Cell Reports, Volume *42*

Supplemental information

ETV2 primes hematoendothelial gene enhancers prior to hematoendothelial fate commitment

Jeffrey D. Steimle, Chul Kim, Megan Rowton, Rangarajan D. Nadadur, Zhezhen Wang, Matthew Stocker, Andrew D. Hoffmann, Erika Hanson, Junghun Kweon, Tanvi Sinha, Kyunghee Choi, Brian L. Black, John M. Cunningham, Ivan P. Moskowitz, and Kohta Ikegami

Figure S1

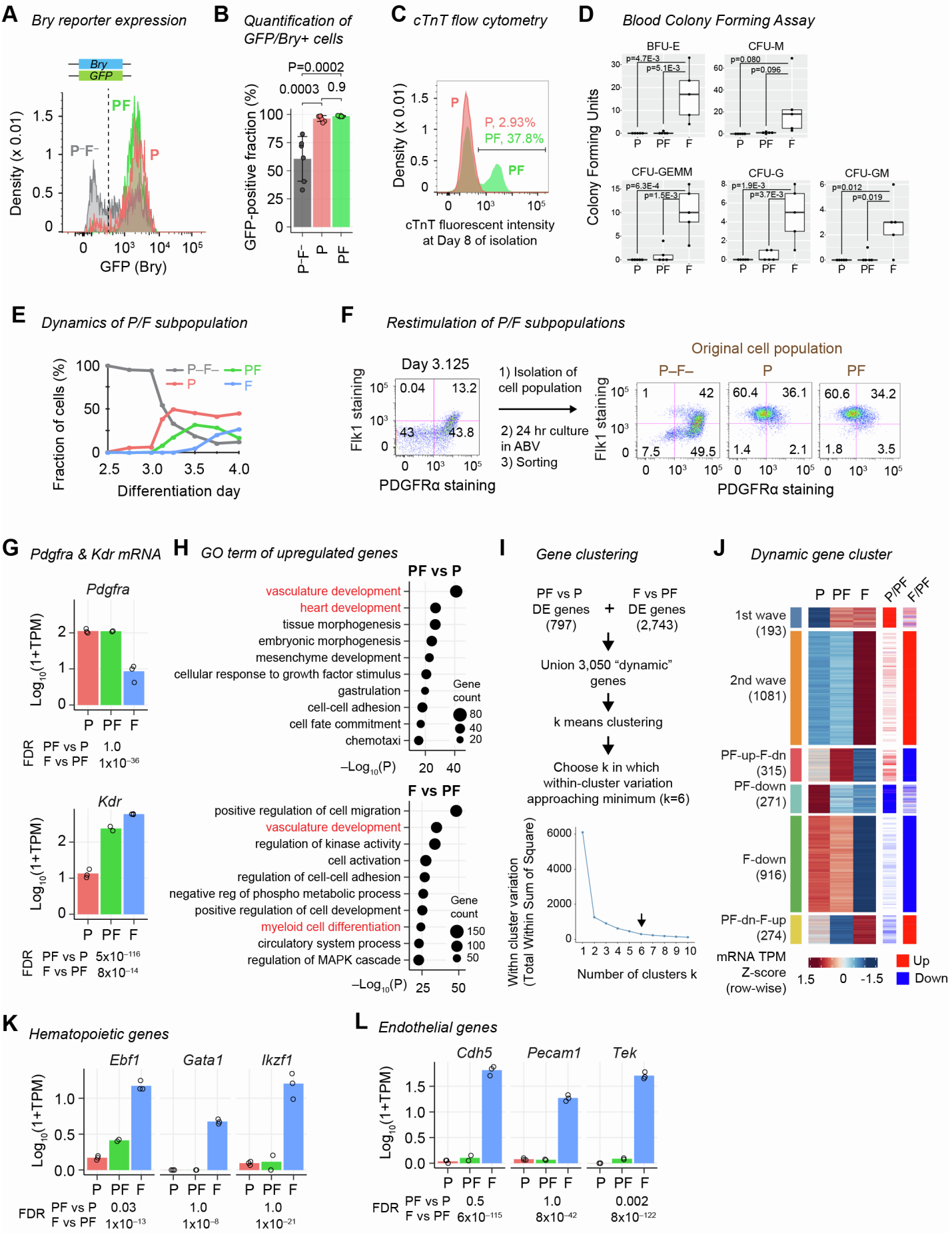

A
Bry reporter expression
Bry
GFP
PF
P–F–
P
Density (x 0.01)
GFP (Bry)
B
Quantification of GFP/Bry+ cells
P=0.0002
0.0003
0.9
GFP-positive fraction (%)
P–F–
P
PF
C
cTnT flow cytometry
P
P, 2.93%
PF, 37.8%
PF
Density (x 0.01)
cTnT fluorescent intensity at Day 8 of isolation
D
Blood Colony Forming Assay
BFU-E
CFU-M
CFU-GEMM
CFU-G
CFU-GM
Colony Forming Units
E
Dynamics of P/F subpopulation
P–F–
PF
P
F
Fraction of cells (%)
Differentiation day
F
Restimulation of P/F subpopulations
Day 3.125
Flk1 staining
PDGFRα staining
1) Isolation of cell population
2) 24 hr culture in ABV
3) Sorting
Original cell population
P–F–
P
PF
G
Pdgfra & Kdr mRNA
Pdgfra
Kdr
Log10(1+TPM)
FDR
PF vs P
F vs PF
H
GO term of upregulated genes
PF vs P
vasculature development
heart development
tissue morphogenesis
embryonic morphogenesis
mesenchyme development
cellular response to growth factor stimulus
gastrulation
cell-cell adhesion
cell fate commitment
chemotaxi
Gene count
–Log10(P)
F vs PF
positive regulation of cell migration
vasculature development
regulation of kinase activity
cell activation
regulation of cell-cell adhesion
negative reg of phospho metabolic process
positive regulation of cell development
myeloid cell differentiation
circulatory system process
regulation of MAPK cascade
I
Gene clustering
PF vs P DE genes (797)
F vs PF DE genes (2,743)
Union 3,050 “dynamic” genes
k means clustering
Choose k in which within-cluster variation approaching minimum (k=6)
Within cluster variation (Total Within Sum of Square)
Number of clusters k
J
Dynamic gene cluster
P/PF
F/PF
1st wave (193)
2nd wave (1081)
PF-up-F-dn (315)
PF-down (271)
F-down (916)
PF-dn-F-up (274)
mRNA TPM Z-score (row-wise)
Up
Down
K
Hematopoietic genes
Ebf1
Gata1
Ikzf1
L
Endothelial genes
Cdh5
Pecam1
Tek

Figure S1. Characterization of PDGFRα/FKL1 subpopulations and their gene expression (related to Figure 1)

(A) Histogram for GFP levels in P/F populations at Day 4 of differentiation of the transgenic mESC line with heterozygous GFP insertion into the *Brachyury* locus.

(B) Quantification of GFP/Brachyury-positive cell fraction in P/F populations (n=6). P, Generalized Linear Model p-value accounting for experiment batches.

(C) A representative flow cytometry plot for cTnT-stained cells at 8 days after isolation. See main **Fig. 1C** for quantification.

(D) Blood colony forming units at 10 days after isolation for individual blood subtypes. BFU-E, erythroid burst-forming unit. CFU-M, megakaryocyte forming unit. CFU-GEMM, granulocyte, erythrocyte, monocyte, megakaryocyte-forming unit. CFU-G, granulocyte-forming unit. CFU-GM, granulocyte, monocyte-forming unit. See main **Fig. 1D** for aggregate plot.

(E) PDGFRα/FKL1 subpopulation composition during differentiation. FACS plots are shown in main **Fig. 1E**.

(F) Re-stimulation of isolated PDGFRα/FKL1 subpopulations in the ABV regimen for 24 hours.

(G) RNA-seq TPMs for *Pdgfra* encoding PDGFRα and *Kdr* encoding FLK1.

(H) Gene Ontology (GO) terms associated with upregulated genes in *PF* over *P* (top) or in *F* over *PF* (bottom). Red, most relevant GO terms.

(I) Algorithm for clustering dynamic genes.

(J) Heatmap for all 3,050 dynamic genes grouped into 6 clusters.

(K) RNA-seq TPMs for representative hematopoietic genes.

(L) RNA-seq TPMs for representative endothelial genes.

Figure S2

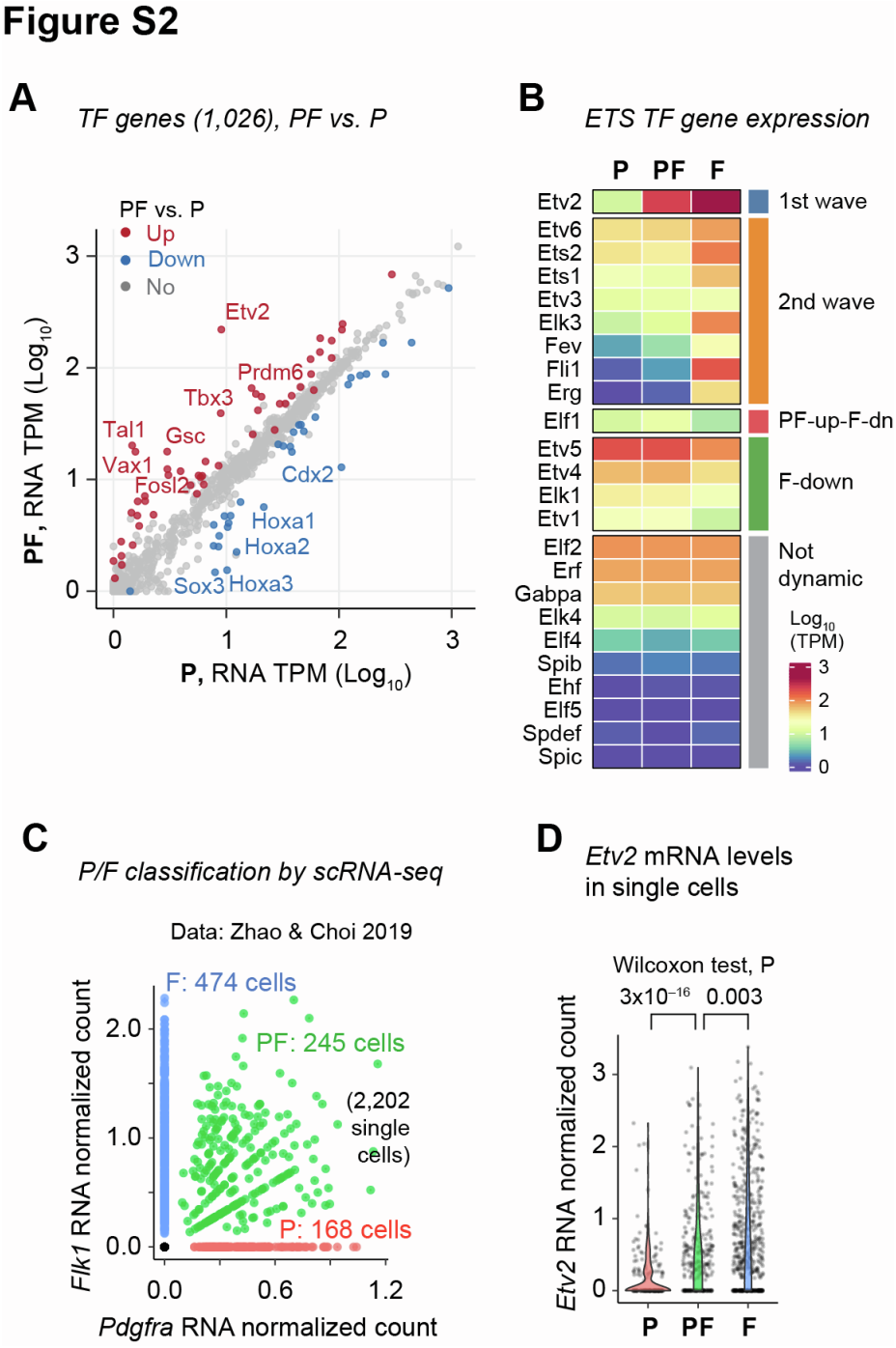


Figure S2. Characterization of *Etv2* expression (related to Figure 2)

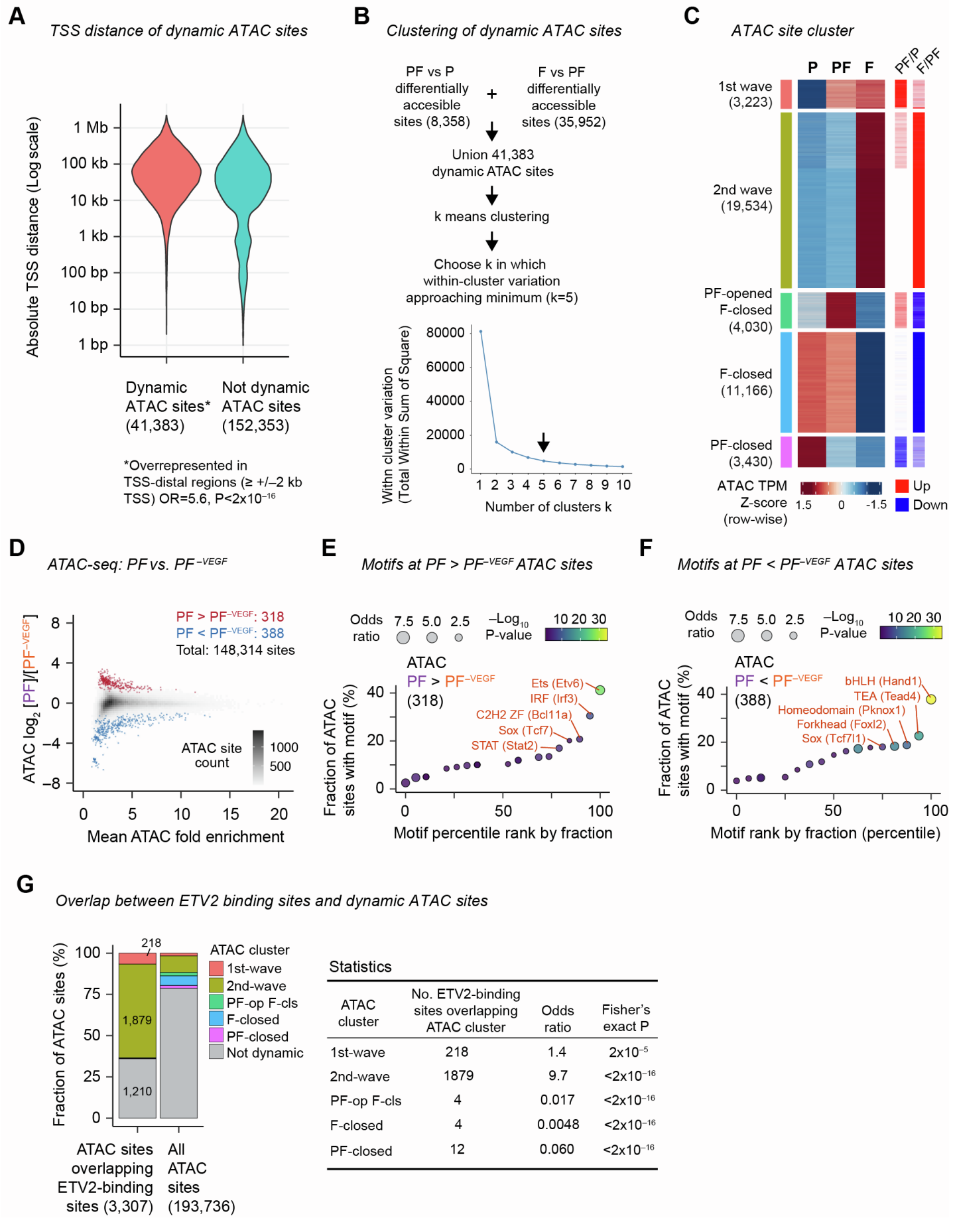
(A) RNA-seq TPMs for all 1,026 transcription factor (TF) genes in the *P* (x-axis) and the *PF* (y-axis) populations.

(B) RNA-seq TPMs for all 24 ETS TF genes.

(C) *Pdgfra* and *Flk1* (*Kdr*) mRNA levels (normalized read counts) of 2,202 single cells (Zhao & Choi, 2019). *P*, cells with *Pdgfra* count > 0 and *Flk1* count = 0. *PF*, cells with *Pdgfra* count > 0 and *Flk1* count > 0. *F*, cells with *Pdgfra* count = 0 and *Flk1* count > 0.

(D) *Etv2* expression levels of single cells categorized as *P*, *PF*, or *F* categorized in **Fig. S2C**.

Figure S3



Statistics

ATAC cluster	No. ETV2-binding sites overlapping ATAC cluster	Odds ratio	Fisher's exact P
1st-wave	218	1.4	$2x10^{-5}$
2nd-wave	1879	9.7	$<2x10^{-16}$
PF-op F-cls	4	0.017	$<2x10^{-16}$
F-closed	4	0.0048	$<2x10^{-16}$
PF-closed	12	0.060	$<2x10^{-16}$

Figure S3. Characterization of accessible chromatin in the P/F populations (related to Figure 3)

(A) Distribution of distance from dynamic or not-dynamic ATAC sites to the nearest transcription start site (TSS). Fisher's exact test examined the fraction of ATAC sites with +/–2 kb of TSS. OR, odds ratio.

(B) Procedure for clustering dynamic ATAC sites.

(C) Heatmap for all dynamic ATAC sites (41,383) grouped into 6 clusters.

(D) MA plot comparing ATAC-seq-derived chromatin accessibility at the 148,314 union ATAC sites in the *PF* population derived with VEGF (*PF*) vs. the *PF* population derived without VEGF (PF^{-VEGF}). The *PF* and PF^{-VEGF} cells are produced in parallel.

(E) TF family motifs overrepresented within the 314 ATAC sites that were more accessible in *PF* relative to PF^{-VEGF}.

(F) TF family motifs overrepresented within the 388 ATAC sites that were less accessible in *PF* relative to PF^{-VEGF}.

(G) (Left) Fraction of 3,307 ETV2-binding sites (pre-selected for those overlapping any ATAC sites) overlapping dynamic ATAC cluster sites. Fraction of all ATAC sites within the dynamic clusters is shown as comparison. (Right) Number of ETV2-binding sites overlapping dynamic ATAC cluster sites and statistical evaluation. PF-op F-cls denotes PF-opened F-closed sites.

FIGURE S4

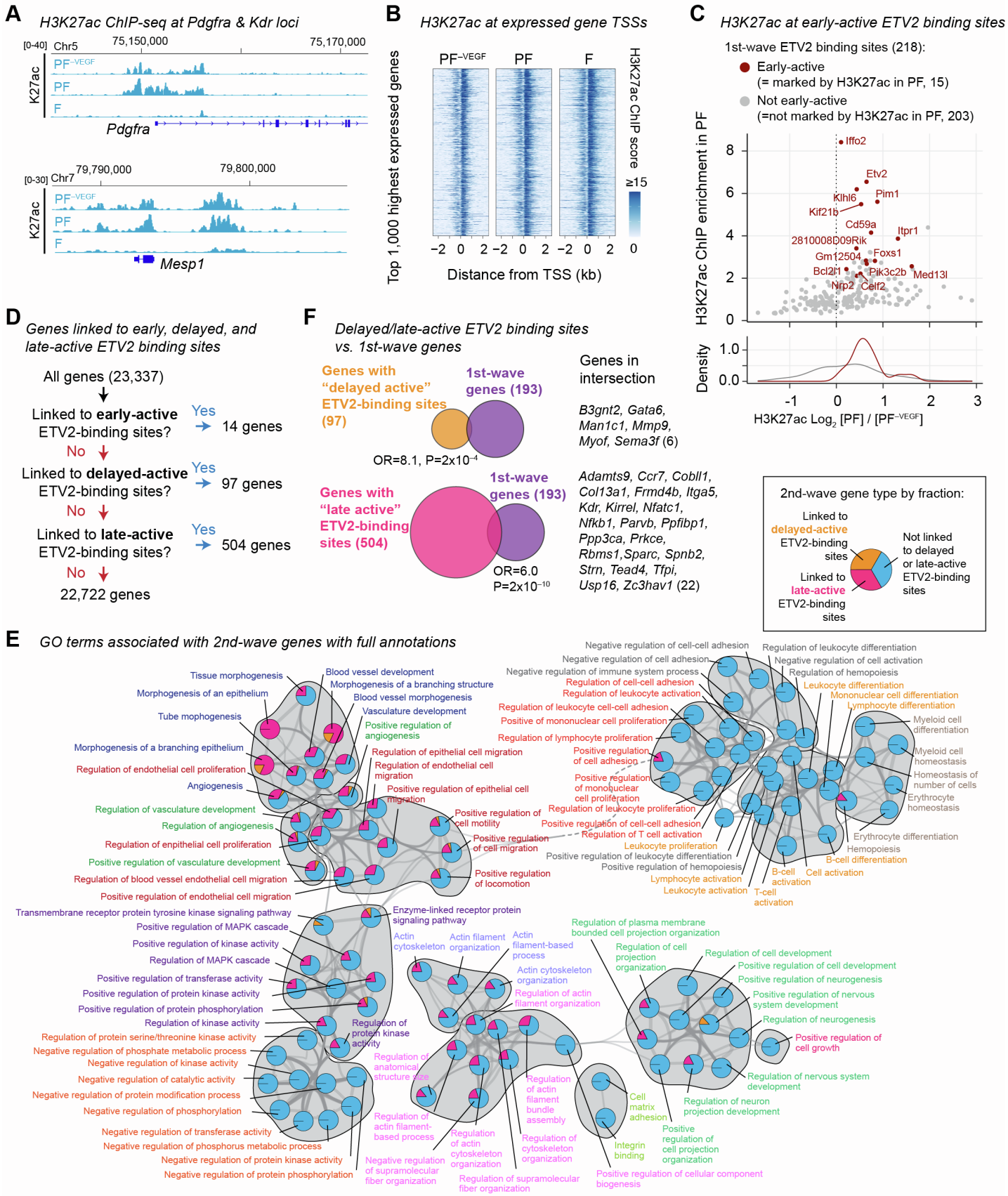


Figure S4. Histone H3K27ac states at ETV2-binding sites (related to Figure 4)

(A) H3K27ac ChIP-seq fold enrichment signal tracks at early mesoderm genes *Pdgfra* and *Mesp1*.

(B) H3K27ac ChIP-seq fold enrichment signals at 1,000 highest and consistently expressed gene TSSs.

(C) VEGF-dependency of H3K27ac levels at the early-active 1st-wave ETV2-binding sites (red) and other 1st-wave ETV2-binding sites (grey). (Top) Plot shows the fold difference of the H3K27ac level between *PF* and PF^{-VEGF} (x-axis) and the H3K27ac level in *PF* (y-axis). (Bottom) Histogram of data points along the x-axis.
(D) Classification of genes by linkage with early-active, delayed-active, and late-active ETV2-binding sites.
(E) Network visualization of GO terms overrepresented in the 2nd-wave genes. Same as **Fig. 4E** but all GO annotations are labeled. For each GO term (circle), the fraction of the GO-associated genes linked to delayed-active (yellow) or late-active (magenta) ETV2-binding sites or not linked to delayed or late-active ETV2-binding sites (blue) are shown as a pie chart.
(F) Relationship between the 1st-wave genes and the genes linked to delayed-active ETV2-binding sites (top) or the genes linked to late-active ETV2-binding sites. OR, odds ratio. Statistics, Fisher's exact test.