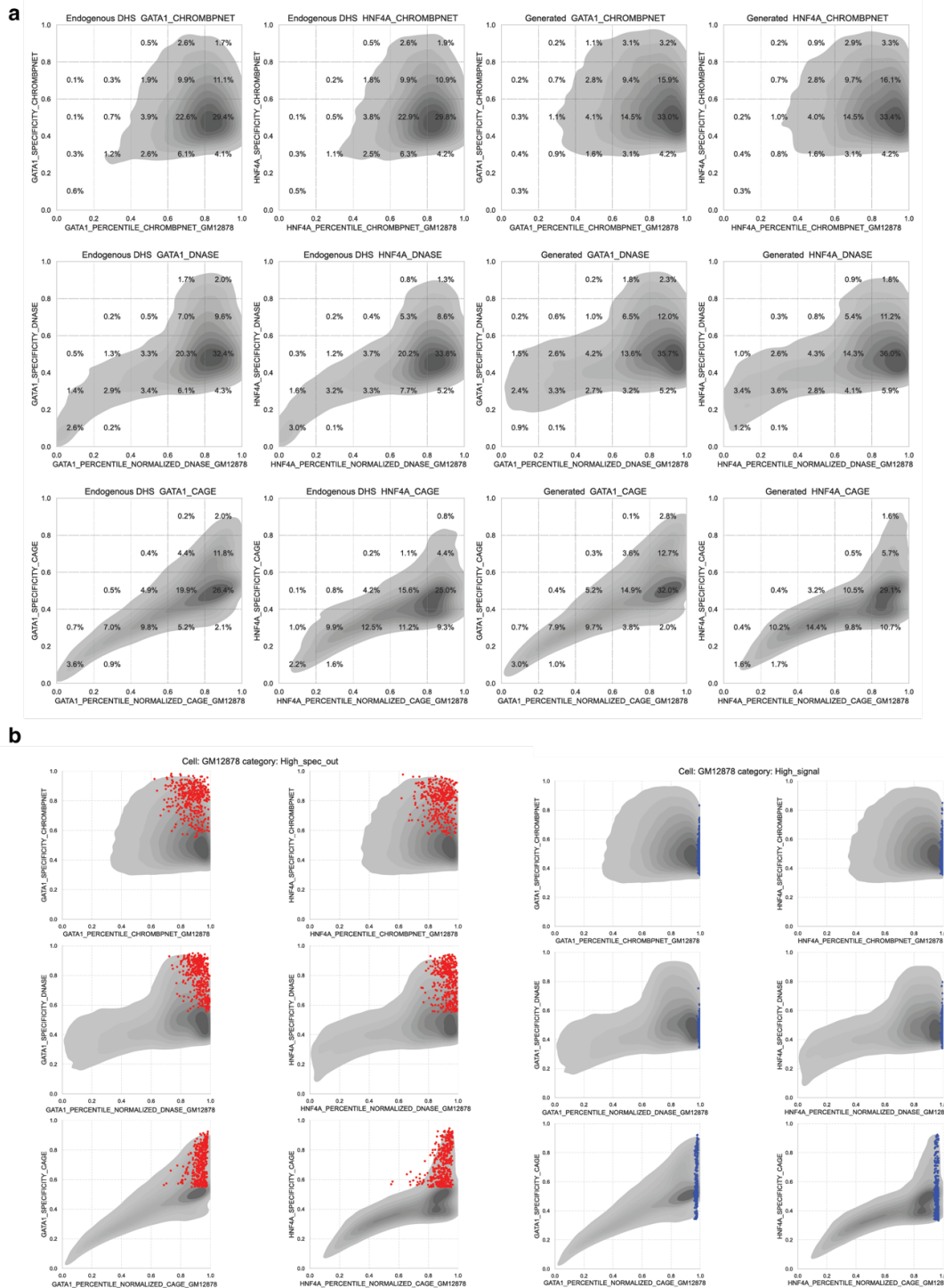**Supplementary Figure 5. Oracle prediction scores with HepG2 cell type-specific sequences across CD19 and GATA1 gene loci.**

Supplementary Figure 6. Oracle prediction scores with K562 cell type-specific sequences across CD19 and HNF4A gene loci.

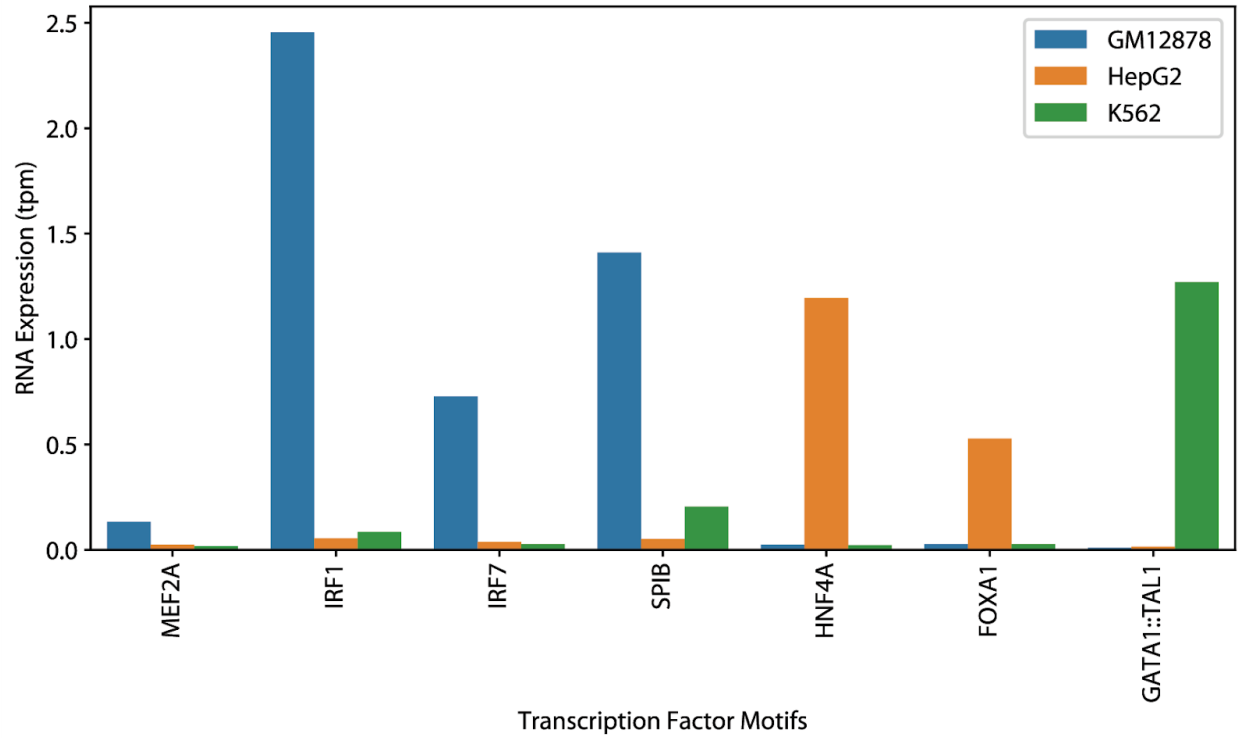**Supplementary Figure 7. Oracle prediction scores with K562 cell type-specific sequences across HNF4A and GATA1 gene loci.**
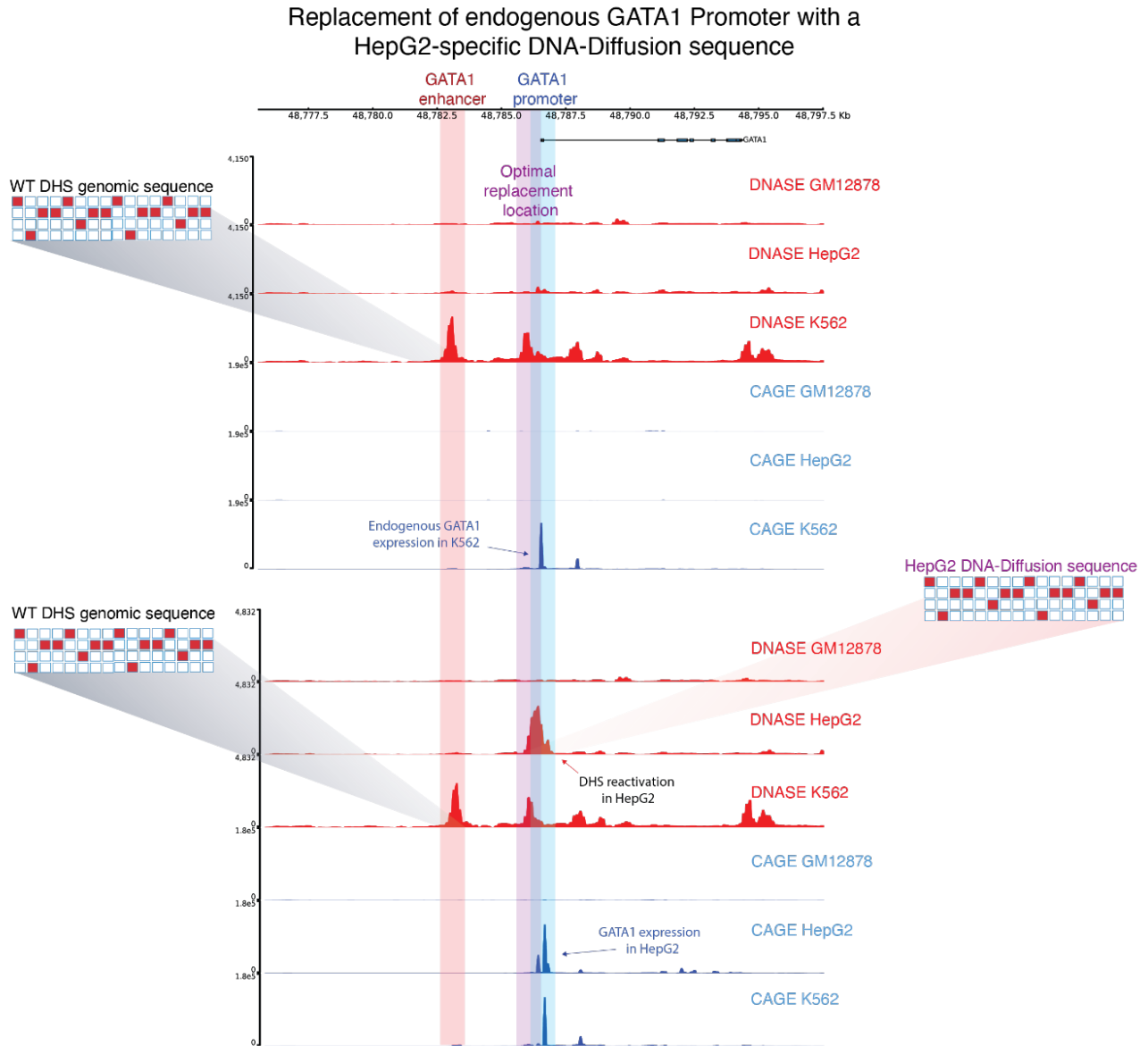
Supplementary Figure 8. Exploring the effect of GATA1 promoter region replacement with HepG2-specific DNA-Diffusion sequence

Supplementary Figure 9. Endogenous RNA expression values for selected cell type-specific TF motifs. Bar plots showing TPM values for cell type specific TFs from **Fig. 6** across the three cell types.