Expanded View Figures

Figure EV1. Expression of core naïve TFs does not explain strain-specific transcriptional programs in ESCs.

A PCA biplot showing individual gene contributions to the first two principal components. Transcript abundance measured by bulk RNA-seq from 3 biological replicates for each strain in ESCs and EpiLCs. Core naïve and primed marker genes are indicated, drive separation between cell states, and are largely not differentially expressed between strains. As previously described (Buecker *et al*, 2014), *Pou5f1* expression is unchanged across cell states.

B Heatmap of bias-corrected deviation *z*-scores of chromatin accessibility at TF motifs ($n = 3$ biological replicates of ESCs and EpiLCs from both B6 and D2 strains).

C MA plot highlighting genes differentially expressed between strain within EpiLCs as determined by pairwise comparison (FDR < 0/05, $\log_2$FC > 1). Core primed TFs are highlighted as not significantly differentially expressed between strains.

D GSEA for functional programs curated within MsigDB using DEGs between B6 and D2 ESCs. Upregulated genes in B6 ESCs were enriched for GO terms in biological processes associated with self-renewal and pluripotency such as regulation of mitotic metaphase and anaphase and metabolism of amino acids and derivatives (significant enrichment in phenotype using 1,000 permutations, FDR < 25%), as well as regulation of chromatin organization (significant for single gene set, nominal *P*-value < 0.05).

E Similar to (D) showing pathways upregulated in D2. D2 ESCs were associated with regulation of G2/M phase transition and anterior–posterior pattern specification (FDR < 25%) in addition to self-renewing programs such as WNT signaling (nominal *P*-value < 0.05).

F PCA biplot indicating selected gene weights contributing to occupancy of cells along naïve pluripotency spectrum from Fig 1F. Core naïve TFs (such as *Klf4*, *Tbx3*, *Nanog*, *Pou5f1*, and *Prdm14*) do not contribute to separation of cells grown in different culture conditions or differential occupancy of pluripotency continuum by B6 and D2 ESCs. Cells grown in conditions lacking 2i have higher expression of primed TFs (i.e., *Fgf5*, *Otx2*, *Wnt8a*). While B6 and D2 both associate with conditions that include 2i, D2 is separated from B6 ESCs by expressing transcription profiles similar to conditions that include serum. Gene weights selected are in top 100 genes contributing to separation along PC2 and PC3; only genes driving separation between B6 and D2 along PC3 are plotted at threshold of 0.15 contribution.

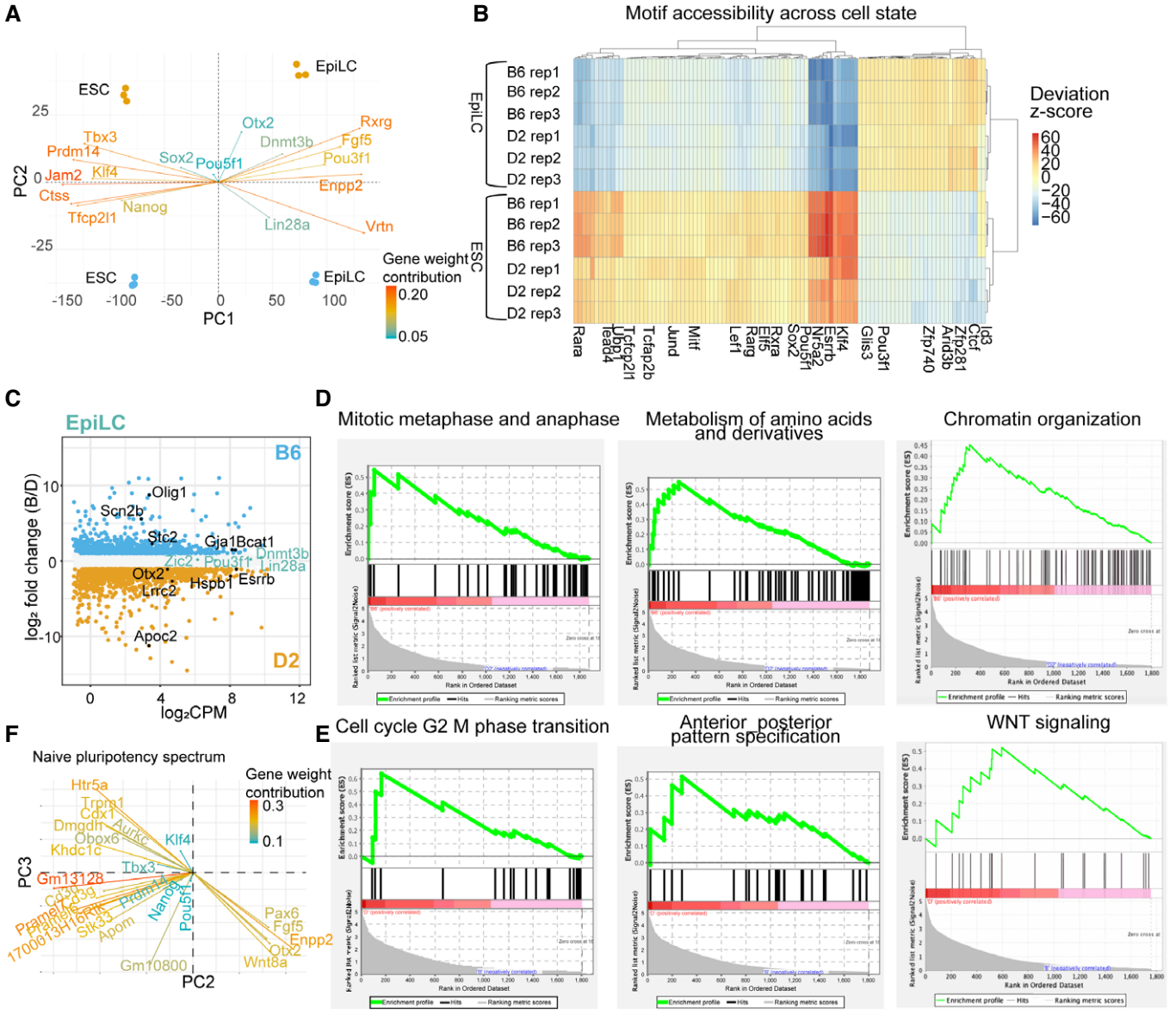

Figure EV1.

Figure EV2. Genetic background induces distinct expression pattern changes across cell state transition.

A–C Heatmaps showing expression patterns for genes detected with a significant difference based on general linearized model (glm, state + strain + stain x strain). Hierarchical clustering of samples (columns) results in proper association of cell states and biological replicates within genotype. (A) Genes with significant difference between cellular states independent of genetic background (n = 10,380 at FDR < 0.05). (B) Genes with significant difference between strains independent of state (n = 7,359 at FDR < 0.05). (C) Genes with significant interaction between state and strain (n = 4,912 at FDR < 0.05).

D–O Candidate gene expression patterns between genotype and state represent expression modules m1-12. Modules resulted from intersection between EBseqHMM-defined expression paths and genes with a significant GxS interaction. Number of genes within each module indicated at top in parentheses.

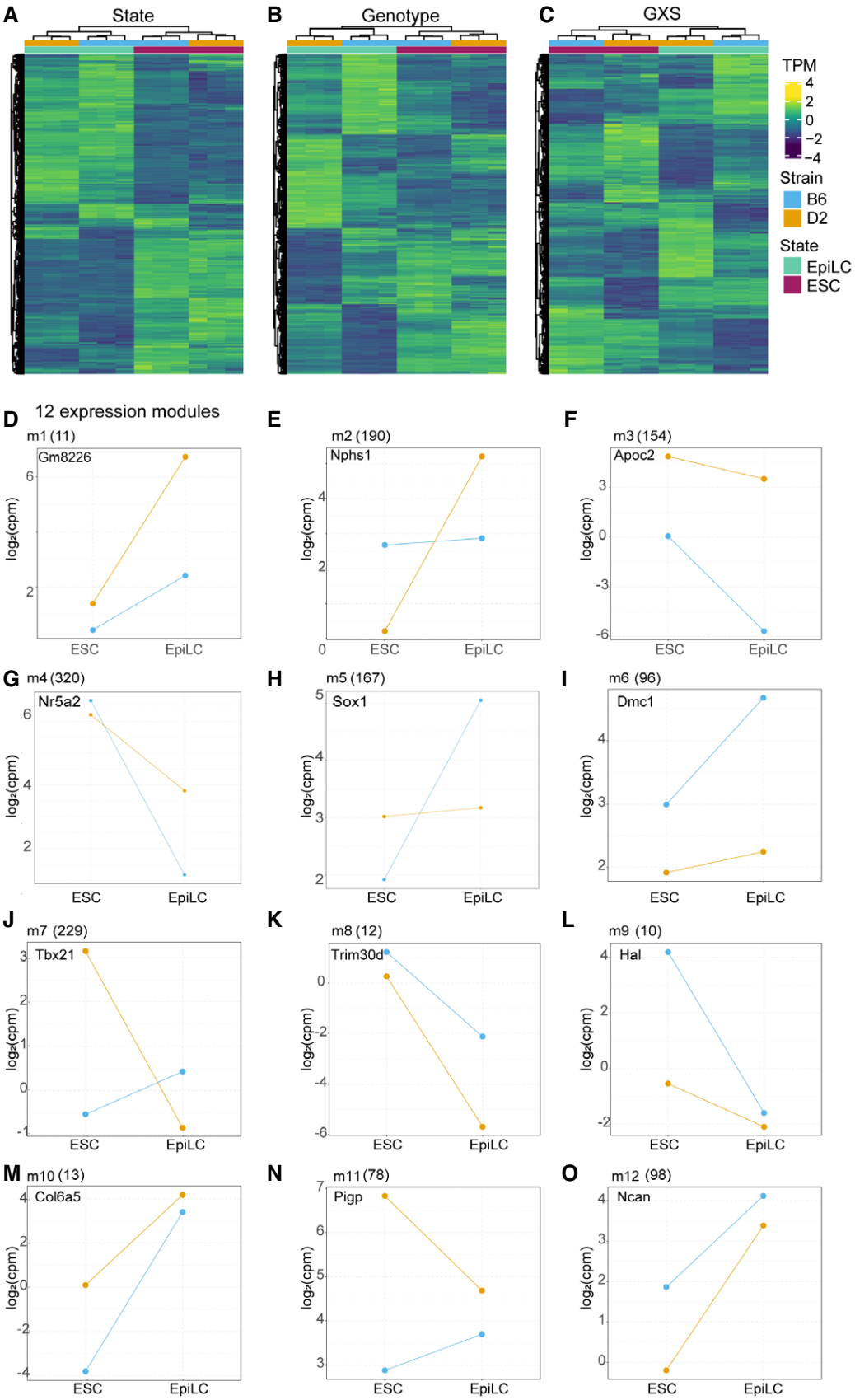

Figure EV2.

Figure EV3. Spontaneous differentiation of ESCs to EBs identifies divergent developmental trajectories between B6 and D2 strains.

A Experimental design to assess reproducibility of EB formation and scRNA-seq characterization of cell proportions. Reproducibility was accessed by growing (i) three biological replicates (originating from independently derived ESC lines), (ii) one ESC cell line grown for three technical replicates, and (iii) sampling replicates (pool of disaggregated EBs sampled three times). Each replicate was barcoded using lipid-modified oligonucleotides, pooled, and loaded onto a single 10X Chromium microfluidic lane.

B UMAP embeddings for all 10 scRNA-seq libraries colored by individual replicate. Each dot represents a single cell. Cells from replicate EBs clustered in proximity demonstrating reproducibility.

C Similar to (B), UMAP embeddings of cell clustering based on transcript abundance for all 10 samples.

D Number of cells per sample shown for each sample in (B).

E Proportion of cells within each cluster color coded by sample in (B). Overall, biological, technical, and sample replicates made up similar proportions of clusters. Because technical and sampling replicates were collected from D2 ESCs, downstream analysis for Fig 3 was performed with only 3 biological replicates for each genotype in order to allow comparisons between similar cell numbers.

F Violin plot of all 10 samples in (B) indicating the distribution of number of transcripts per sample.

G Violin plot indicating number of transcripts per cluster from 6 samples (3 biological replicates per strain) represented in UMAP shown in Fig 3B. Clusters 10 and 13 had fewer transcripts compared with all other clusters. Attempts at annotation of these two clusters mostly identified cellular processes instead of obvious cellular lineages.

H UMAP embeddings showing individual cells for clusters 4, 5, and 9 from Fig 3B. Trajectory of pseudotime analysis based on gene expression, using Slingshot, is overlaid on UMAP. Cluster 4 represents yolk sac blood island cells, which are common progenitors of vascular endothelium (cluster 5) and primitive erythrocytes (cluster 9).

I, J Pseudocolor plots for replicate FACS analysis on cells from EBs for both B6 and D2 strains. Cells from technical and biological replicates were labeled with anti-EpCAM (I) and anti-SOX1 antibody (J). Gated population indicates percentages of $EpCAM^+$ (Q3) or $SOX1^+$ cells. Replicate 1 is redisplayed here from Fig 3E to show with the other replicates.

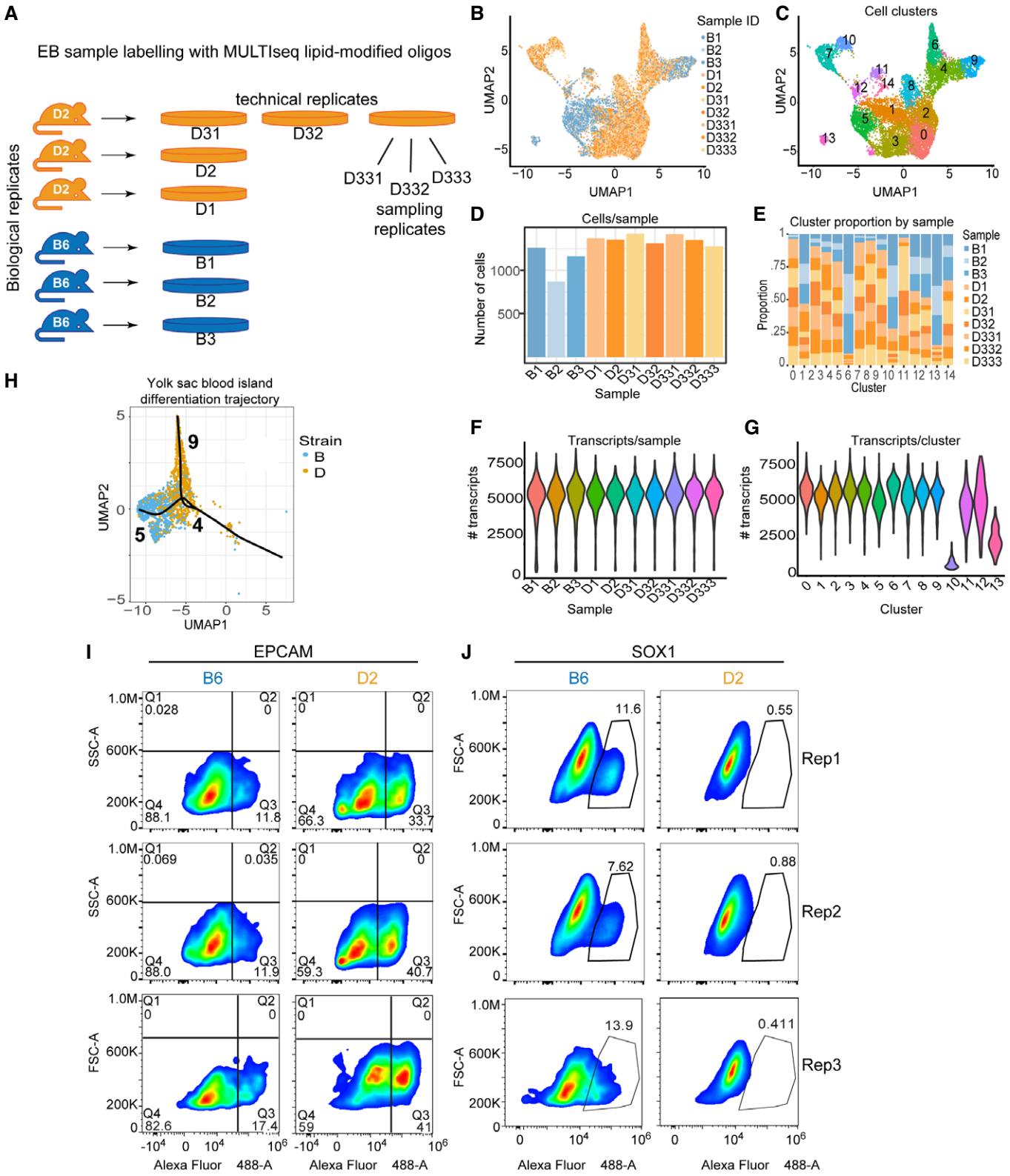

Figure EV3.

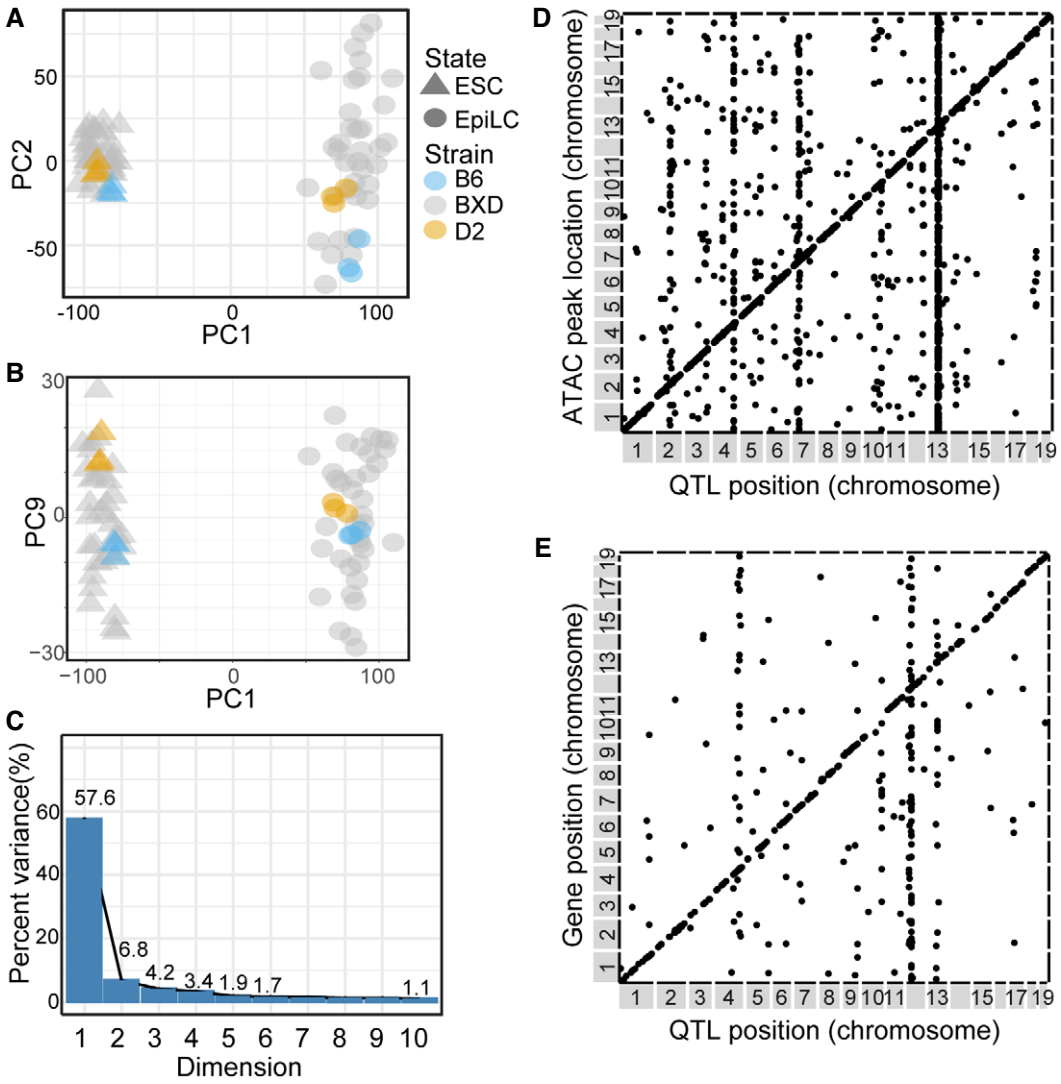

Figure EV4. QTL mapping chromatin accessibility and gene expression in EpiLCs.

A Scatterplot of PC1 and PC2 showing variance of total transcript abundance from RNA-seq for 33 distinct strains of BXDs and 3 biological replicates each for B6 and D2 strains in ESCs and EpiLCs.

B Similar to (A) showing variance in PC1 versus PC9.

C Scree plot shows variance explained for PC1-10. Visual inspection of PC2-9 (total variation = 22.2%) identified that, while biological replicates of parental strains tightly cluster, they are separated by variance captured by the PC, suggesting these PCs represent genetic contribution to transcript abundance within cell state. Individual BXDs can have transcriptomes with greater variation compared to the inbred parents, as is often the case in genetic populations with segregating alleles.

D Scatterplot of genomic locations of individual caQTL (x-axis, LOD > 5) versus location of ATAC peak being regulated (y-axis) for EpiLCs. Genetic effects that act locally (n = 4,463), i.e., in *cis*, fall on the diagonal line, while distal-acting QTL (n = 1,172) lie off the diagonal.

E Similar to (D), plotting eQTL position versus target gene location (n = 318 local-eQTL, n = 200 distal-eQTL).

Figure EV5. B6 haplotype at Chr 12 eQTL increases expression of genes found within the genotype-by-state neurogenesis module.

A Phenotype-by-genotype plots for all 19 Chr 12 eQTL targets that are found within the GxS neurogenesis module (m5). In each individual case, having the B6 haplotype at Chr 12 resulted in higher expression of the eQTL target gene, consistent with B6's differentiation propensity toward neuroectoderm (mean ± SEM).

B LOD plot for Chr 12 neurogenesis module eigengene. The eigengene is represented by the first principal component of the expression matrix for the genes indicated in panel (A). The modules maximum LOD score (5.63) maps to the proximal end of Chr 12 with a 1.5 LOD-drop confidence interval between 14753136 and 29439595, the same interval that the individual gene targets map to. Dashed lines represent empirical *P*-value thresholds based on 1,000 permutations (gray < 0.05, red < 0.01).

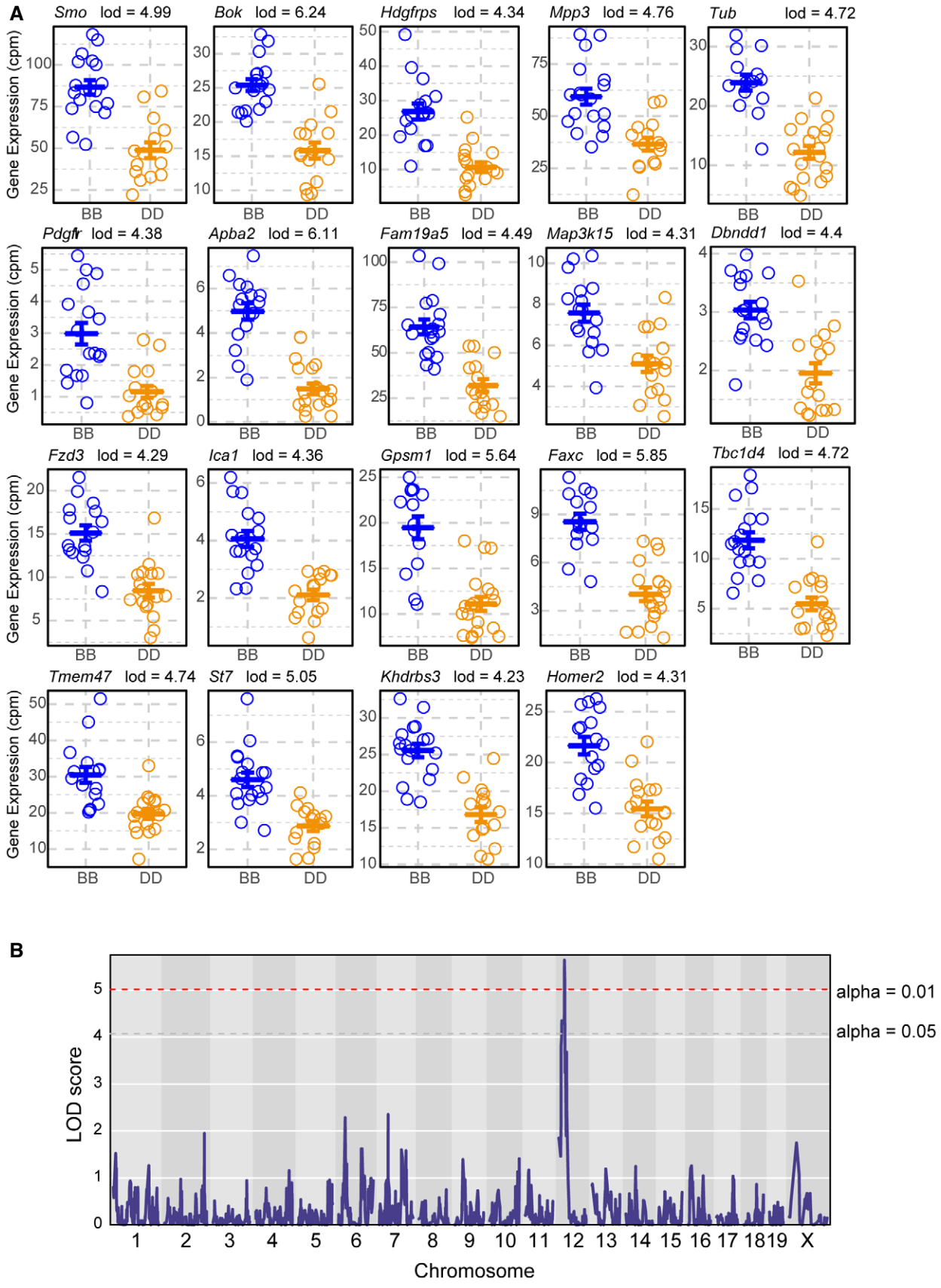

Figure EV5.