

Supplementary information, Figure S4

Figure S4

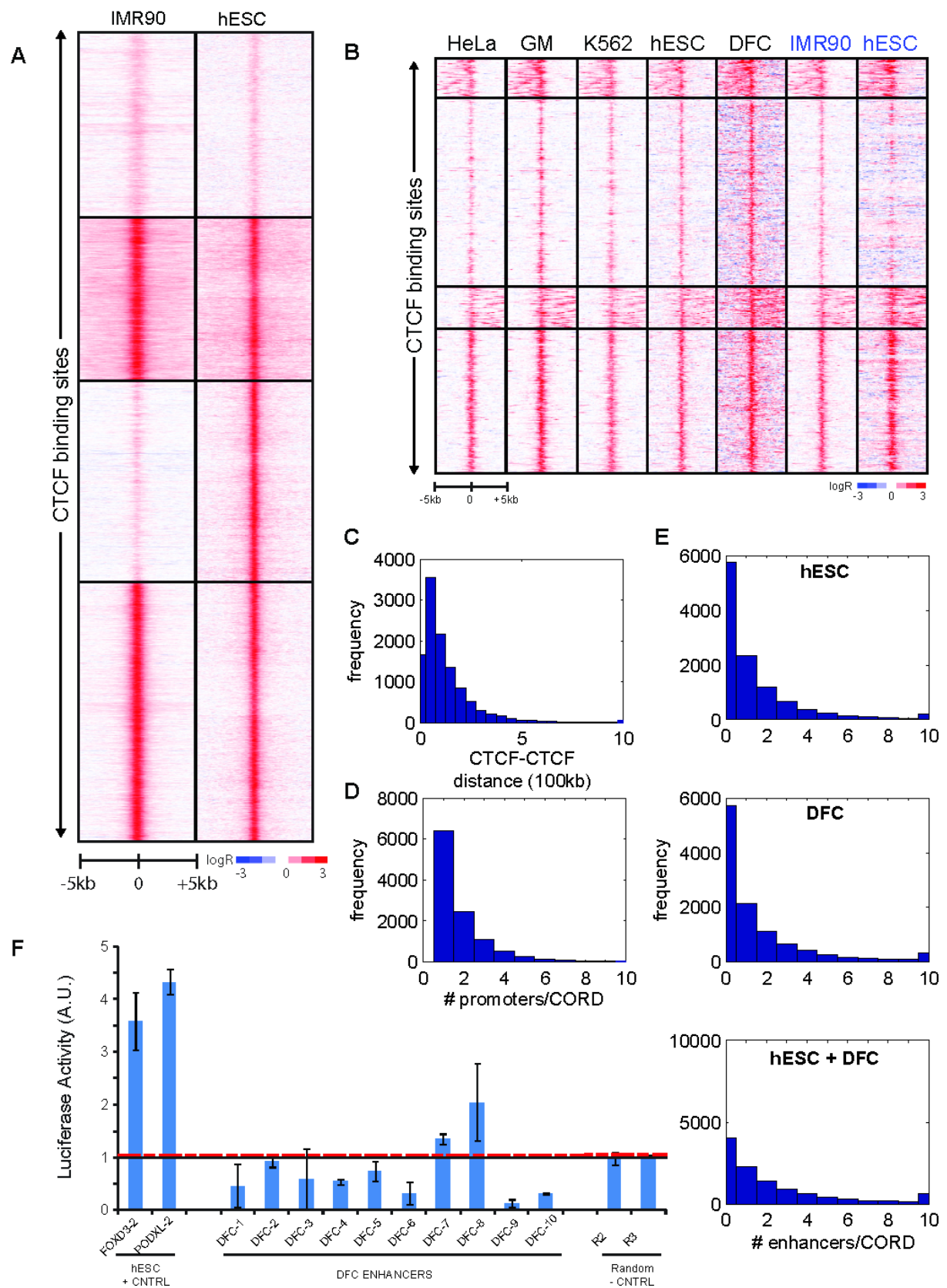


Figure S4, related to Figure 3 Additional analysis of CTCF binding site analysis and CTCF-defined domains. **(A)** Clustergram at 29 880 combined CTCF binding sites determined from IMR90 (Kim *et al.*, 2007), HeLa (Wendt *et al.*, 2008), and CD4 cells (Barski *et al.*, 2007), and assessed for in genome-wide CTCF binding sites found in IMR90 and hES cells. **(B)** k-means cluster gram of CTCF binding sites in the ENCODE region from HeLa, GM06990 (GM), K562 leukemic cells, hESCs, and DFCs from Heintzman *et al.*, 2009. These are compared to ENCODE regions extracted from genome-wide data in IMR90 cells (Kim *et al.*, 2007) and hESCs presented here (blue). **(C)** The genome-wide distribution of adjacent CTCF-CTCF distances. **(D)** The genome-wide distribution of the number of promoters per CORD. **(E)** The genome-wide distribution of adjacent enhancer-enhancer distances for: (top) hESC predicted enhancers only, (middle) DFC predicted enhancers only, and (bottom) combined predictions. **(F)** Reporter assays of enhancer function at 10 predicted DFC-specific enhancers, 2 randomly chosen genomic regions as negative controls, and 2 hESC-specific enhancers as positive controls, cloned downstream of a luciferase gene and tested for activity in hESCs. The dashed red line indicates a *P*-value cutoff of 1%.