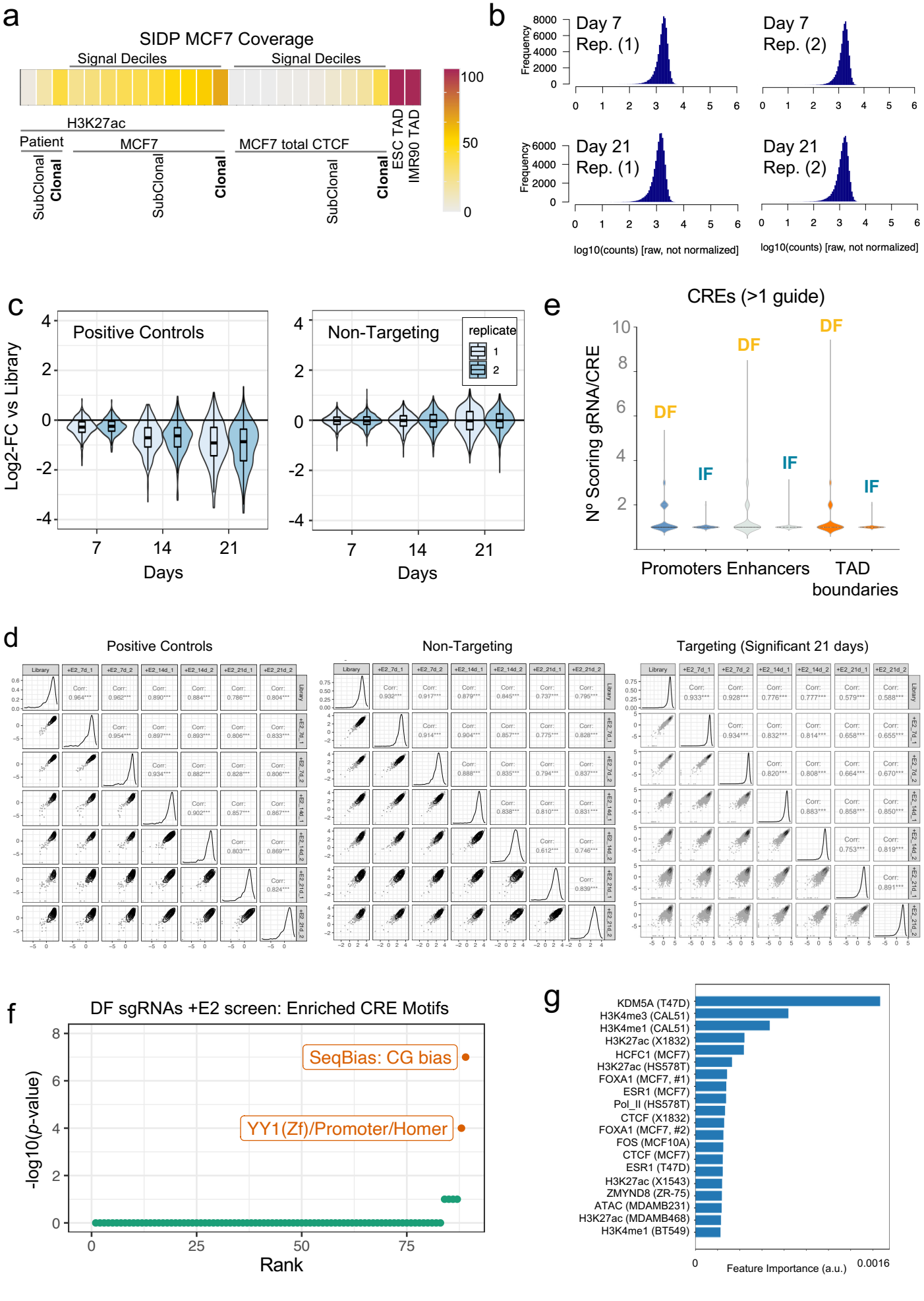
Supplementary Figure S1



Supplementary Figure 1. SIDP identifies Cis Regulatory regions in MCF7 with high reproducibility. (a) SIDP coverage (percentage) of the specific partitions of the human CREs considered in this study. (b) Histograms showing the distribution of counts per sgRNAs (log10) for two replicates of sgRNAs in pool 1, at day 7 and day 21 post-infection (MCF7 +E2). (c) Box plots overlaid on violin plots showing the log2-fold-change of positive controls (left panel) and non-targeting sgRNAs (right panel) in two replicates of MCF7 cells, at 7, 14 and 21 days, as compared to the initial library. (d) Violin plots showing the distribution of the number of significantly scoring sgRNAs per CRE, for DF (decreased frequency; yellow) and IF (increased frequency; blue) sgRNAs, across three different genomic partitions (promoters, putative enhancers, and CTCF-clusters associated to TAD boundaries). (e) Correlograms (considering either positive controls, non-targeting sgRNAs, or targeting sgRNAs scoring significantly at 21 days post-infection) between the TMM-normalized, log2-transformed counts across replicates (7, 14 and 21 days post-infection; +E2). The initial library is also included. Scatterplots of each pair of samples are drawn on the left part of grid. Pearson correlation is displayed on the right (p-value: *** < 0.001; ** < 0.01; * < 0.05; . < 0.10). Distribution of values for each sample is shown on the diagonal. (f) Motif analysis of CREs associated with significantly exhausted sgRNAs identifies YY1 as a putative TF enriched in functional CREs. (g) Feature of a deepSEA model trained to distinguish scoring from non-scoring sgRNA highlights specific epigenetic and TF-binding signatures associated with functional CREs. importance