nature cardiovascular research

Article

https://doi.org/10.1038/s44161-025-00660-y

Pioneer factor ETV2 safeguards endothelial cell specification by recruiting the repressor REST to restrict alternative lineage commitment

Received: 14 August 2024

Accepted: 24 April 2025

Published online: 10 June 2025

Check for updates

Danyang Chen[1], Xiaonuo Fan[1], Ninghe Sun[2,3], Kai Wang[4,5], Liyan Gong[4,5], Juan M. Melero-Martin[4,5,6] & William T. Pu[1,6]

Mechanisms of cell fate specification are central to developmental biology and regenerative medicine. ETV2 is a master regulator for the endothelial cell (EC) lineage specification. Here we study mechanisms by which ETV2 overexpression in human induced pluripotent stem-cell-derived mesodermal progenitors efficiently specifies ECs. We used CUT&RUN, scRNA-seq and scATAC-seq to characterize the molecular features of EC differentiation mediated by ETV2. We defined the scope of ETV2 pioneering activity and identified its direct downstream target genes. Induced ETV2 expression both directed specification of endothelial progenitors and suppressed acquisition of alternative fates. Functional screening and candidate validation revealed cofactors essential for efficient EC specification, including the transcriptional activator GABPA. Notably, the transcriptional repressor REST was also necessary for efficient EC specification. ETV2 recruited REST to repress non-EC lineage genes. Our study provides an unparalleled molecular analysis of EC specification at single-cell resolution and highlights the important role of pioneer factors to recruit repressors that suppress commitment to alternative lineages.

During mammalian development, a single multipotent progenitor forms diverse cell lineages. Elucidating the mechanisms responsible for cell lineage specification and differentiation is a fundamental goal of developmental biology, and a better understanding of these mechanisms will have direct applications in regenerative medicine. We previously reported that expression of ETV2, a master regulator of endothelial cell (EC) specification, directs mesoderm progenitor cells (MPCs) to rapidly and efficiently differentiate into ECs[1]. Here we further optimized this powerful lineage specification model system and deployed it to dissect mechanisms by which a master transcriptional regulator programs progenitor cells to specify progeny lineages.

DNA occupancy by nucleosomes is a key epigenetic barrier that constrains cells to a restricted set of transcriptional states and lineages[2]. Pioneer transcription factors (TFs) can occupy binding sites within nucleosomal DNA and make it accessible for binding by non-pioneer factors[3]. Pioneer TF binding sites are in part determined by pioneer TF interaction with non-pioneer TFs[4–7]. The ability of pioneer TFs to make previously closed chromatin accessible for binding by non-pioneer TFs makes them central to lineage specification, as they allow cells to

[1]Department of Cardiology, Boston Children's Hospital, Boston, MA, USA. [2]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA. [3]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA. [4]Department of Cardiac Surgery, Boston Children's Hospital, Boston, MA, USA. [5]Department of Surgery, Harvard Medical School, Boston, MA, USA. [6]Harvard Stem Cell Institute, Cambridge, MA, USA. e-mail: william.pu@cardio.chboston.org

surmount epigenetic constraints to differentiate into additional cell types[8]. ETV2 is a pioneer TF that is critical for EC specification[9–11]. This ETS-family TF is transiently expressed in mesodermal EC progenitors from embryonic day (E) 7.5–E9.5 of mouse development[12]. Mice lacking ETV2 fail to develop blood or vasculature[13,14] and induced pluripotent stem (iPS) cells lacking ETV2 do not form ECs[1]. Forced expression of ETV2 directs iPS cells, fibroblasts and MPCs to differentiate into ECs[1,15,16]. Among ETV2's identified target genes are hallmark EC genes *Tie2*, *Lmo2* and *Fli1* (refs. 10,14,17). Despite this large body of data, the mechanisms by which ETV2 reprograms cells into ECs remain unclear.

Here we developed a highly efficient and rapid iPS cell-based system to study ETV2-driven EC specification and differentiation from MPCs. We used this system with single-cell approaches to determine ETV2's impact on MPC differentiation trajectories. We identified enhancers and genes regulated by ETV2 pioneering activity and uncovered TFs that collaborate with ETV2 to direct EC specification. We discovered that ETV2 recruits REST, a transcriptional repressor, to silence non-EC lineage genes, enhancing EC specification by blocking alternative fates. Collectively, our study provides molecular insights into the mechanisms by which ETV2 drives EC specification.

Results

Efficient ETV2-directed differentiation of MPCs to ECs

We first established a system to rapidly, efficiently and reproducibly induce iPS cell-derived MPCs to differentiate into ECs. We previously showed that modified messenger RNA-mediated expression of ETV2 in iPS cell-derived MPCs efficiently directs their differentiation to iPS cell-derived ECs[1]. To increase the reproducibility and flexibility of programmed ETV2 expression, we engineered a stable clonal iPS cell line, TRE3G-ETV2, which inducibly expressed ETV2 upon treatment with doxycycline (Dox) (Fig. 1a). Treatment of iPS cells with Wnt activator CHIR99021 for 2 days induced formation of MPCs. On day 2 (D2), CHIR99021 was removed, and cells were cultured for two additional days in the presence of angiogenic growth factors (VEGFA, FGF2 and EGF) and the TGFβ inhibitor SB431542. Alternative CHIR99021 or Dox treatments that we tested were less efficient (Extended Data Fig. 1a–f). Varying the dose of Dox revealed a steeply nonlinear response consistent with the dose–response of reverse tet-activator systems[18] (Extended Data Fig. 1g,h). These results yielded an optimized 4-day protocol in which ETV2 induction by 1 μg ml^{-1} Dox from D2–D3 resulted in highly efficient MPC-to-EC differentiation by D4. In the absence of exogenous ETV2 induction (Dox$^-$), this protocol yielded ~30% CDH5 and PECAM1 double-positive iPS cell-derived ECs (Fig. 1a,b and Extended Data Fig. 1a–c). Exogenous ETV2, induced by Dox (Dox$^+$) from D2–D3 efficiently generated iPS cell-derived ECs (93% CDH5$^+$PECAM1$^+$) (Fig. 1a,b). ETV2 immunostaining and FACS showed that 1 μg ml^{-1} Dox induced ETV2 expression in ~87% of D3 Dox$^+$ cells, whereas only ~30% of the Dox$^-$ cells express VEGFA-induced endogenous ETV2 on day 3.5 (D3.5 Dox$^-$). Among ETV2-expressing cells, ETV2 protein level per cell was on average 2.3–3-fold higher in D3 Dox$^+$ compared to D3.5 Dox$^-$ conditions (Extended Data Fig. 1i,j).

The iPS cell-derived ECs generated using this optimized protocol exhibited expected EC phenotypes: they expressed EC markers (Extended Data Fig. 2a), aligned in response to flow (Extended Data Fig. 2b), migrated in the wound scratch assay (Extended Data Fig. 2c), produced nitric oxide (Extended Data Fig. 2d), took up acetylated low-density lipoprotein (AcLDL; Extended Data Fig. 2e) and upregulated *ICAM1*, *VCAM1* and *SELE* in response to the inflammatory cytokine tumor necrosis factor (Extended Data Fig. 2f). Using RNA-seq, we compared the transcriptomes of the generated iPS cell-derived ECs to primary cells isolated from human tissues[19]. The iPS cell-derived ECs grouped with ECs and away from non-ECs (Extended Data Fig. 2g) and had the greatest transcriptional similarity with arterial ECs (aECs). These studies confirmed that ETV2-directed, iPS cell-derived ECs possess functional properties of ECs in vitro and are transcriptionally most similar to aECs.

scRNA-seq analysis of iPS cell-derived EC differentiation

To better characterize the products of control (Dox$^-$) and Dox$^+$ differentiation, we acquired high-quality single-cell transcriptomes from 25,502 Dox$^-$ and Dox$^+$ cells at differentiation D4 (Fig. 1c). The majority of Dox$^+$ and Dox$^-$ cells at D4 were transcriptionally distinct (Fig. 1c). Dox$^-$ cells predominantly clustered in subpopulations with signatures of venous EC (vEC; *NR2F2*) and cardiac progenitors (*HAND2*) (Fig. 1d,e). In contrast, most Dox$^+$ cells contributed to two large, related clusters (aEC-1 and aEC-2) that highly expressed pan-EC markers (*CDH5*) and aEC genes such as *GJA4* (Fig. 1d,e). Gene Ontology (GO) enrichment analysis of the marker genes of each cluster further revealed their biological function and cell identity (Extended Data Fig. 2h). These results show that Dox$^+$ and Dox$^-$ differentiations yielded distinct and largely non-overlapping cell types, with Dox cells nearly exclusively forming aECs and control cells predominantly forming vECs and cardiac progenitors.

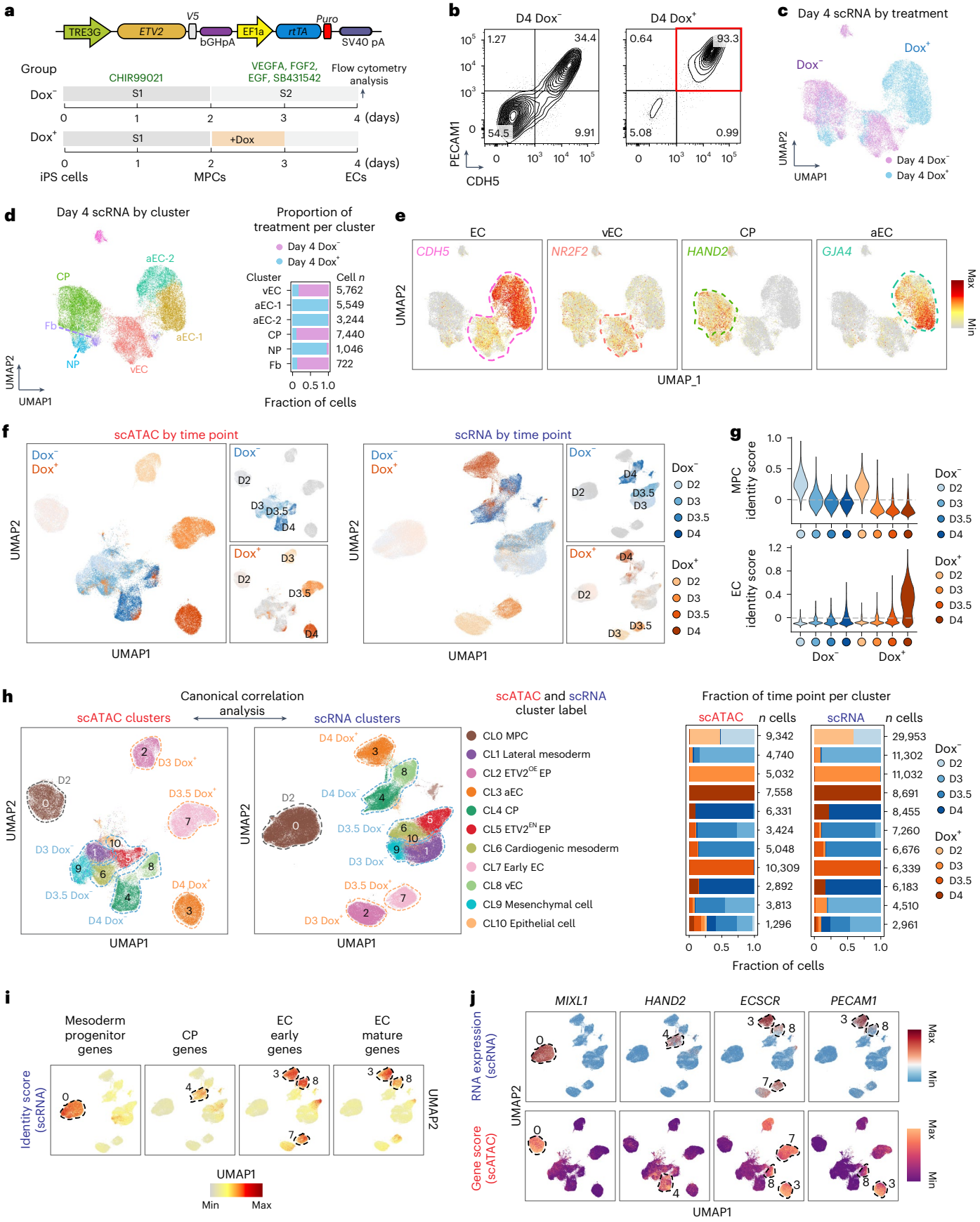
To control for the effect of Dox and the TRE3G-ETV2 transgene, we treated wild-type and TRE3G-ETV2 cells with different Dox concentrations (Extended Data Fig. 2i,j). In the absence of Dox, wild-type cells (0 μg ml^{-1}) and TRE3G-ETV2 cells (0 μg ml^{-1}) clustered similarly to each other, excluding the effect of the transgene on differentiation. Wild-type cells treated with 0, 0.1 or 1.0 μg ml^{-1} Dox clustered similarly to each other, indicating that Dox itself does not affect EC differentiation. Meanwhile, TRE3G-ETV2 cells treated with 0.1, 0.5 or 1.0 μg ml^{-1} Dox clustered similarly and separate from wild-type cells. These results show that the Dox dose within this range does not influence the types of cells generated.

Single-cell analysis of iPS cell-derived EC differentiation dynamics

To assess chromatin and gene expression dynamics during EC specification, we performed separate single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq) and single-cell RNA sequencing (scRNA-seq) at D2, D3, D3.5 and D4 from the Dox$^+$ and Dox$^-$ differentiation conditions in biological duplicate. After stringent filtering and quality control, we obtained 59,785 scATAC-seq and 105,563 scRNA-seq cells with high quality (Extended Data Fig. 3). We performed integrative analysis of the Dox$^+$ and Dox$^-$ cells at the four time points for the scRNA-seq and scATAC-seq datasets (Fig. 1f). The D2 cells, representing mesoderm progenitor cells, from Dox$^+$ and Dox$^-$ time courses grouped together as expected, suggesting minimal batch effects. In contrast, Dox$^+$ cells at each subsequent time point largely

Fig. 1 | Integrated single-cell transcriptomic and epigenomic analysis of EC differentiation. a, Overview of the TRE3G-ETV2 system (top) and timeline of 4-day EC differentiation protocol showing timing of Dox addition (1 μg ml^{-1}). Cells were treated without Dox (Dox$^-$) or with Dox (Dox$^+$) for 24 h. **b**, Flow cytometry analysis for PECAM1 and CDH5 expression at D4 of differentiation. **c**, Cells were collected from Dox$^-$ and Dox$^+$ groups on D4 for scRNA-seq. scRNA-seq data were displayed as a UMAP projection and labeled by group. **d**, scRNA-seq data labeled by inferred cell types. Bar plots show the proportion of cells from each treatment across the clusters. CP, cardiac progenitor; Fb, fibroblast; NP, nephron progenitor. **e**, Feature plots of selected markers of each cluster. **f**, Time course of scATAC-seq and scRNA-seq data from D2, D3, D3.5 and D4 projected onto UMAP space showing days of differentiation of both Dox$^-$ and Dox$^+$ cells. Cells are colored by sample collection time. **g**, Cell identity scores of mesoderm progenitor and EC gene sets within each time point. **h**, UMAP from scATAC-seq and scRNA-seq colored by identified cell clusters (left). Biological classification of the scRNA-seq and inferred scATAC-seq clusters (middle). Contribution of each group to each scATAC-seq and scRNA-seq cluster (right). **i**, Feature plots showing gene-set identity scores. Each cell's value is shown, mapped onto the scRNA-seq UMAP plot. **j**, Feature plots showing gene expression (top row) and ATAC gene scores (bottom row) of cell identity markers. Each cell's value is shown, mapped onto either the scRNA-seq UMAP plot (top row) or the scATAC-seq UMAP plot (bottom row). ETV2^{OE}, ETV2 overexpressing; ETV2^{EN}, endogenous ETV2-expressing.

https://doi.org/10.1038/s44161-025-00660-y



https://doi.org/10.1038/s44161-025-00660-y

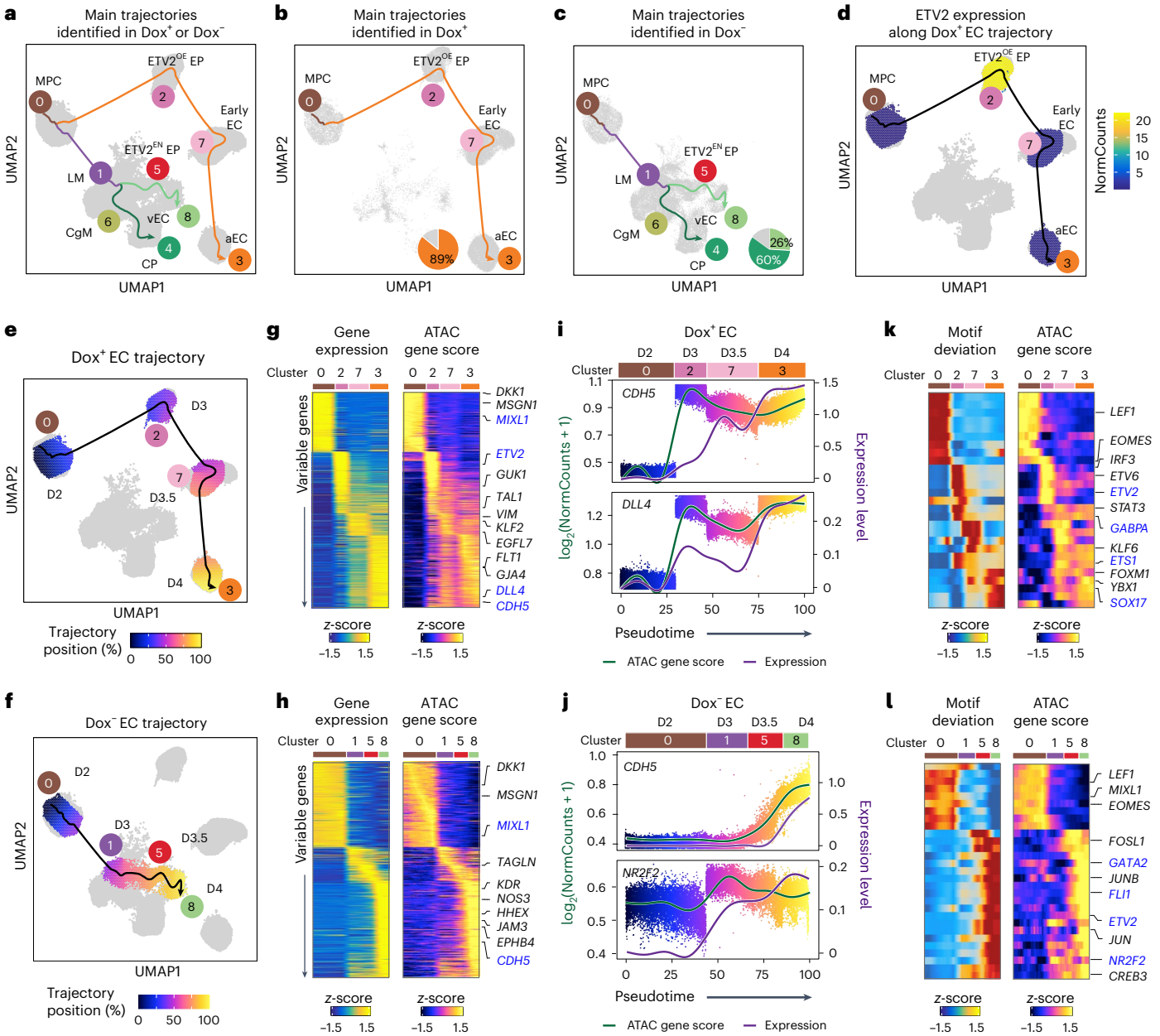


Fig. 2 | Reconstruction of EC differentiation trajectories with and without ETV2 overexpression. a–**c**, UMAP of scATAC-seq cells highlighting the dominant trajectories (**a**) identified in Dox$^+$ (**b**) and Dox$^-$ (**c**) groups. The cell clusters correspond to those in Fig. 1h. Pie charts in **b** and **c** show the percentage of cells on the trajectory in the indicated cluster at D4. **d**, Single-cell *ETV2* expression across pseudotime in the Dox$^+$ EC trajectory. **e**,**f**, UMAP visualization of the trajectory of EC differentiation with (**e**) or without (**f**) ETV2 overexpression. Each cell is colored by the pseudotime and each cluster is annotated by sample collection time. The smoothed arrow represents a visualization of the interpreted trajectory in the UMAP embedding. **g**,**h**, Heatmaps showing scRNA-seq gene expression (left) and scATAC-seq gene score (right) of genes with correlated variable activity in cells identified to be in the differentiation trajectory, as ordered by pseudotime. Cell clusters are labeled according to their position along the pseudotime. **i**,**j**, Gene expression and ATAC gene score dynamics of *CDH5* and *DLL4* in the Dox$^+$ EC trajectory (**i**) or *CDH5* and *NR2F2* gene in Dox$^-$ EC trajectory (**j**) across pseudotime. **k**,**l**, Heatmap of TF regulators for which ATAC gene score is positively correlated with chromVAR TF deviation (motif deviation) across the Dox$^+$ EC trajectory (**k**) or Dox$^-$ EC trajectory (**l**). Cell clusters are labeled according to their position along the pseudotime.

clustered in a different uniform manifold approximation and projection (UMAP) space than Dox$^-$ cells (Fig. 1f). This indicates that ETV2 guides EC specification by reshaping the gene expression and genomic accessibility landscape.

To investigate the differentiation dynamics, we analyzed gene expression at each time point (Extended Data Fig. 4a). Activation of Wnt by CHIR99021 differentiated iPS cells into *MIXL1*$^+$ and *DLL3*$^+$ mesoderm at D2, while suppressing ectoderm and endoderm formation (Extended Data Fig. 4a). Composite cell identity scores demonstrated loss of MPC identity by D3 in both Dox$^+$ and Dox$^-$ conditions (Fig. 1g, top). EC identity progressively increased thereafter (Fig. 1g, bottom). Early EC markers including *KLF2*, *KDR*, *ECSCR* and *HHEX* were acquired by Dox$^+$ cells by D3 (Extended Data Fig. 4a). Subsequently, Dox$^+$ cells expressed mature EC markers including *PECAM1*, *FLT1*, *ESAM* and *JAM3* at D3.5 and D4. At D4, Dox$^+$ cells expressed aEC markers (*GJA4*). Dox cells did not commit to hematopoietic lineage (Extended Data Fig. 4a), consistent with other studies showing that ETV2 is not sufficient to promote hematopoietic lineage specification[15,16,20]. The shift toward EC identity was only weakly apparent in Dox$^-$ cells (Fig. 1g, bottom). Neither Dox$^+$ nor Dox$^-$ cells expressed capillary marker genes (Extended

Data Fig. 4a). The difference between Dox$^+$ and Dox$^-$ cells was further corroborated by analysis of the scATAC-seq signal at *MIXL1*, *ESAM* and *GJA4* (Extended Data Fig. 4b–d).

Unsupervised clustering of the combined scRNA-seq datasets defined 11 distinct clusters (CL0–CL10) (Fig. 1h and Supplementary Tables 1 and 2). To integrate the chromatin accessibility and transcriptomic profiles into a unified cell-type-resolved regulatory atlas, we performed canonical correlation analysis[21] to match scATAC-seq cells with their nearest neighbor cells in the annotated scRNA-seq data. Most cells in scATAC-seq clusters mapped to a single corresponding scRNA-seq cluster (Extended Data Fig. 4e), and corresponding scRNA-seq and scATAC-seq clusters comprised concordant proportions across cell groups and time points (Fig. 1h; Fraction of time point per cluster). Dox$^+$ cells overlapped with Dox$^-$ cells at D2 and subsequently diverged, forming a single dominant cluster at each subsequent time point (Fig. 1h): at D3, CL2 (ETV2 overexpressing endothelial progenitor (EP), hereafter refer as ETV2^{OE} EP); at D3.5, CL7 (early EC); and at D4, CL3 (aEC). In contrast, the Dox$^-$ cells had higher cellular heterogeneity: at D3, CL1 (lateral mesoderm (LM)) and CL9 (mesenchymal cell (MC)); at D3.5, CL5 (endogenous ETV2-expressing EP, refer as ETV2^{EN} EP) and CL6 (cardiogenic mesoderm (CgM)); and at D4, CL4 (cardiac progenitor) and CL8 (vEC).

Consistent with their assigned labels, identity scores calculated for MPC, cardiac progenitor and EC gene sets selectively highlighted CL0, CL4, and CL3, CL7 and CL8, respectively (Fig. 1i). Known signature genes for each cell population, including *MIXL1* for mesodermal progenitors, *HAND2* for cardiac progenitors, *ECSCR* for early ECs identity and *PECAM1* for mature EC identity, were expressed most highly in the respective clusters (Fig. 1j, top row). The scATAC-seq gene accessibility scores of these signature marker genes were concordant with marker expression (Fig. 1j, bottom row). Based on the correlation between chromatin accessibility and gene expression, we identified 77,679 significant peak-to-gene linkages including promoter and putative distal enhancers across differentiation (Extended Data Fig. 4f and Supplementary Table 3).

Collectively, these results revealed transcriptomic and epigenomic signatures of MPC differentiation to ECs in native (Dox$^-$) and ETV2^{OE} (Dox$^+$) conditions. ETV2^{OE}-induced differentiation yielded distinct and more homogeneous cell states compared to native differentiation.

ETV2^{OE} and ETV2^{EN} EC differentiation trajectories

To investigate the cell differentiation dynamics of MPCs (CL0) with or without ETV2 overexpression, we reconstructed cell trajectories in Dox$^+$ and Dox$^-$ groups (Fig. 2a). Under Dox$^+$ treatment, MPCs followed a single major trajectory to differentiate into ETV2^{OE} EPs (CL2; D3), then into early ECs (CL7; D3.5), and finally into aECs (CL3; D4; Fig. 2b,d) with high efficiency (89%). Under Dox$^-$ conditions, MPCs primarily differentiated into lateral mesoderm (CL1; D3; Fig. 2a,c). Trajectories then bifurcated, with one terminating on cardiac progenitors (CL4; D4), and the other on vECs (CL8; D4). Based on the relative proportion of cells in vECs versus cardiac progenitor clusters at D4, most Dox$^-$ cells (60%) followed the cardiac rather than the EC trajectory, and as a result EC differentiation was inefficient in Dox$^-$ cells (Fig. 2c).

We calculated the pseudotime position of each cell along the Dox$^+$ or Dox$^-$ EC trajectories (Fig. 2e,f) and used this to analyze the dynamics of gene expression and chromatin accessibility during the differentiation process (Fig. 2g,h). In both trajectories, the expression of the mesoderm marker *MIXL1* decreased with pseudotime, as did its ATAC gene score, a metric designed to infer expression from scATAC-seq data (Methods). In the Dox$^+$ EC trajectory, *ETV2* peaked at D3, followed by subsequent upregulation of pan-EC gene *CDH5* and then aEC gene *DLL4* (ref. 22) (Fig. 2g,i and Extended Data Fig. 5a). Compared to the Dox$^+$ EC trajectory, *ETV2* was only weakly upregulated at D3.5 in the Dox$^-$ EC trajectory that terminated on vECs (Extended Data Fig. 5a), and *CDH5* upregulation occurred more weakly and later in pseudotime (Fig. 2h,j). The peak level of ETV2 expression along the pseudotime was 3.6-times higher in Dox$^+$ compared to Dox$^-$ conditions (Extended Data Fig. 5a), consistent with our measurements of ETV2 protein levels (Extended Data Fig. 1i,j). The Dox$^-$ trajectory was also characterized by upregulation of *NR2F2*, a TF important for vEC differentiation[23] (Fig. 2j). A parallel analysis of the Dox$^-$ trajectory that terminated on cardiac progenitors (Extended Data Fig. 5b) demonstrated upregulation of cardiac progenitor genes (*MEF2C* and *HAND2*) at late pseudotimes, and no induction of *ETV2* and *CDH5* (Extended Data Fig. 5c). Notably, across these analyses, chromatin accessibility changes (ATAC gene score) generally preceded gene expression changes in pseudotime, suggesting that activation of *cis*-regulatory elements (CREs) was followed by later gene upregulation (Fig. 2i,j).

To gain insights into TFs driving gene expression and chromatin accessibility changes along these trajectories, we identified TFs with a strongly correlated ATAC gene score and chromVAR motif deviation[24], a method for estimating accessibility within peaks containing the motif (Methods), across the Dox$^+$ (Fig. 2k) or Dox$^-$ (Fig. 2l and Extended Data Fig. 5d) trajectories. In the Dox$^+$ EC trajectory, ETV2 motif activity and ATAC gene score sharply increased in ETV2^{OE} EP and declined subsequently (Fig. 2k). The motif deviation of several other ETS-family TFs including *GABPA* and *ETS1* showed similar patterns. The motif accessibility and ATAC gene score of *SOX17*, an aEC-associated TF[25], rapidly increased in cells belonging to the aEC cluster (CL3). In the Dox$^-$ EC trajectory, the signature of activation of *ETV2* and other ETS-family TFs (*FLI1*) was observed at late pseudotimes. The GATA-family TF *GATA2* demonstrated increased motif activity and ATAC gene score at later pseudotimes, along with a signature for increased activity of the vEC TF *NR2F2* (Fig. 2l). In the Dox$^-$ cardiac progenitor trajectory, at late pseudotimes we observed increased motif accessibility and ATAC gene score for cardiac progenitor markers, such as *HAND2* and *MEF2C* (Extended Data Fig. 5c,d).

Together, these analyses show that under Dox$^-$ conditions, most cells followed a trajectory to other mesoderm fates including cardiac progenitors. Relatively fewer MPCs differentiated into EPs and finally to ECs with venous features. In contrast, ETV2 overexpression suppressed alternative fates and drove highly efficient differentiation to EPs, which subsequently differentiated into ECs with arterial features.

TF regulatory networks that orchestrate EC differentiation

To identify the cell-type-resolved TFs that orchestrate EC differentiation, we analyzed the scATAC-seq data to identify differentially accessible peaks and the TF motifs enriched among differential peaks in each cell cluster (Fig. 3a). We further used scRNA-seq data to perform single-cell gene regulatory network analysis using SCENIC[26] (Fig. 3b).

Motifs enriched in a cluster's differentially accessible peaks identified candidate TFs that regulate gene expression and chromatin accessibility in the cluster (Fig. 3c and Supplementary Table 4). For example, the binding motifs for *EOMES* and *LEF1*, well established to regulate mesoderm formation[27,28], were overrepresented in MPC (CL0; Fig. 3c). ETV2^{OE} EP (CL2) had a distinct TF motif profile characterized by enrichment in ETS-family binding motifs, including *ETV2* and *GABPA*, as well as motifs for STAT family members (Fig. 3c). The ETV2^{EN} EP cluster in Dox$^-$ cells (CL5) and the early EC cluster in Dox$^+$ cells (CL7) shared motif enrichment patterns. These clusters lacked ETS-family motifs and instead were highly enriched for AP-1 TFs such as *JUN* and *FOSL1* and were moderately enriched for GATA-family motifs. The motifs for *FOXO1* and *SOX17*, established aEC TFs[25,29], were enriched in Dox$^+$ aEC (CL3) (Fig. 3c).

Complementary to the motif analysis, analysis of the expression of TFs and their predicted downstream genes by SCENIC identified top TF regulons within each cluster (Fig. 3d and Supplementary Table 5). Top-ranked regulons included *ETV2*, *GABPA*, *REST* and *STAT6* in ETV2^{OE} EPs (CL2); *JUN* and *CREB5* in early ECs (CL7) and *JUNB* in ETV2^{EN} EPs (CL5); *FOXO1*, *KLF6* and *SOX17* in aECs (CL3); and *NR2F2* in vECs (CL8)

https://doi.org/10.1038/s44161-025-00660-y

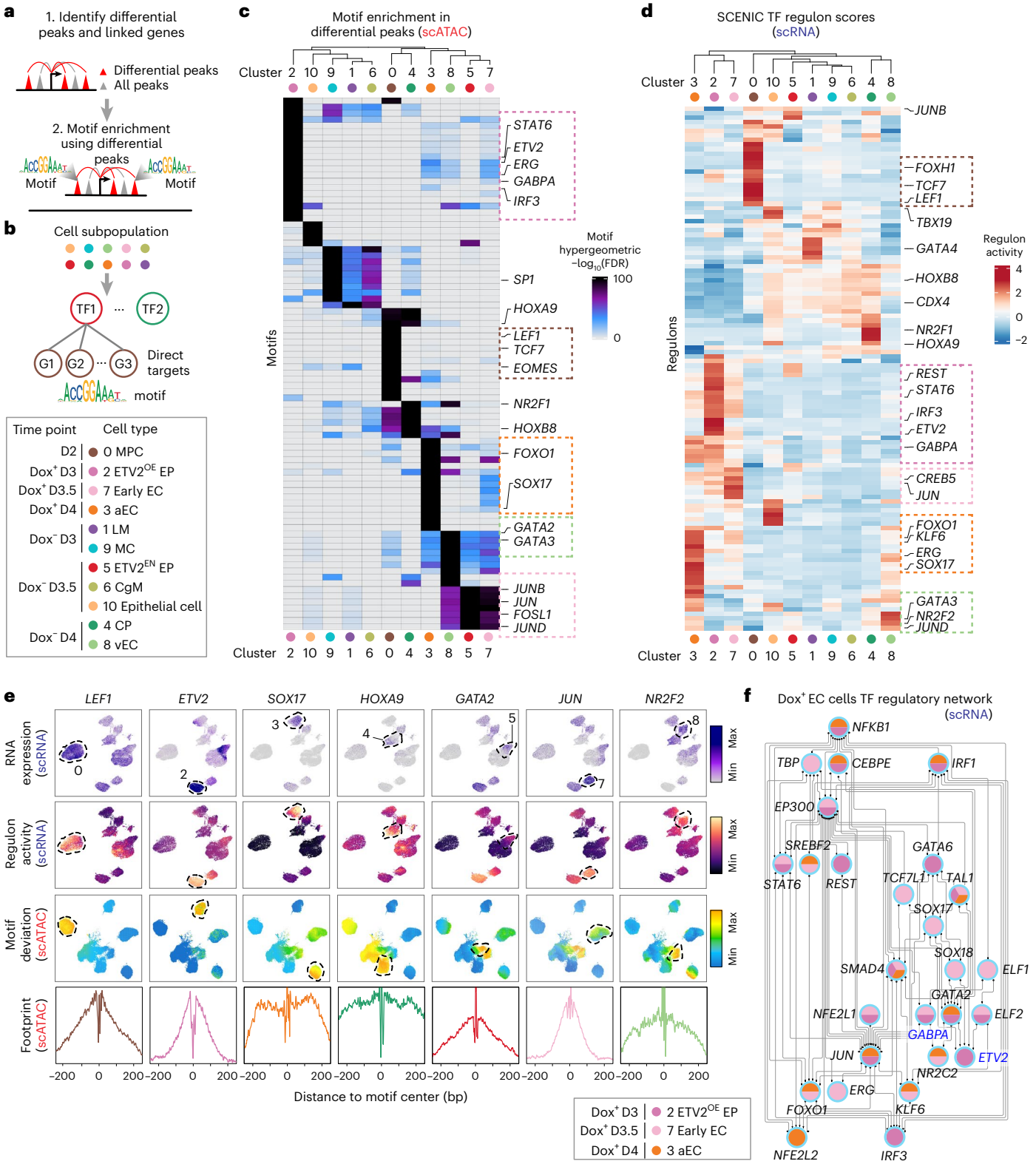


Fig. 3 | Integrative scATAC-seq and scRNA-seq analyses nominate cell-type-resolved TFs that regulate EC specification. a, Schematic for identifying differential peaks and performing motif enrichment within differential peaks using scATAC-seq datasets. **b**, Schematic for identifying TF regulons within different cell clusters using scRNA-seq datasets. **c**, TF motif enrichment in differentially accessible peaks across each *k*-means cluster. Color indicates the motif enrichment ($-\log_{10}$(FDR)) based on the one-sided hypergeometric test. Selected TFs are labeled within dotted boxes that are colored to match the corresponding cell cluster. **d**, Heatmap of SCENIC TF regulon activity across all cell clusters, showing cell-type-specific transcriptional regulatory patterns. Selected TFs are highlighted within dotted boxes as in **c**. **e**, Representative TFs with cell-type selective regulatory activity. Row 1, RNA expression plotted on scRNA-seq UMAP; row 2, regulon activity plotted on scRNA-seq UMAP; row 3, chromVAR TF deviation (motif deviation) plotted on scATAC-seq UMAP; row 4, Tn5 bias-adjusted TF footprints. Gene expression was positively correlated with motif deviation across differentiation. **f**, Regulatory network of TFs in Dox$^+$ clusters. Nodes represent TFs enriched in these clusters by motif enrichment (**c**) or SCENIC analyses (**d**). Edges represent protein–protein associations from STRING. Node colors indicate Dox$^+$ EC clusters.

https://doi.org/10.1038/s44161-025-00660-y

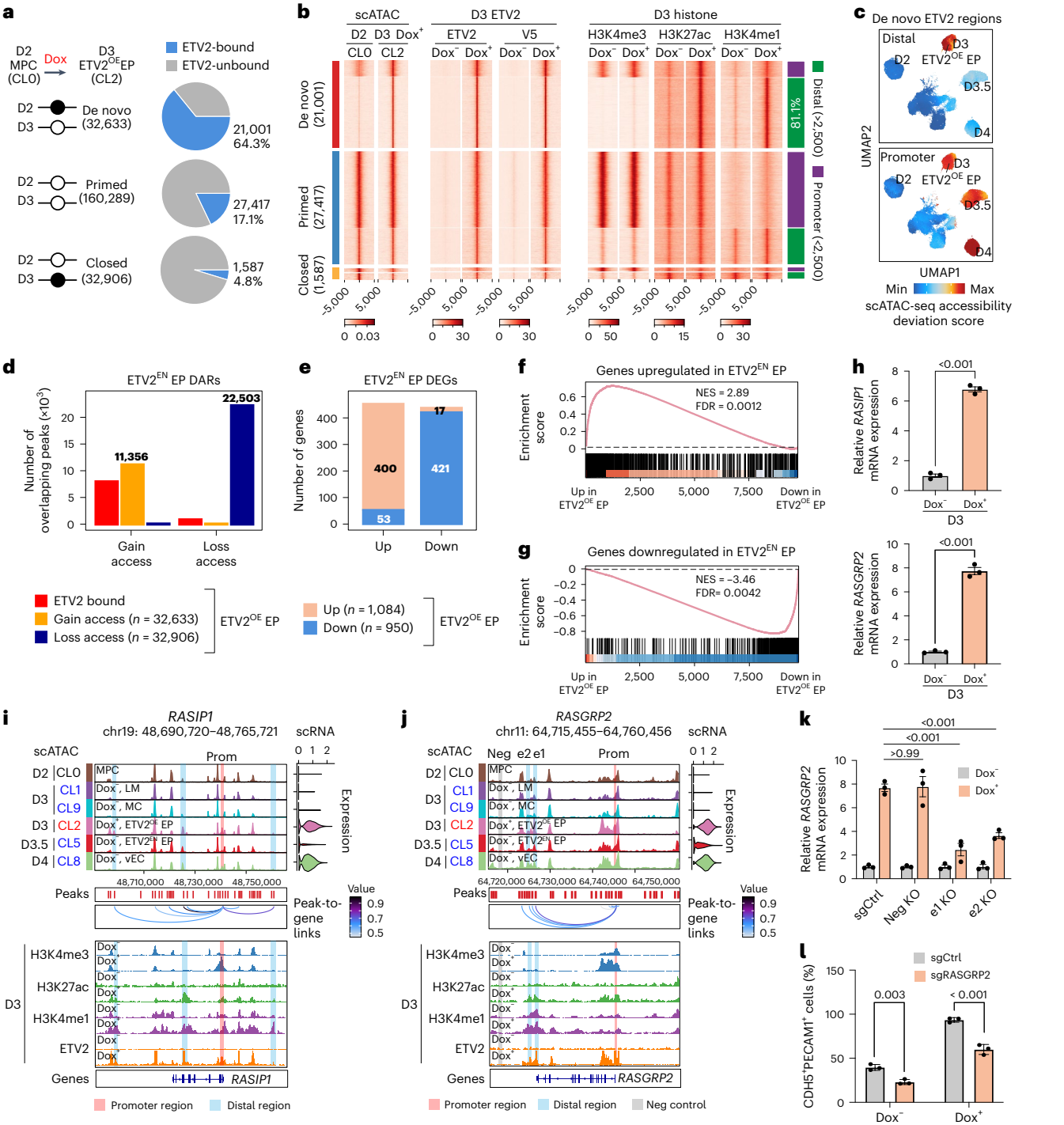


Fig. 4 | ETV2 promotes EC specification by modulating EC gene expression and GTPase signaling. a, Accessible regions from scATAC data in D2 MPC (CL0) and D3 ETV2^{OE} EP (CL2) were classified into de novo (gained accessibility), primed (persistent accessibility) and closed (lost accessibility) regions. Pie charts show the percentage of ETV2-bound and unbound sites in each region class. **b**, Heatmap showing signal for the indicated chromatin features at all high-confidence ETV2 binding sites. Regions are grouped by region class (de novo, primed and closed) and location with respect to transcription start site (TSSs) (promoter and distal). **c**, Accessibility of ETV2-bound de novo promoter (≤2.5 kb from TSS) and distal regions (>2.5 kb to TSS) during EC specification. Accessibility deviation scores were mapped onto scATAC-seq UMAP plots. **d**, The plot shows regions with increased (gain access) or decreased (loss access) accessibility in ETV2^{OE} EP (versus MPC) that were shared with ETV2^{EN} EP regions with increased or decreased accessibility. Overlap with ETV2-bound regions in ETV2^{OE} EP is also shown. DARs, differentially accessible regions. **e**, Number of up- and downregulated genes in ETV2^{OE} EP (versus MPC) were shared with those in ETV2^{EN} EP. **f**,**g**, Gene set enrichment analysis (GSEA) of sets of up- or downregulated genes in ETV2^{EN} EP among genes differentially expressed in ETV2^{OE} EP (versus MPC). Normalized enrichment score (NES) and FDR are calculated using GSEA. **h**, RT–qPCR analysis of *RASIP1* and *RASGRP2* mRNA expression in each group on D3. $n = 3$ independent biological replicates. Two-tailed unpaired Student's t-test. **i**,**j**, Browser track showing chromatin features across clusters for the *RASIP1* (**i**) and *RASGRP2* (**j**) loci. Peak-to-gene loops are based on the correlation between peak accessibility and gene expression. Gene expression is shown for each cluster from scRNA-seq. **k**, RT–qPCR measurement of *RASGRP2* expression after Cas9-mediated deletion of negative (neg), enhancer 1 (e1) and enhancer 2 (e2) regions highlighted in **j**. $n = 3$ independent biological replicates. Two-way analysis of variance (ANOVA) with Tukey's multiple comparisons. **l**, Quantification of the percentage of CDH5^{+}PECAM1^{+} cells generated at D4 by differentiation of sg*RASGRP2* or sgCtrl iPS cells under Dox^{-} and Dox^{+} conditions. $n = 3$ independent biological replicates. Two-tailed unpaired Student's t-test. Graphs in **h**, **k** and **l** show mean ± s.e.m.

(Fig. 3d). By comparing the concordance between TF RNA expression level, TF regulon activity and TF motif activity and footprint, we corroborated the regulatory activity of *LEF1* in MPCs (CL0), *ETV2* in ETV2^{OE} EPs (CL2), *JUN* in early ECs (CL7), *GATA2* in ETV2^{EN} EPs (CL5), *SOX17* in aECs (CL3) and *NR2F2* in vECs (CL8) (Fig. 3e).

We used the top-ranked TFs from motif-based and regulon analyses in Dox$^+$ EC clusters (CL2, CL7 and CL3 in Fig. 3c,d) and protein–protein data from STRING[30] to construct the TF regulatory network that regulates Dox$^+$ EC differentiation (Fig. 3f). Two prominent TFs in ETV2^{OE} EP were *ETV2* and *GABPA*. These two TFs shared targets, including genes upregulated in Dox$^+$ cells and associated with vessel structure and physiological function such as *RASIP1*, *RASGRP2* and *ARHGEF12* (Extended Data Fig. 6).

Collectively, this analysis identified key TFs predicted to regulate EC specification in native and ETV2^{OE}-induced MPC differentiation.

ETV2 modulates dynamic chromatin accessibility

To investigate the direct role of ETV2 in inducing chromatin and enhancer remodeling during EC specification, we measured ETV2 chromatin occupancy in D3 Dox$^+$ cells using CUT&RUN[31] (Extended Data Fig. 7a). Measurements made in biological duplicate using ETV2 and V5 antibodies yielded 43,277 high-confidence peaks ('ETV2 regions') shared between replicates and between ETV2 and V5 antibodies (Extended Data Fig. 7b,c). Most ETV2 regions (57%) contained a specific ETS motif (Extended Data Fig. 7d). ETV2 occupancy was barely detectable in D3 Dox$^-$ cells (Extended Data Fig. 7e,f), consistent with little ETV2 expression at this time point and demonstrating the specificity of the chromatin occupancy signal. Genome browser views of known ETV2 target genes *TJP2*, *ELK3* and *LMO2* demonstrated representative ETV2 binding at their enhancers and promoters (Extended Data Fig. 7g,h).

To focus on the direct effects of ETV2, we analyzed chromatin accessibility changes between D2 MPC (CL0) versus D3 Dox$^+$ ETV2^{OE} EP (CL2). Forced ETV2 expression opened 32,633 'de novo' regions in Dox$^+$ conditions that gained accessibility at D3 (Fig. 4a). Regions were accessible at both D2 and D3 and were referred to as 'primed' accessible regions. Forced ETV2 expression also closed 32,906 regions that lost accessibility at D3 (Fig. 4a).

In D3 Dox$^+$ cells, the majority (21,001, 64.3%) of de novo peaks were associated with ETV2 binding, demonstrating the pioneering activity of ETV2 to bind and open chromatin. In comparison, much smaller fractions of primed (27,417, 17.1%) and closed (1,587, 4.8%) regions were ETV2 bound (Fig. 4a). The ETV2 motif was the top-ranked motif among de novo and primed ETV2 regions (Extended Data Fig. 7i). Cofactors have been shown to modulate pioneer TF binding[4–7]. In addition to the ETV2 motif, we found significant enrichment for binding motifs corresponding to ETS member *GABPA* and TFs from other families, including *IRF3* and *STAT6* (Extended Data Fig. 7i). The ETS–FOX heterodimer motif[32] was also highly enriched, particularly among de novo ETV2 regions. This motif was previously found to be highly enriched in p300-bound EC enhancers in vivo[33] and to be sufficient to drive enhancer EC activity during embryonic vascular development[34].

Post-translational histone modifications associated with active promoters or enhancers, measured by H3K4me3, H3K27ac and H3K4me1 CUT&RUN, increased on nucleosomes neighboring de novo accessible ETV2 regions (Fig. 4b), suggesting the formation of active enhancers or promoters. The majority of de novo accessible ETV2 regions (81.1%) were located in distal regions (Fig. 4b). The remaining de novo accessible ETV2 regions were enriched near promoters, considering that promoters comprise a small fraction of the genome. Most de novo accessible ETV2 regions located in distal intergenic regions were transiently accessible in ETV2^{OE} EP and became inaccessible at D3.5 following Dox withdrawal (Fig. 4c, top). In contrast, de novo accessible ETV2 regions located near promoters retained accessibility even after Dox withdrawal (Fig. 4c, bottom). This suggests that ETV2-driven establishment of enhancers instructs early steps in EC specification, whereas ETV2-associated promoter activation is more stable through EC differentiation.

Taken together, these results demonstrate that ETV2's pioneering activity reshape the epigenetic landscape and identify enhancers and promoters directly regulated by ETV2.

ETV2-regulated enhancers and target genes

To investigate ETV2 regulation of gene expression, we analyzed differentially expressed genes (DEGs) between D2 MPC (CL0) and D3 Dox$^+$ ETV2^{OE} EP (CL2) (Supplementary Table 6). Genes associated with de novo ETV2 regions were enriched among genes upregulated in D3 Dox$^+$ (Extended Data Fig. 7j). Among upregulated genes, 64% (694 of 1,084) had a de novo ETV2 peak within 50 kb upstream or downstream of their transcription start site, versus 30% (295 of 950) of downregulated genes. Dox upregulated DEGs were highly enriched for functional terms related to GTPase activity and vasculature development (Extended Data Fig. 7k). For example, ETV2 overexpression activated *EGFL7* and *FLT1*, which are required for angiogenesis, EC migration, and endothelial integrity[35,36]. In contrast, downregulated DEGs were enriched for functional terms related to embryonic organ development, cardiac ventricle development and muscle cell differentiation (Extended Data Fig. 7l). To assess whether these findings are relevant in native differentiation, we compared the overlap in chromatin accessibility and gene expression changes between ETV2^{OE} EP in D3 Dox$^+$ and ETV2^{EN} EP in D3.5 Dox$^-$. On one hand, we found that 11,356 of the ETV2^{OE} EP open sites (32,633) were opened in the ETV2^{EN} EP, 71% of which were bound by ETV2 (Fig. 4d and Supplementary Table 7). Twenty-two thousand five hundred three of the ETV2^{OE} EP closed sites (32,906) were also closed in the ETV2^{EN} EP. On the other hand, genes upregulated in ETV2^{EN} EP were highly enriched among upregulated genes in ETV2^{OE} EP (Fig. 4e,f) and genes downregulated in ETV2^{EN} EP were highly enriched among downregulated genes in ETV2^{OE} EP (Fig. 4e,g).

We further investigated enrichment of GTPase-related functional terms by ETV2^{OE}. We confirmed the upregulation of GTPase regulators *RASIP1* (Ras interacting protein 1) and *RASGRP2* (RAS guanyl-releasing protein 2) in D3 Dox$^+$ cells by quantitative PCR with reverse transcription (RT–qPCR) (Fig. 4h). We also identified candidate CREs that might regulate these genes, based on ETV2 binding, ETV2-dependent enhancer features (chromatin accessibility, H3K27ac, and H3K4me1), and strong peak-to-gene links predicted based on correlation between region accessibility and gene expression (Fig. 4i,j). To test the requirement for the predicted CREs to activate these genes, we selected two adjacent CREs to *RASGRP2* (e1 and e2; Fig. 4j) for Cas9-mediated deletion. Deletion of either e1 or e2 strongly attenuated *RASGRP2* upregulation by Dox at D3 (Fig. 4k). Furthermore, we confirmed that Cas9-mediated *RASGRP2* knockout reduced EC specification efficiency in both Dox$^+$ and Dox$^-$ conditions (Fig. 4l).

These results identify enhancers and genes directly regulated by ETV2 to promote EC specification.

CRISPR screening uncovers regulators of EC specification

The analyses of regulons and TF motifs involved in EC specification and differentiation (Fig. 3) and of direct ETV2 target genes (Fig. 4) nominated transcriptional regulators that may collaborate with ETV2 to drive EC lineage commitment. To experimentally assess the contribution of these 510 candidate genes to EC lineage commitment induced by ETV2 overexpression, we performed a focused, pooled CRISPR screen using a library containing six sgRNAs targeting each candidate gene[37]. We transduced TRE3G-ETV2 iPS cells with lenti-Cas9-GFP and the lenti-sgRNA-mCherry library lentivirus (Fig. 5a). Green fluorescent protein (GFP) and mCherry double-positive cells were sorted and differentiated under Dox$^+$ conditions. At D4, CDH5$^+$PECAM1$^+$ cells were isolated by FACS. Next-generation sequencing of single guide RNAs (sgRNAs) within the double-positive ECs were compared to D2

https://doi.org/10.1038/s44161-025-00660-y

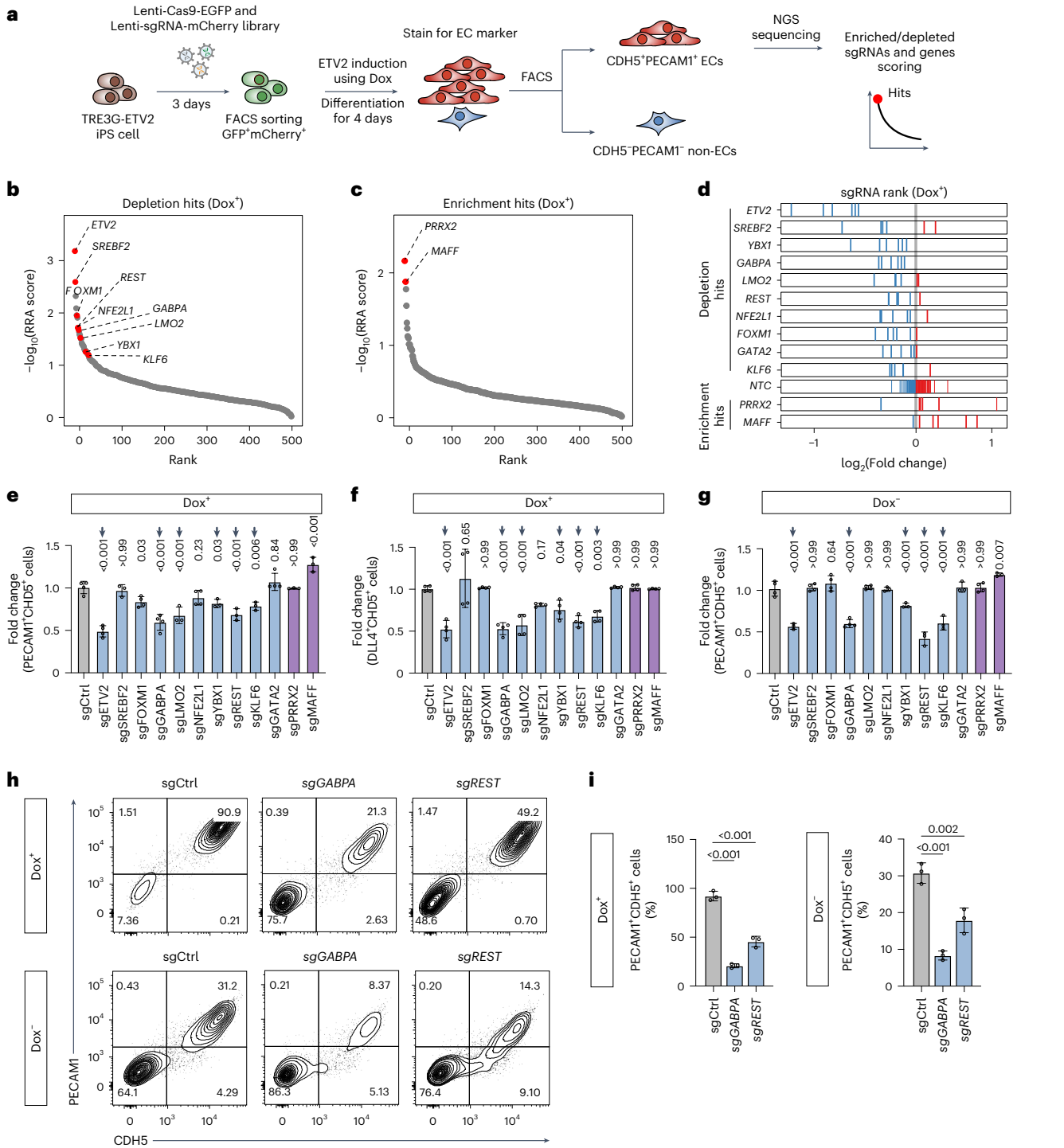


Fig. 5 | CRISPR-Cas9 screening identifies GABPA and REST as key regulators of ETV2-directed EC specification. a, Schematic of the CRISPR screening approach for identifying TF regulators of EC specification. Oligonucleotide synthesis was used to generate a pooled lenti-sgRNA-mCherry library containing 3,150 gRNAs that targeted candidate transcriptional regulators. TRE3G-ETV2 iPS cells were transduced with lenti-Cas9-GFP and the lenti-sgRNA-mCherry library to establish stably transduced cell pools. The cells were then differentiated with Dox. FACS for CDH5 and PECAM1 was used to separate ECs from non-ECs. Enriched and depleted sgRNA were identified by next-generation sequencing (NGS) of sgRNAs. **b**,**c**, Depletion hits (**b**) and enrichment hits (**c**). Each gene was ranked based on the MAGeCK robust ranking aggregation (RRA) score. The *y* axis represents $-\log_{10}$(RRA score). The selected top hits are labeled in red. **d**, Top screening candidate genes, showing relative rank for each individual sgRNA (six sgRNAs per gene). The *x* axis represents $\log_2$ fold change of sgRNA. Blue and red bars indicate sgRNA depletion or enrichment, respectively. **e**,**f**, Validation of screening hits. TRE3G-ETV2 iPS cells were transfected with a sgCtrl or sgRNA pool targeted to the indicated gene. Differentiation efficiency under Dox+ conditions to PECAM1+CDH5+ ECs (**e**, each group $n = 3$ independent biological replicates, except sgCtrl, sg*ETV2*, sg*FOXM1*, sg*GABPA*, sg*NFE2L1* and sg*GATA2*: $n = 4$ independent biological replicates) or DLL4+CDH5+ aECs (**f**, each group $n = 4$ independent biological replicates) was measured by flow cytometry. **g**, Differentiation efficiency under control conditions (Dox−) to PECAM1+CDH5+ ECs was measured by flow cytometry. Each group $n = 3$ independent biological replicates, except sgCtrl, sg*SREBF2*, sg*FOXM1*, sg*GABPA*, sg*LMO2* and sg*PRRX2*: $n = 4$ independent biological replicates). **h**,**i**, Flow cytometry analysis (**h**) and quantification (**i**) for percentage of PECAM1+CDH5+ cells in sg*Ctrl*, sg*GABPA* and sg*REST* group under Dox+ and Dox− conditions on D4. $n = 3$ independent biological replicates. One-way ANOVA with Dunnett's multiple comparison test to sgCtrl (**e**–**g**,**i**). Graphs in **e**–**g** and **i** show mean ± s.d.

https://doi.org/10.1038/s44161-025-00660-y

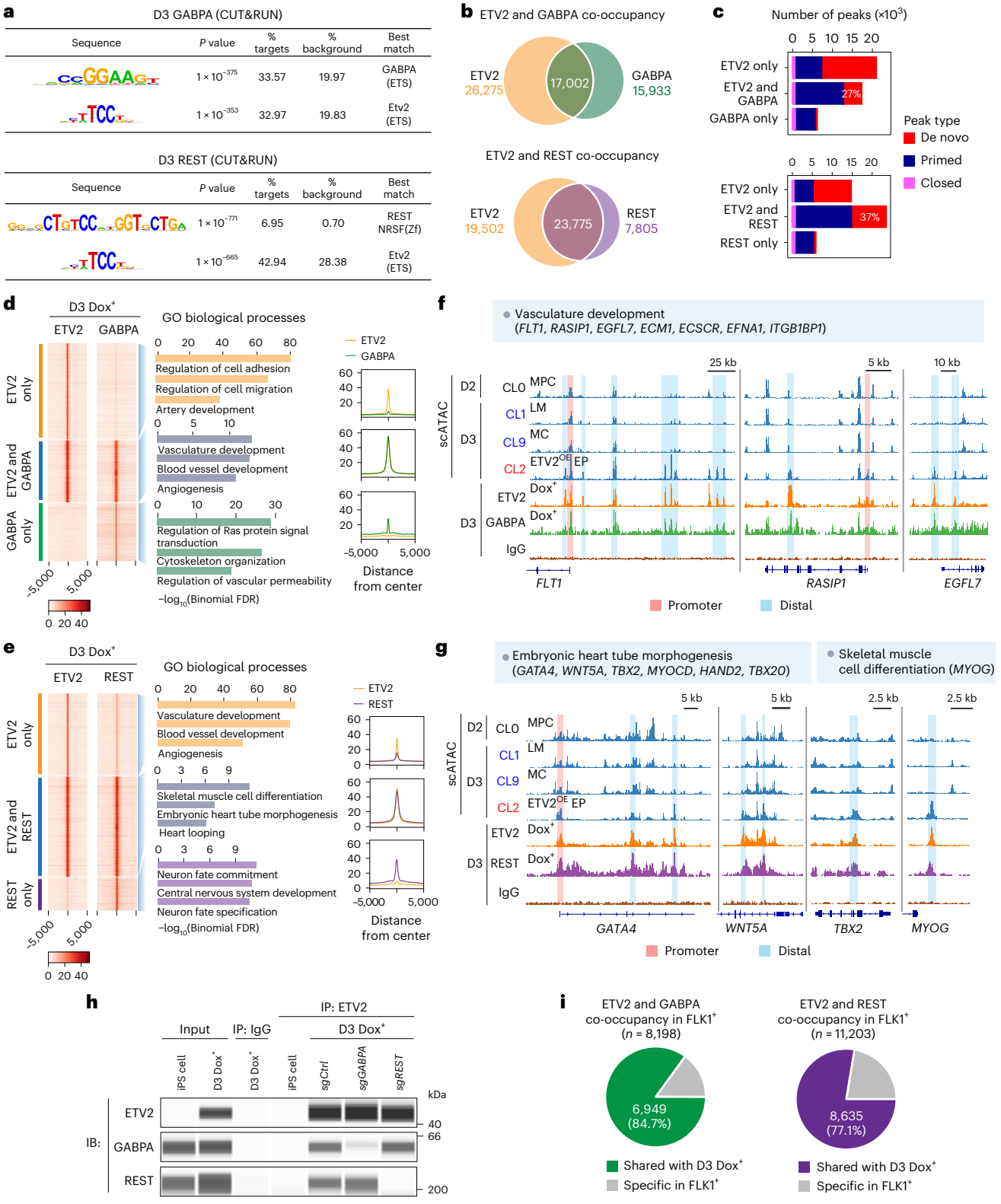


Fig. 6 | ETV2 recruits GABPA and REST to enhancers. a, The two top enriched motifs in all GABPA regions or REST regions in D3 Dox$^+$ cells. Motif enrichment *P* values are calculated using a one-sided binomial test in HOMER. **b**, Venn diagrams showing the overlap between GABPA and ETV2 binding peaks (top: hypergeometric test $P = 1.1 \times 10^{-216}$), or between REST and ETV2 binding peaks (bottom: hypergeometric test $P = 0$) in D3 Dox$^+$ cells. **c**, Overlap of GABPA and ETV2 region (top) or REST and ETV2 regions (bottom), separated into de novo, primed and closed subsets based on the scATAC-seq data presented in Fig. 4b. **d**,**e**, Heatmaps showing GABPA (**d**) or REST (**e**) binding signals at ETV2 regions, demonstrating ETV2 co-occupancy with GABPA or REST (left). Regions with different TF occupancy patterns had distinct enriched GO biological processes terms (GREAT; $-\log_{10}$(binomial FDR), binomial test) (middle). Mean CUT&RUN signals in each region (right). **f**,**g**, Genome browser views showing scATAC-seq signals in D2 (CL0), D3 Dox$^-$ (CL1, CL9), D3 Dox$^+$ (CL2, ETV2^{OE} EP) cells and ETV2, GABPA (**f**) or REST (**g**) binding signals in D3 Dox$^+$ cells at regions of representative genes. **h**, Co-IP using ETV2 or IgG control antibodies, followed by capillary immunoblotting (IB) for ETV2, GABPA or REST, in iPS cell or D3 Dox$^+$ sgCtrl, sg*GABPA* or sg*REST* cells. Data are representative of $n = 3$ biologically independent experiments. **i**, Number of ETV2 and GABPA (left) and ETV2 and REST (right) co-occupied sites that were found in FLK1$^+$ cells.

mesoderm cells as input to identify sgRNAs and their associated genes that altered MPC-to-EC differentiation (Fig. 5a).

The screen identified genes whose knockout significantly inhibited (depleted hits) or promoted (enriched hits) EC differentiation induced by ETV2 overexpression (Fig. 5b–d). Several of these genes were previously inactivated in mice. Some, including *KLF6* and *FOXM1*, were previously implicated in vascular development (Supplementary Table 8), but their requirement for EC specification has been little studied. We selected ten depleted and two enriched hits for individual validation. For each gene, we transduced TRE3G-ETV2 iPS cells with lenti-Cas9-GFP and a lenti-sgRNA-mCherry pool that expressed the gene's six sgRNAs (Fig. 5d). We did not observe changes in pluripotency marker SSEA-4 in the CRISPR-treated iPS cells pools compared to non-targeting control (sg*Ctrl*) (Extended Data Fig. 8a). We next tested the CRISPR-treated iPS cell pools for EC differentiation under Dox$^+$. Knockout of *GABPA*, *LMO2*, *YBX1*, *REST* and *KLF6* impaired EC specification efficiency, whereas knockout of *MAFF* increased it (Fig. 5e). These genes were also required for efficient differentiation of DLL4$^+$CDH5$^+$ aECs (Fig. 5f). Under Dox$^-$ conditions, we likewise observed that *GABPA*, *YBX1*, *REST* and *KLF6* were required for efficient EC differentiation and that *MAFF* limited EC differentiation efficiency (Fig. 5g and Extended Data Fig. 8b), demonstrating the relevance of these genes to native EC specification from iPS cells.

To improve knockout efficiency and reduce potential off-target effects, we knocked out *GABPA*, *REST*, *KLF6*, *LMO2* and *YBX1* in the ETV2-inducible iPS cells using the gene's three most effective sgRNAs (Fig. 5d). Parallel treatment with the non-targeting gRNA yielded control cells. These iPS cells propagated and sustained the expression of core pluripotency TFs (Extended Data Fig. 8c,d). We differentiated these iPS cells to MPCs and did not observe a notable change in mesoderm marker genes (Extended Data Fig. 8c,d), indicating that mesoderm formation was not affected. Consistent with the results from CRISPR-treated cell pools, *GABPA*, *REST* and *KLF6* knockout significantly reduced the fraction of PECAM1$^+$CDH5$^+$ ECs on D4 compared to control in either Dox$^+$ or Dox$^-$ conditions (Fig. 5h,i and Extended Data Fig. 8e). Examination of the expression profile of these genes along the Dox$^+$ EC trajectory (Fig. 2e) showed that *REST* and *GABPA* maintained expression levels during the peak of *ETV2* expression, whereas *KLF6* was upregulated at later pseudotimes (Extended Data Fig. 8f). Western blotting confirmed depletion of GABPA or REST with no change in ETV2 levels (Extended Data Fig. 9a,b). Taken together, our loss of function screen and validation experiments suggest that *GABPA* and *REST* are functional regulators of native and ETV2^{OE}-induced EC specification.

GABPA and REST co-occupy ETV2 regions in ETV2^{OE} cells

GABPA, an ETS-family TF that is required for early embryogenesis, has not been functionally analyzed in EC specification. Similarly, the function of *REST*, a well-characterized transcriptional repressor[38,39], in EC specification has not been explored.

To test the hypothesis that GABPA and REST regulate EC lineage commitment by cooperating with ETV2, we measured their genomic occupancy using CUT&RUN on D3 Dox$^+$ cells (Supplementary Tables 9 and 10). A notable CUT&RUN signal was not detected in cells with deficiency of GABPA or REST (Extended Data Fig. 9c,d), indicating the assays' high specificity. GABPA and full-length REST motifs were among the top two most significantly enriched motifs (Fig. 6a), demonstrating the high quality of the data. Notably, ETV2 was another top enriched motif in the GABPA or REST regions (Fig. 6a). Overall, 51% of GABPA regions and 75% of REST regions overlapped with ETV2 regions (Fig. 6b); we defined overlapping chromatin occupancy signals as 'co-occupancy'. Indeed, a substantial fraction of the regions co-occupied by ETV2 and GABPA or REST were de novo accessible regions in D3 ETV2^{OE} EPs (ETV2 and GABPA, 27%; ETV2 and REST, 37%) (Fig. 6c); however, few GABPA or REST regions without ETV2 (GABPA only and REST only) were within de novo accessible chromatin regions, indicating that GABPA and REST have little independent pioneering activity. To determine whether REST or GABPA were bound to REST-ETV2 or GABPA-ETV2 de novo co-occupied regions before ETV2 induction, we performed REST and GABPA CUT&RUN on D2 cells. We found that most regions co-occupied by REST-ETV2 or GABPA-ETV2 in D3 Dox$^+$ cells were not bound by REST or GABPA at D2 (Extended Data Fig. 9f,g). The CUT&RUN occupancy signal for ETV2, GABPA and REST was substantially stronger at co-occupied regions compared to regions occupied by each TF alone (Fig. 6d,e), suggesting collaborative binding. The results are consistent with previous studies demonstrating that pioneer TF activity is modulated by their interaction with non-pioneering TFs[4–7].

We analyzed the functional terms enriched among genes neighboring regions occupied by ETV2 and GABPA (Fig. 6d). ETV2 and GABPA co-occupied regions revealed enrichment for ontologies involved in vasculature development, and angiogenesis. For example, GABPA and ETV2 co-occupied regions near several upregulated genes involved in vasculature development, such as *FLT1*, *RASIP1* and *EGFL7* (Fig. 6f), indicating that ETV2 and GABPA cooperate to participate in early vascular development. We found that GABPA and ETV2 co-occupied their promoters and enhancers, which were also opened following ETV2 overexpression in scATAC-seq data (CL2 versus CL0; Fig. 6f).

ETV2 and REST regions displayed distinct functional categories between ETV2 plus REST, ETV2-only and REST-only subsets (Fig. 6e). ETV2-only regions were enriched for GO terms related to vascular development, REST-only regions were related to neuron fate commitment and central nervous system development, and ETV2 plus REST co-occupied regions were associated with the embryonic heart tube morphogenesis, heart looping and skeletal muscle cell differentiation. Among the REST and ETV2 co-occupied regions were de novo accessible regions of several cardiac (for example, *GATA4* and *TBX2*) and muscle genes (for example, *MYOG*; Fig. 6g), which were silenced in mature ECs, indicating that ETV2 recruits REST to suppress non-EC lineage genes. ETV2 and REST co-occupied a subset of VISTA-validated cardiac enhancers[40] or chromatin regions that were differentially accessible in fetal heart and iPS cell-derived cardiomyocytes[41] (Extended Data Fig. 9h).

Motif analysis indicated that ETV2 and REST co-occupied regions frequently enriched the ETV2 but not REST motif, whereas the REST motif was found in REST-only regions (Extended Data Fig. 9e). This observation suggested that REST is recruited to ETV2 and REST co-occupied regions by protein–protein interaction with ETV2. To test this hypothesis, we assessed ETV2-REST interaction using co-immunoprecipitation (co-IP) assays. REST co-immunoprecipitated with ETV2, as did GABPA (Fig. 6h). These data suggest that REST and GABPA co-occupy sites with ETV2 through protein–protein interactions.

To assess whether these correlations hold in ETV2-expressing cells during native differentiation, we FACS sorted D3.5 Dox$^-$ cells by their expression of FLK1 (VEGFR2), as FLK1$^+$ cells contain ETV2^{EN} EPs[13,42]. We performed CUT&RUN for ETV2, REST and GABPA on the FLK1$^+$ populations. Of the 8,198 ETV2 and GABPA co-occupied sites in FLK1$^+$ cells, 6,949 (84%) were found in D3 Dox$^+$ cells (Fig. 6i). In FLK1$^+$ cells, we observed the co-occupancy of GABPA and ETV2 near vascular genes *FLT1* and *RASIP1* (Extended Data Fig. 9i). Similarly, of the 11,203 ETV2 and REST co-occupied sites in FLK1$^+$ cells, 8,635 (77%) were found in D3 Dox$^+$ cells (Fig. 6i). The ETV2 and REST co-occupied de novo CREs near cardiac (for example, *GATA4*) and muscle genes (for example, *MYOG*) were also bound by ETV2 or REST in FLK1$^+$ cells but showed no detectable ETV2 or REST signals in FLK1$^-$ cells (Extended Data Fig. 9j). We further confirmed that GABPA and REST co-immunoprecipitated with ETV2 in FLK1$^+$ cells (Extended Data Fig. 9k).

Collectively, these results demonstrate that ETV2 opens chromatin and recruits GABPA and REST to regulate target genes in ETV2^{OE} EPs and ETV2^{EN} EPs.

https://doi.org/10.1038/s44161-025-00660-y

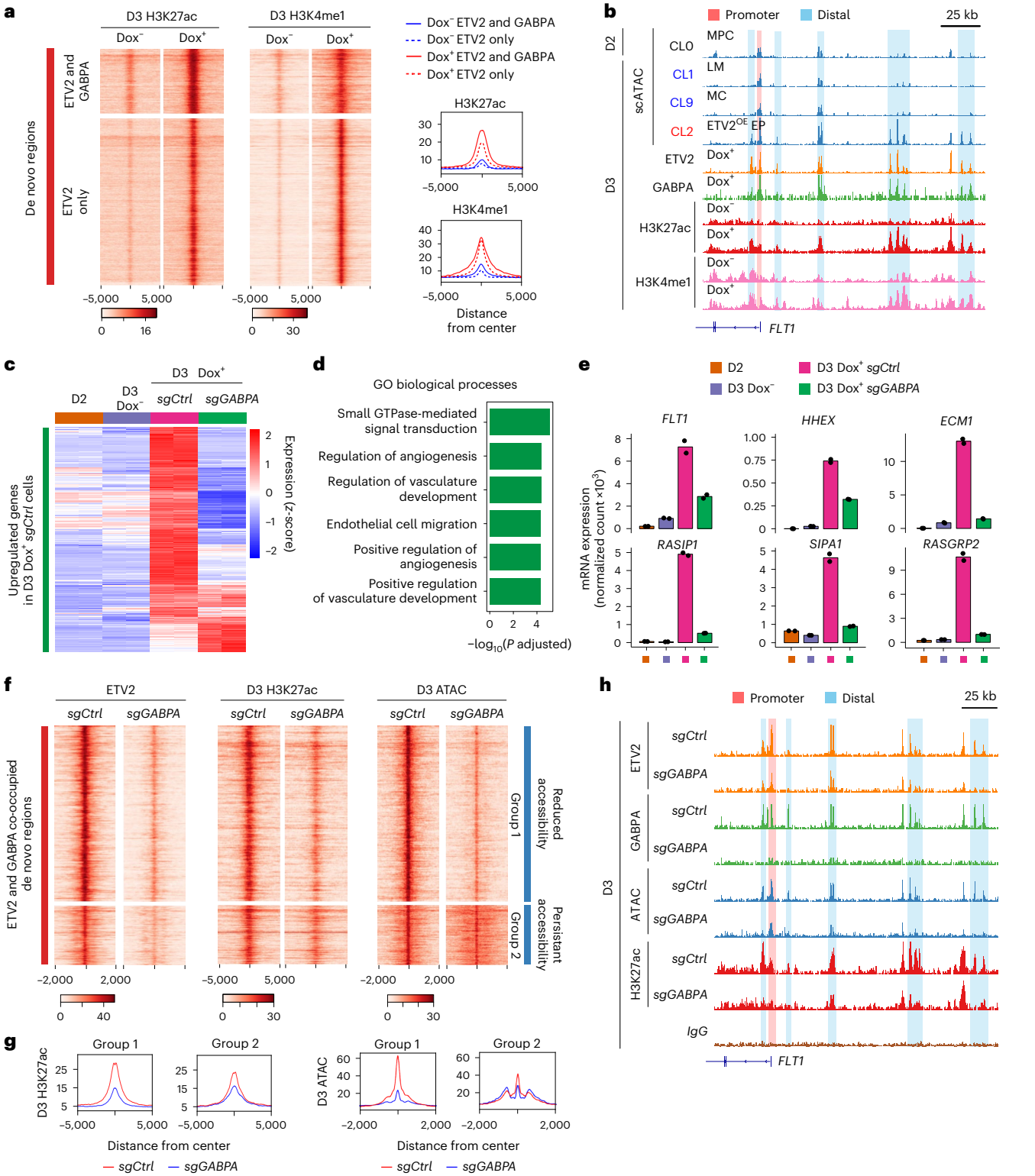


Fig. 7 | GABPA collaborates with ETV2 to open chromatin and promote EC gene expression. a, Heatmap showing signal for the indicated histone features at de novo ETV2 regions, with GABPA co-occupancy (ETV2 and GABPA) or without GABPA co-occupancy (ETV2-only) (left). Mean H3K27ac and H3K4me1 CUT&RUN signals in the class of regions (right). **b**, Genome browser view showing scATAC, ETV2, GABPA, H3K27ac and H3K4me1 signals at the *FLT1* locus. **c**, Heatmap of the expression of selected genes in each group. Selected genes were upregulated in D3 Dox$^+$ sg*Ctrl* (versus D3 Dox$^-$). Each row is *z*-score normalized. **d**, GO term analysis of genes that were downregulated in D3 Dox$^+$ sg*GABPA* cells (versus D3 Dox$^+$ sg*Ctrl*). Fisher's exact test, corrected for multiple testing using the Benjamini–Hochberg method. **e**, Examples of genes downregulated in D3 Dox$^+$ sg*GABPA* cells. These genes were co-occupied by ETV2 and GABPA in D3 Dox$^+$ sg*Ctrl* cells. $n = 2$ biologically independent samples. **f**, Heatmap showing the indicated chromatin features from D3 Dox$^+$ sg*Ctrl* and sg*GABPA* cells at ETV2 and GABPA co-occupied de novo accessible regions. Region subsets (group 1) had reduced accessibility in sg*GABPA* cells, and group 2 had similar accessibility in D3 Dox$^+$ sg*Ctrl* and sg*GABPA* cells. **g**, Aggregation plots showing mean H3K27ac CUT&RUN signals (left) and average ATAC signal (right) from region subsets in D3 Dox$^+$ sg*Ctrl* and sg*GABPA* cells. **h**, Representative genome browser view at *FLT1* showing ETV2 binding, GABPA binding, ATAC-seq and H3K27ac deposition in D3 Dox$^+$ sg*Ctrl* and sg*GABPA* cells.

GABPA and ETV2 collaboratively activate early EC genes

GABPA was shown to be required for formation of enhanceosome complexes[43]. In Dox$^+$ conditions, ETV2 and GABPA co-occupied regions had stronger H3K27ac and H3K4me1 signals than ETV2-only regions (Fig. 7a). Indeed, strong histone signals were observed for genes associated with the ETV2 and GABPA co-occupied de novo regions (Fig. 7b), such as *FLT1*, suggesting that GABPA helps ETV2 install functional enhancers.

To unravel the contribution of GABPA to transcriptional regulation by ETV2, we analyzed gene expression in D2 cells, D3 Dox$^-$ cells, D3 Dox$^+$ sg*Ctrl* and sg*GABPA* cells (Fig. 7c and Extended Data Fig. 10a). Of note, most genes upregulated by ETV2 overexpression (D3 Dox$^+$ sg*Ctrl* versus D3 Dox$^-$) were strongly downregulated in the D3 Dox$^+$ *GABPA* knockout cells (Fig. 7c): genes in this intersection comprised 73% of ETV2^{OE}-upregulated genes and 67% of *GABPA* knockout-downregulated genes (Extended Data Fig. 10b and Supplementary Table 11). ETV2 and GABPA co-occupied most (76%) of these overlapping genes. *GABPA* knockout-downregulated genes were enriched for GO terms related to the regulation of angiogenesis (*FLT1*, *HHEX* and *ECM1*) and small GTPase-mediated signal transduction (*RASIP1*, *SIPA1* and *RASGRP2*; Fig. 7d,e). These genes were associated with the ETV2 and GABPA co-occupied de novo regions, supporting the participation of GABPA in ETV2-mediated gene activation.

To investigate the effect of *GABPA* knockout on ETV2 occupancy and on chromatin accessibility, we conducted ETV2 CUT&RUN and ATAC-seq in D3 Dox$^+$ sg*Ctrl* and sg*GABPA* cells. The knockout of GABPA strongly reduced ETV2 binding at de novo regions where GABPA was co-occupied (Fig. 7f and Extended Data Fig. 10c), indicating that GABPA regulates ETV2 pioneering activity directly, in keeping with previous studies that demonstrated the impact of non-pioneer factors on sites of chromatin opening by pioneer factors[4–7]. Consistent with altered ETV2 pioneering activity, GABPA depletion reduced accessibility of 46.5% of D3 Dox$^+$ de novo ETV2 regions (Extended Data Fig. 10d). Among ETV2 and GABPA co-occupied de novo accessible regions, 75% had reduced accessibility upon depletion of GABPA, corresponding to strongly decreased H3K27ac signals (group 1; Fig. 7f,g). For example, we observed that in *GABPA* knockout cells, ETV2 occupancy decreased in many enhancers of *FLT1*, and their chromatin accessibility and H3K27ac signals decreased (Fig. 7h).

These results demonstrate that GABPA collaborates directly with ETV2 to shape the chromatin accessibility landscape of ETV2^{OE} EP cells, install functional enhancers and activate downstream endothelial genes.

REST recruitment is required to repress non-EC genes

REST binds to a 21-bp consensus sequence known as repressor element 1 (RE1) and represses gene transcription[38,39,44]. In Dox$^+$ conditions, ETV2 and REST co-occupied regions had less H3K27ac and H3K4me1 signal than ETV2-only regions (Fig. 8a). Their de novo co-occupied sites showed reduced H3K27ac marks on neighboring nucleosomes (Fig. 8b), such as *GATA4*, suggesting that REST represses gene expression via reducing histone modifications.

To define which ETV2 target genes and their CREs are coregulated by ETV2 and REST, we compared gene expression profiles of D2 cells, D3 Dox$^+$ sg*Ctrl* and sg*REST* cells (Fig. 8c and Extended Data Fig. 10e). Notably, many of the genes downregulated in D3 Dox$^+$ cells during EC differentiation (D3 Dox$^+$ sg*Ctrl* versus D3 Dox$^-$) were markedly upregulated in *REST* knockout cells (Fig. 8c): the intersection comprised 72% of D3 Dox$^+$ downregulated genes and 44% of *REST* knockout upregulated genes (Extended Data Fig. 10f and Supplementary Table 12). ETV2 and REST co-occupied the majority (66%) of these overlapping genes. GO terms enriched among these genes co-occupied by ETV2 and REST were related to heart specification, and affected genes, included *GATA4*, *GATA5*, *SMAD6*, *GREB1L*, *SMARCD3* and *WNT5A*, all known to regulate heart development (Fig. 8d,e). These results show that ETV2 opens chromatin and recruits REST to repress the expression of non-EC lineage genes, especially cardiac genes, during EC differentiation. Cardiac differentiation was also a major alternative trajectory in Dox$^-$ conditions (Fig. 2c), consistent with the need to actively repress this fate to achieve efficient EC specification.

We investigated the effect of *REST* knockout on ETV2 occupancy and chromatin accessibility by performing ETV2 CUT&RUN and ATAC-seq in D3 Dox$^+$ sg*Ctrl* and sg*REST* cells. Further supporting the cooperation of REST and ETV2 during EC specification, *REST* knockout reduced ETV2 occupancy at de novo accessible regions co-occupied by REST (Fig. 8f and Extended Data Fig. 10g). Reflective of reduced ETV2 pioneering activity, REST inactivation reduced the accessibility of 34.6% of de novo accessible ETV2 regions (Extended Data Fig. 10h). Among ETV2 and REST co-occupied de novo accessible regions, 53% had reduced accessibility in *REST* knockout (group 1, Fig. 8f,g). We examined the CREs of GATA4, a key regulator of cardiac differentiation. The ETV2 and REST co-occupied de novo GATA4 CREs were less accessible in the absence of REST and showed reduced ETV2 binding (Fig. 8h). Loss of REST repression is consistent with GATA4 upregulation in D3 Dox$^+$ *REST* knockout cells (Fig. 8e). REST knockout resulted in increased formation of cTnT$^+$ or ACTN2$^+$ cardiomyocytes during differentiation under Dox$^+$ conditions (Fig. 8i,j).

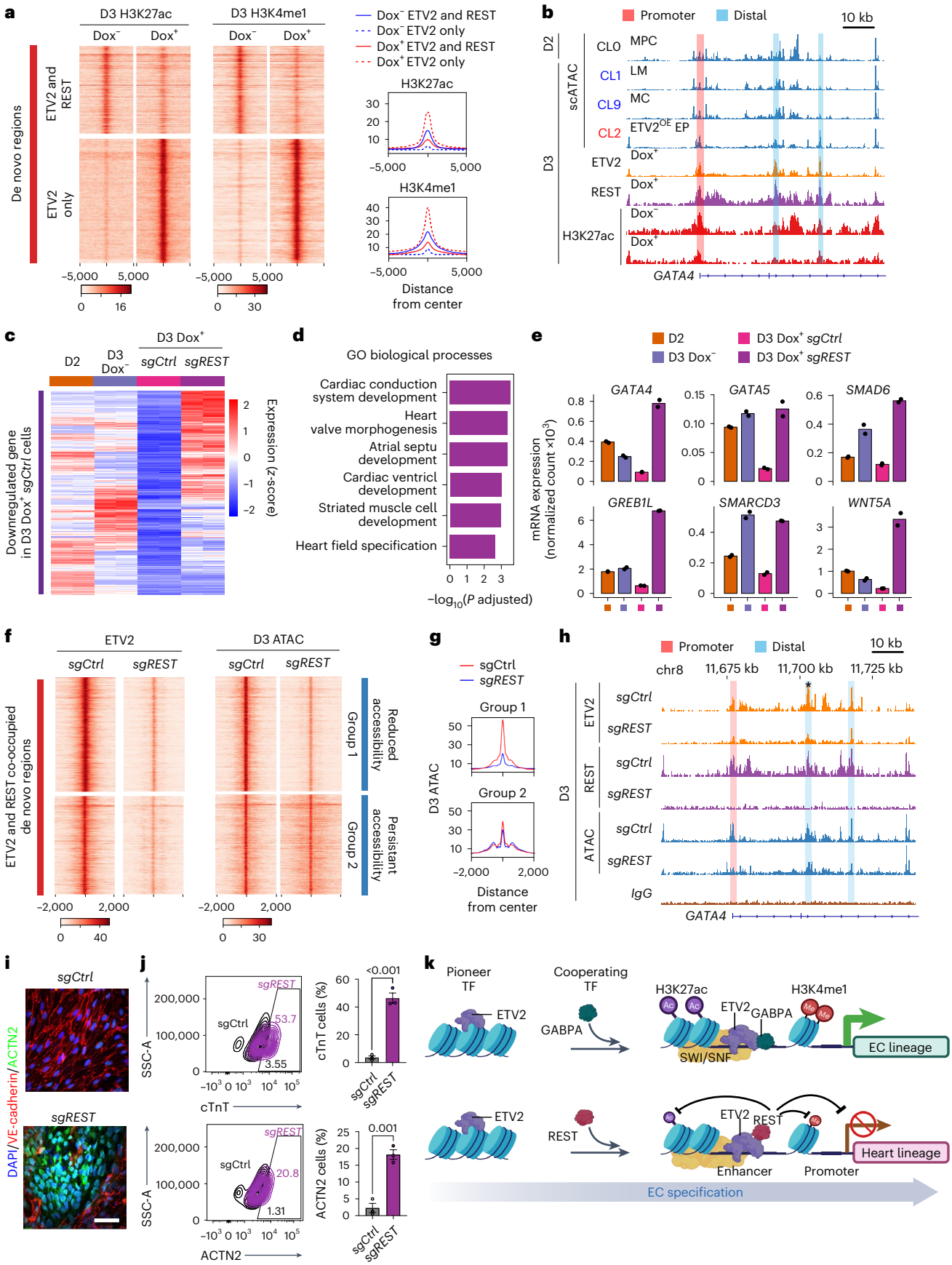
Our findings support the model that ETV2 pioneering activity is influenced by REST, and that ETV2 recruitment of REST is required to repress the cardiac lineage during EC specification. Our study shows that REST typifies a class of lineage TFs with a gatekeeper function capable of restricting plasticity through engaging pioneer TFs, which is essential for cell fate choices determination.

Discussion

To gain insights into mechanisms of cell specification, we developed a highly efficient system in which overexpression of the pioneer factor

Fig. 8 | REST collaborates with ETV2 to open chromatin and repress gene expression. a, Heatmap showing signal for the indicated histone features at de novo ETV2 regions, with REST co-occupancy (ETV2 and REST) or without REST co-occupancy (ETV2-only) (left). Mean H3K27ac and H3K4me1 CUT&RUN signals in the class of regions (right). **b**, Genome browser view showing scATAC, ETV2, REST and H3K27ac signals at the *GATA4* locus. **c**, Heatmap of the expression of selected genes in each group. Selected genes were downregulated in D3 Dox$^+$ sg*Ctrl* (versus D3 Dox$^-$). Each row is *z*-score normalized. **d**, GO term analysis of genes that were upregulated in D3 Dox$^+$ sg*REST* cells (versus D3 Dox$^+$ sg*Ctrl*). Fisher's exact test, corrected for multiple testing using the Benjamini–Hochberg method. **e**, Examples of genes upregulated in D3 Dox$^+$ sg*REST* cells. These genes were co-occupied by ETV2 and REST in D3 Dox$^+$ sg*Ctrl* cells. $n = 2$ biologically independent samples. **f**, Heatmap showing the indicated chromatin features for D3 Dox$^+$ sg*Ctrl* and sg*REST* cells at ETV2 and REST co-occupied de novo regions. Region subsets (group 1) had reduced accessibility in sg*REST* cells, and group 2 had similar accessibility in sg*Ctrl* and sg*REST* cells. **g**, Aggregation plots showing average ATAC signal from region subsets in D3 Dox$^+$ sg*Ctrl* and sg*REST* cells. **h**, Representative genome browser view at *GATA4* showing ETV2 binding, REST binding and ATAC-seq in D3 Dox$^+$ sg*Ctrl* and sg*REST* cells. The VISTA-validated cardiac enhancers region is marked with an asterisk. **i**,**j**, Formation of cardiomyocytes under Dox$^+$ differentiation conditions. sg*Ctrl* and sg*REST* cells were analyzed at differentiation D10. **i**, ACTN2 (green), VE-cadherin (red) and DAPI (blue) immunostaining. Representative of $n = 3$ biologically independent experiments. Scale bar, 50 µm. **j**, Flow cytometry analysis of cTnT and ACTN2 expression in sg*Ctrl* (black contour) and sg*REST* (purple contour) cells. $n = 3$ independent biological replicates. Data are mean ± s.e.m. Two-tailed unpaired Student's *t*-test. **k**, Model of pioneer factor ETV2 safeguards EC specification by recruiting the transcription co-activator GABPA to promote EC lineage commitment and repressor REST to restrict alternative lineage commitment. Recruitment of cooperating TFs stabilizes ETV2 binding and participates in altering the local chromatin landscape. Image in **k** created with BioRender.com.

https://doi.org/10.1038/s44161-025-00660-y



ETV2 in MPCs reproducibly drives EC specification in two days. Using this system, we established a single-cell transcriptome and accessible chromatin atlas to capture the transcriptional and epigenetic dynamics of EC specification. We demonstrated that ETV2 overexpression overcomes two key bottlenecks in EC specification: (1) formation of EC progenitors and (2) diversion of cells into non-EC fates. We defined ETV2-bound de novo accessible regions, which depend upon ETV2 binding and pioneering activity to open chromatin, establish active enhancer marks and activate EC genes. Functional CRISPR screening of candidate TFs that collaborate with ETV2 to promote efficient EC specification revealed several factors with previously unrecognized roles in EC specification. Among these were the transcriptional activator GABPA and the transcriptional repressor REST (Fig. 8k).

Pioneer factors promote lineage specification by binding recognition sites within nucleosomal DNA, making the DNA accessible so that typical TFs and other epigenetic factors can bind, recruit transcriptional activators and stimulate the expression of genes that promote lineage commitment[3]. A previous study showed that ETV2 recruits BRG1, the catalytic subunit of the SWI/SNF chromatin remodeling complex, to remodel chromatin[11]. BRG1 may augment ETV2 pioneering activity and increase accessibility of ETV2-bound regions, but it does not obviate the need for TFs to initially bind nucleosomal DNA, marking the site for further epigenetic modification. This canonical pioneering role of ETV2 was exemplified by its interaction with GABPA. GABPA co-occupied regions with ETV2, potentially through direct protein–protein interactions that modulate and enhance ETV2 pioneering activity. This co-occupancy enhanced installation of functional enhancers at de novo accessible regions. Although *ETV2* expression is transient during EC specification, other angiogenic ETS-family TFs such as *FLI1* that are activated by *ETV2* may continue to occupy and activate ETV2-pioneered regions, previously described as an 'ETS switch'[10]. Our data suggest that this ETS switch mechanism occurs most frequently in promoter regions, which demonstrated sustained accessibility after ETV2 downregulation.

Our study revealed an additional facet of ETV2 pioneering activity. We found that ETV2 recruits the transcriptional repressor REST to silence non-EC lineage genes, particularly cardiac genes. In keeping with a previous scRNA-seq study[45], under native conditions without ETV2 overexpression, the majority of cells acquired cardiac and other non-EC fates, in part accounting for inefficient EC specification. Previous studies showed that ETV2 inactivation in mouse embryos and murine embryonic stem cells/embryonic bodies resulted in expansion of cardiac and other mesodermal lineages[46,47]. These data indicate that ETV2 balances mesodermal lineage commitment. We showed that REST ablation reduced EC specification and upregulated cardiac genes. ETV2 physically interacted with REST and recruited it to ETV2-bound de novo accessible regions, where it repressed expression of alternative lineage genes. The function of pioneer TFs in repressing gene expression and thereby altering cell fate choices is just beginning to be studied[48]. We propose that recruitment of transcriptional repressors may be a general feature of pioneer factors that limits transcriptional activation by some newly opened chromatin regions to suppress commitment to alternative lineages. Further studies will be important to define whether GABPA and REST cooperate with ETV2 to control EC specification and differentiation during fetal development.

Early in vascular development, ECs adopt arterial or venous identities, each characterized by distinct transcriptional programs. Developmentally, aECs emerge slightly earlier than vECs in murine embryos[49]. NOTCH, high VEGF and ERK signaling promote aEC differentiation, whereas PI3K and the TF NR2F2 promote vEC differentiation[50]. We observed that high and early induction of ETV2 promoted a distinct EC specification pathway that terminated on ECs with arterial features, whereas lower and later endogenously expressed ETV2 yielded ECs with venous features. This observation suggests that the level or timing of ETV2 during EC specification may be an additional factor that biases arteriovenous differentiation. Whether the level and timing of ETV2 expression participates in arteriovenous lineage choices during normal in vivo vascular development requires further investigation.

Our study identified candidate ETV2-dependent CREs that regulate GTPase genes. These regions are decorated by ETV2, active enhancer marks H3K27ac and H3K4me1, as well as strong peak-to-gene links. Notably, this list of putative ETV2-dependent distal regulatory elements, along with a broader map of ETV2-regulated CREs and their target genes, will be highly valuable for future research on EC specification.

Vascular regeneration is a promising therapeutic strategy for diseases that impair vascular function, including atherosclerosis and diabetes[51]. ETV2 overexpression reprograms iPS cells, iPS cell-derived MPCs, and somatic cells such as fibroblasts into functional ECs[1,15,52]. Although ETV2-directed reprogramming of iPS cell-derived MPCs rapidly and efficiently yields ECs (this study and ref. 1), forced expression of ETV2 in fibroblasts was much less efficient[15,52]. The mechanisms responsible for the difference in reprogramming efficiency are uncertain and might include a less-malleable chromatin environment or altered expression of key cofactors (like REST and GABPA). Investigating these differences will be an important future area of investigation. Our atlas of efficient ETV2-directed MPC-to-EC specification will serve as a roadmap for future work to optimize EC differentiation and reprogramming strategies for therapeutic vascular regeneration.

Methods

Cell culture

The human BJ273 iPS cell line was sourced from Boston Children's Hospital's Stem Cell Core[53]. TRE3G-ETV2 iPS cells were engineered from this parental line transfecting them with piggyBac expression plasmid (SBI, PB210PA-1) and the piggyBac transposon vector depicted in Fig. 1a. Stable cell lines were selected with puromycin (1 μg ml^{-1}), and clones were tested for pluripotency and Dox responsiveness.

iPS cells were maintained in mTeSR1 medium (Stem Cell Technologies, 85850) on Matrigel (Corning, 354277) precoated plates at 37 °C with 5% CO_2, and the medium was changed daily. iPS cells were passaged using TrypLE Select (Life Technologies, 12563011) and seeded with Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Stem Cell Technologies, 72308) for the first day. Cells were frozen using mFreSR (Stem Cell Technologies, 05854) using a CoolCell LX (Corning, 432003) overnight at −80 °C and then in vapor-phase liquid nitrogen for long-term storage.

iPS cell-derived EC differentiation

iPS cells were differentiated into ECs using previously described differentiation medium[1]: Advanced DMEM/F12 (Thermo Fisher, 12634010), 1× GlutaMAX supplement (Thermo Fisher, 35050061) and 60 μg ml^{-1} L-ascorbic acid (Sigma-Aldrich, A8960). For differentiation, iPS cells were dissociated into single cells with TrypLE Select and plated on Matrigel precoated plates at a density of 60,000 iPS cells per cm^2 in mTeSR1 supplemented with 10 μM Y-27632. At day 0, iPS cells were briefly washed with PBS and first differentiated into MPCs using a differentiation medium supplemented with CHIR99021 (6 μM, Stem Cell Technologies, 100-1042) for 48 h. At day 2, MPCs were dissociated and then seeded on a 60-mm Matrigel-coated dish in differentiation medium supplemented with VEGFA (50 ng ml^{-1}, PeproTech, 100-20), fibroblast growth factor 2 (FGF2, 50 ng ml^{-1}, PeproTech, 100-18B), EGF (10 ng ml^{-1}, PeproTech, AF-100-15), and SB431542 (10 μM, Selleckchem, S1067) for another 2 days. For the Dox$^+$ group, cells were treated with Dox (1 μg ml^{-1}, Sigma-Aldrich, D3072) for 24 h. The medium was changed every day throughout this protocol.

RNA isolation and RT–qPCR analysis

Total RNA was extracted using TRIzol reagent (Invitrogen, 15596026) and Zymo Research RNA Clean & Concentrator kits (Zymo Research, ZR1013). Trizol aqueous phase was transferred to a Zymo-Spin IC

Column and treated with DNase I. RNA was reverse transcribed to cDNA using SuperScript III First-Strand Synthesis SuperMix (Thermo Fisher, 11752050). RT–qPCR was carried out on a Bio-Rad CFX96 Touch Deep Well Real-Time PCR instrument using the SYBR Green PCR Master Mix (Thermo Fisher, 4368708). The primer sequences for qPCR analysis are listed in Supplementary Table 13. Relative gene expression was normalized to GAPDH and analyzed by the $2^{-\Delta\Delta Ct}$ method.

Flow cytometry

Cells were dissociated using TrypLE Select, washed and resuspended in FACS buffer (PBS with 0.5% BSA). For surface antigens, live cells were stained with commercially available antibodies (Supplementary Table 14) for 30 min on ice in the dark. For staining TFs, True-Nuclear Transcription Factor Buffer Set (BioLegend, 424401) was used. Stained cells were washed twice with FACS buffer, filtered into a 35-µm strainer-capped tube (Falcon, 352235) and analyzed by a BD LSRFortessa. FACS was performed on a BD FACSMelody Cell Sorter (BD Biosciences). Spectral overlap was compensated for using UltraComp eBead Plus Compensation Beads (Life Technologies, 01-3333-42) with single fluorophore-conjugated antibodies. Flow cytometry data were analyzed using FlowJo software (v.10.7.1).

Purification and expansion of ECs

ECs obtained from differentiation of iPS cells were dissociated using TrypLE Select and sorted using PECAM1-conjugated magnetic microbeads (Miltenyi Biotec, 130-091-935) according to the manufacturer's instructions. The purified PECAM1-positive cells were seeded on 1% gelatin-coated plates and maintained in EGM2 medium (Promocell, C-22111) supplemented with 1× GlutaMAX supplement and 10 µM SB431542. ECs were expanded and used for functional assays.

Shear stress-induced polarization of iPS cell-derived ECs

The resultant ECs were cultured on a 1% gelatin precoated glass-bottom plate with EGM2 medium for 24 h, either in static culture or a rotator. Confluent monolayers of ECs were subjected to orbital shear stress at a rotating frequency of 150 rpm using an orbital shaker positioned inside a cell culture incubator. After 24 h of static or shear stress conditions, immunostaining was performed to quantify the degree of cell alignment. After imaging, quantification of cell orientation angles was performed using ImageJ.

Nitric oxide production assay

To determine the ability of the cells to produce nitric oxide, cells were placed in fresh medium containing 1 µM DAF-FM diacetate (Thermo Fisher, D23844) or vehicle (dimethylsulfoxide (DMSO)). Cells were cultured for 30 min and then collected for flow cytometric analysis. As a negative control, iPS cells were incubated in mTeSR1 containing DMSO or DAF-FM diacetate. Geometric mean fluorescence intensities were calculated using FlowJo.

AcLDL uptake assay

To assess the ability of ECs to uptake AcLDL, ECs were incubated at 37 °C for 4 h with AcLDL conjugated to Alexa Fluor 488 (5 µg ml^{-1}, Thermo Fisher, L23380) in complete EGM2 medium. Following incubation, the ECs were washed with PBS to remove the excess stain and microscopic images were taken. Representative image shown is selected among multiple independent experiments.

Immunofluorescent staining

Immunofluorescent staining was performed as previously described[54]. In brief, cells were washed with PBS followed by fixation with 4% paraformaldehyde for 15 min at room temperature. PBS containing 0.1% Triton X-100 was used for permeabilization. Then, cells were blocked for 1 h in blocking buffer (3% bovine serum albumin in PBS). After blocking, cells were incubated with primary antibodies (Supplementary Table 14) in blocking buffer at 4 °C overnight. Cells were then washed three times followed by incubation with appropriate fluorescent secondary antibodies at room temperature for 2 h. Finally, cells were stained with a 4,6-diamidino-2-phenylindole (DAPI) solution and mounted in ProLong Glass Antifade Mountant (Thermo Fisher, P36982). Images were taken using a Olympus FV3000R confocal microscope. Analysis of microscopy images was carried out using Fiji (ImageJ, v.2.1.0).

Wound-healing scratch assay

Wound-healing scratch assays were performed as described previously[55]. ECs were seeded into six-well plates and grown to reach about 90% confluency on the day of the experiment. Monolayers were scratched using a pipette tip and washed three times with complete EGM2 medium. Scratches were imaged using a Keyence automated epifluorescent microscope equipped with a ×10 objective. Cells were imaged after 8 and 16 h to determine their capacity to migrate in the cell free gap.

Co-immunoprecipitation

Cells were lysed with Pierce IP Lysis Buffer (Thermo Scientific, 87787) supplemental with Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific, 78440). The lysates were centrifuged for 30 min at 16,000*g* at 4 °C, and the supernatants were transferred. Then, 50 µl of lysate was used as the input. Co-IP was performed using the Dynabeads Protein G (Life Technologies, 10004D) following the manufacturer's protocol. In brief, Dynabeads Protein G were incubated with primary antibody (Supplementary Table 14) or IgG at 4 °C overnight. The antibody was crosslinked to the magnetic beads before immunoprecipitation using the crosslinking reagent BS_3. Approximately 1 mg of cell lysate was added to the magnetic bead–antibody complex and incubated with rotation for 4 °C overnight. The magnetic bead–antibody–antigen complex was pulled down and proteins were eluted with 0.1 M glycine–HCl, pH 2.8. The pH of samples was adjusted by adding 1 M Tris, pH 8.0.

Immunoblotting

Total protein was extracted from cells using RIPA lysis buffer (Santa Cruz Biotechnology, SC-24948). The lysates were centrifuged for 30 min at 16,000*g* at 4 °C, and the supernatants were transferred. Protein concentration was determined using the Pierce BCA Protein Assay kit (Life Technologies, 23227). Sample preparation was performed following the default Simple Western protocol. Protein separation and immunoassay were conducted with the automated WES capillary-based electrophoresis system (WES; ProteinSimple). Target proteins were identified using primary antibodies (Supplementary Table 14) and immunoprobed using an HRP-conjugated secondary antibody (ProteinSimple). Chemiluminescent signals were detected and analyzed using Compass for SW (v.6.1.0) software.

Lentivirus preparation and transduction

All lentiviral plasmids were prepared to the PureLink HiPure Plasmid Filter Midiprep kit (Life Technologies, K210014). Viruses were packaged in 293T cells by co-transduction with lenti-Cas9-GFP (Addgene, 63592) or lenti-sgRNA-mCherry with psPAX2 (Addgene, 12260) and pMD2.G (Addgene, 12259) using polyethyleneimine. The culture medium was exchanged after 12 h and supernatants were collected 48 h and 72 h after transfection. The supernatants were passed through a 0.45-µm PVDF filter (EMD Millipore, SLHVM33RS) and pooled. The virus particles were concentrated by ultracentrifugation, resuspended in PBS and kept at −80 °C in small aliquots. Virus titers were determined with a qPCR Lentivirus Titration kit (Applied Biological Materials, LV900). iPS cells were transduced with different amounts of lentivirus depending on the objective of the experiment. Protamine sulfate (4 µg ml^{-1}) was used to enhance infection efficiency. After 6 h of incubation with lentivirus, cells were washed to remove residual lentivirus.

CRISPR screening library construction

pCF221 (Addgene, 121669) plasmid encoding the U6-sgRNA and mCherry was digested with BsmBI (NEB, R0580) and gel extracted (Zymo Research, D4008). Six sgRNAs were designed using Benchling (www.benchling.com) against individual exons of each gene. All 3,150 sgRNA oligonucleotides (Supplementary Table 15) with adaptor sequences, including 90 non-targeting control sgRNAs, were synthesized by Twist Bioscience and pooled in equal molarity. The pooled sgRNA oligonucleotides were then amplified by PCR and cloned into the BsmBI-digested pCF221 vector using HiFi DNA Assembly Master Mix (NEB, E2621S). Ligated vectors were electroporated into Endura Electro Competent Cells (Lucigen, 60242-1). Two electroporations collectively yielded ~400,000 colonies (more than 100× coverage). Colonies were scraped into LB medium, cultured for an additional 5 h and plasmid DNA was collected by ZymoPURE II Plasmid Purification kit, Maxiprep (Zymo Research, D4202). Deep sequencing was performed to verify the identity and representation of all sgRNAs in the pooled plasmids. The library pool was transfected into HEK293T cells with psPAX2 and pMD2.G. The culture medium was collected and purified to yield the concentrated lentiviral library.

Pooled CRISPR screening and sample collection

iPS cells were first infected with lenti-Cas9-GFP at multiplicity of infection ~0.7 to ensure stable expression of Cas9 components. After FACS selection, GFP positive cells were infected with lenti-sgRNA-mCherry at multiplicity of infection ~0.3 to that each sgRNA was introduced into ~1,000 cells to maximize sensitivity. Three days after infection, GFP and mCherry double-positive iPS cells were selected by FACS. The transduced iPS cell populations were seeded into 10-cm dishes in mTeSR1 with 10 μM Y-27632. The next day, the medium was replenished with differentiation medium following the Dox differentiation protocol. On D2, MPCs were dissociated and then seeded. A subset of the cells was collected as the reference time point (input). On D4, cells were dissociated using TrypLE Select. Typically, cells were collected and stained for anti-PECAM1 and anti-CDH5 for 30 min. Cells were then sorted on a BD FACSMelody Cell Sorter (BD Biosciences). Approximately 4 million ECs ($PECAM1^+CDH5^+$) were collected through FACS for each biological replicate in the screen. Cells were subsequently spun down and the cell pellet was collected for genomic DNA extraction, amplification of sgRNA and deep-sequencing analysis.

Genomic extraction and screening library preparation

Genomic DNA was extracted from sorted cells using a Quick-DNA Miniprep Plus kit (Zymo Research, D4068). sgRNA sequencing library construction was performed using a two-step protocol. In brief, sgRNA cassette in genomic DNA was amplified by Q5 High-Fidelity DNA Polymerase (NEB, M0491S) followed by purification. Then the PCR product (200 bp) was used as input for a second step introducing the standard Illumina adapters using an NEBNext UltraDNA Library Prep kit (NEB, E7645S). The final Illumina libraries were pooled and subjected to high-throughput sequencing analysis. The read count of each sgRNA was calculated with no mismatches and aligned to the sequence of reference sgRNA using a Python script with modifications[56]. Data were analyzed using MAGeCK software[57] with default parameters. Genes were ranked by the $-\log_{10}$(robust ranking aggregation score). In each group, four biological replicates were used in MAGeCK analysis.

CRISPR-Cas9 genome editing of iPS cells

In brief, the three most effective guide RNAs targeting each gene in our CRISPR screening experiment were chosen for individual gene knockout. The sgRNA oligonucleotides were ordered from Integrated DNA Technologies and cloned individually into pCF221 (Addgene, 121669). Each completed sgRNA vector was verified by Sanger sequencing. Then, plasmids containing the guide RNAs were co-transfected with psPAX2 and pMD2.G into HEK293T to generate the sgRNA lentivirus. iPS cells were first transduced with lenti-Cas9-EGFP, followed by transduction with sgRNA-mCherry lentivirus. Three days after transfection, GFP^+ and $mCherry^+$ cells were sorted using a BD FACSMelody Cell Sorter (BD Biosciences) and cultured for further differentiation. Genomic DNA from cell lysate was used for PCR genotyping and validated using primers flanking the expected deletion site. sgRNA target sequences and primers used for PCR and sequencing are listed in Supplementary Table 13.

For enhancer deletion, two sgRNAs flanking enhancer 1 or enhancer 2 of RASGRP2 were designed using Benchling to generate deletions of enhancer regions (~500 bp), which centered around an ETV2 peak. sgRNAs were designed to delete neighboring regions as a negative control. The sgRNAs were cloned into the pCF221 plasmid and used for lentivirus generation. The Cas9-expressing iPS cells were transduced with the pairs of sgRNA lentivirus to generate the enhancer deletion iPS cells.

Bulk RNA-seq and data analysis

The NEBNext Ultra II Directional RNA Library Prep kit for Illumina (NEB, E7760) was used for RNA sequencing library construction according to the manufacturer's instructions. Libraries were then sequenced on an Illumina NovaSeq6000. Quality control was performed using FastQC (v.0.11.9). The reads were trimmed using TrimGalore (v.0.6.7) to remove detected adaptors. Reads were mapped to hg38 reference with Hisat2 (v.2.1.0)[58]. The expression count matrix of each sample was calculated using FeatureCounts (v.2.0.1) of the Subread package[59]. The gene counts were normalized using DESeq2 (ref. 60) for visualization and differential expression analyses. Comparison to RNA-seq of cell lines was performed using data from the Gene Expression Omnibus (GSE138734)[19]. Batch effects were removed using limma[61].

Bulk ATAC-seq and data analysis

ATAC-seq were performed as described previously[62]. Libraries were then sequencing on an Illumina NextSeq 550. ATAC-seq data analysis used the following tools. Quality control was performed using FastQC (v.0.11.9). The reads were trimmed using TrimGalore (v.0.6.7) to remove detected adaptors. These reads were uniquely aligned to the reference genome hg38 using Bowtie2 (v.2.3.4.3). The mapped reads were filtered to retain reads with mapping quality of at least 30 and remove mitochondrial reads by SAMtools (v.1.13). Picard tools were then used to remove duplicate reads, and peaks were identified using MACS2 (v.2.1.1).

CUT&RUN

CUT&RUN was performed following the EpiCypher CUTANA protocol. Concanavalin A beads were activated with cold Bead Activation Buffer. Then, 500,000 cells per reaction were collected followed by two washes in the RT Wash Buffer. Cells were incubated with concanavalin A beads then resuspend in a cold Antibody Buffer. Then, 1 μg of primary antibody (Supplementary Table 14) or IgG (Epicypher, 13-0042) was added per reaction and incubated overnight at 4 °C. Cold Digitonin Buffer and pAG-MNase (Epicypher, 15-1016) were added to bind to antibody-tagged chromatin. Subsequently, targeted chromatin was digested by the addition of 100 mM $CaCl_2$. Stop Buffer was added to stop MNase enzyme activity. The fragmented chromatin was purified using a CUTANA DNA Purification kit (EpiCypher, 14-0050). For low input cell numbers, the purified DNA was collected by phenol–chloroform followed by ethanol purification. The final product was quantified with a Qubit fluorometer.

The library preparation used the NEBNext Ultra II DNA Library Prep kit for Illumina (NEB, E7645). In brief, less than 30 ng of fragmented DNA was used for library construction. The product was cleaned up with SPRIselect beads (Beckman Coulter, B23317). The final products were quantified with Qubit and Agilent Tapestation. Libraries were subsequently sequenced on an Illumina NovaSeq6000.

CUT&RUN data analysis

For CUT&RUN sequencing data, raw paired-end reads were trimmed and uniquely aligned to the reference genome hg38 using Bowtie2 (v.2.3.4.3)[63]. The mapped reads were filtered to retain reads with mapping quality of at least 30 and remove mitochondrial reads by SAMtools (v.1.13). Spurious signals were removed by subtracting pre-packaged 'suspect lists' regions[64,65]. After removal of PCR duplication reads by Picard tools MarkDuplicates (v.2.27.4), peaks were identified using MACS2 (v.2.1.1)[66] and using the IgG control sequencing data as background, with the 'narrow peak' option for TFs and H3K4me3 and the 'broad peak' option for H3K4me1. The signal tracks were generated by bamCoverage in Deeptools (v.3.5.2)[67]. Heatmap density plots and metagene plots were generated using ±2 kb or ±5 kb around the peak center by Deeptools. Peaks were compared between replicates, and only the intersection of replicates were kept for further analyses using BEDTools (v.2.30.0)[68]. The annotatePeaks and findMotifsGenome tools with default parameters from HOMER[69] were used to annotate peaks and motif enrichment analysis, respectively. Intersecting peaks between different TFs were also identified using BEDTools, and heatmaps were generated using Deeptools. The genome tracks of peaks were visualized by the IGV software (v.2.12.3)[70].

scRNA-seq library preparation

Cells were dissociated by TrypLE Select for 5 min at 37 °C. Cells were centrifuged at 300*g* for 5 min at 4 °C to remove the debris in the supernatant. The pellets were further washed in chilled 1× DPBS (Life Technologies, 14190136) using a wide-bore pipette tip and centrifuged at 300*g* for 5 min. One million cells were gently resuspended in 200 µl chilled 1× DPBS, then 800 µl chilled methanol was added drop by drop. Cells were incubated for 30 min at −20 °C. Cells were kept at −80 °C and used within 1 month.

To continue with scRNA-seq, fixed cells were put on ice for 5 min and centrifuged at 1,000*g* for 5 min at 4 °C. Cells were gently resuspended with 200 µl 3× SSC Buffer (Sigma-Aldrich, S6639) supplemented with 0.04% BSA (Life Technologies, AM2616), 1 mM dithiothreitol (Sigma-Aldrich, 43816) and 0.2 U μl^{-1} RNase Inhibitor (Roche Diagnostics, 3335399001) and filtered through a 40-µm Flowmi Cell Strainer (Bel-Art, H13680-0040) to obtain single cells. The cell suspension from all time points was adjusted to 2,500 cells per µl using the Countess II Automated Cell Counter. scRNA-seq libraries were generated using the 10x Genomics Chromium Single Cell 3′ Library v.3.1 (PN-1000268), Chip G Single Cell kit (PN-1000120) and Dual Index kit TT Set A (PN-1000215) following the manufacturer's protocol. In brief, approximately 10,000 cells in less than 4 µl suspension buffer to minimize buffer carryover were added to Master Mix and then loaded into a Chromium Controller. Libraries were assessed using an Agilent Tapestation and quantified using a KAPA Library Quantification kit (Roche, KK4824). Libraries were subsequently sequenced on an Illumina NovaSeq6000 with a sequencing depth of at least 50,000 reads per cell with paired-end 150 bp.

scATAC-seq library preparation

Cells were dissociated by TrypLE Select for 5 min at 37 °C. Cells were centrifuged at 300*g* for 5 min at 4 °C to remove the debris in the supernatant. The cells were gently resuspended with chilled mFreSR and frozen using a CoolCell LX overnight at −80 °C. Cells were kept in vapor-phase liquid nitrogen for long-term storage.

To continue with scATAC-seq, cryovials were quickly thawed in a water bath at 37 °C for 1 min and centrifuged at 300*g* for 5 min. Cells were gently resuspended with 1 ml 1× PBS supplemented with 0.04% BSA twice and filtered through a 40-µm Flowmi Cell Strainer to obtain single cells. Nuclei isolation was performed following the protocol from 10x Genomics (CG000169). scATAC-seq libraries were generated using the 10x Genomics Chromium Single Cell ATAC Kit v2 (PN-1000390), Chip H Single Cell kit (PN-1000161) and Single Index kit N Set A (PN-1000212) following the manufacturer's protocol. Approximately 10,000 nuclei were loaded into a Chromium Controller. Libraries were assessed using an Agilent Tapestation and quantified using a KAPA Library Quantification kit (Roche, KK4824). Libraries were subsequently sequenced on an Illumina NovaSeq6000 (paired-end 50 bp, sequencing depth >50,000 reads per cell).

scRNA-seq data analysis

Data processing. FASTQ files were processed with the 10x Genomics Cell Ranger pipeline (v.6.1.2). The raw reads were individually aligned to human reference genome GRCh38 (reference 2020-A) and quantified unique molecular identifiers counted by 'cellranger count' for each gene in each cell. The count matrix of all samples was imported into the R environment (v.4.1) using the function 'Read10X' in Seurat R package (v.4.1.0)[71]. We removed cells with <1,200 genes, >8,000 genes or >10% mitochondrial reads. Only high-quality single-cell transcriptomes were used for subsequent analysis. The count matrix was classified into the Dox^+ and Dox^- group with a total 105,563 cells retained. A list of the number of cells in each sample is provided in Supplementary Table 16.

Data integration and clustering. Standard Seurat data processing and normalization steps were performed for each group: SCTransform and RunPCA. A total of 3,000 genes were selected as integration features and used to identify 'integration anchors'. The data from both groups were integrated using IntegrateData functions in Seurat for batch correction. Then, we performed Seurat function RunPCA, FindNeighbors, FindClusters and RunUMAP on the integrated data. UMAP embedding was computed for the integrated data using the top 40 principal components. Integration of transcriptomes from all time points or D4 of Dox^- and Dox^+ was performed. DoubletFinder[72] was used with default parameters to remove potential doublets. The overlapping embedding of Dox^- and Dox^+ on D2 suggested effective removal of batch effects. Expression of marker genes was denoted by different color shades and superimposed over the entire cell population, as visualized on a UMAP projection. Gene-set signature scores (identity scores) were computed using the AddModuleScore function in Seurat.

Differential gene expression tests. Each cluster's DEGs were identified using the FindAllMarkers function with default parameters (logfc.threshold set to 0.25). For pairwise comparisons, DEGs were identified using the FindMarkers functions. Selected marker genes were displayed using the Dotplot function. All heatmaps were generated using Seurat's DoHeatmap function.

scATAC-seq data analysis

Data processing and clustering. FASTQ files were processed with the 10x Genomics Cell Ranger ATAC pipeline (v.2.1.0). Raw reads were individually aligned to the human reference genome GRCh38 (reference 2020-A). The command 'cellranger-atac count' was used to obtain the fragment file for each sample. Downstream analysis was performed in R (v.4.2.0), Seurat (v.4.1.0), Signac (v.1.9.0)[73] and ArchR (v.1.0.2)[74]. The union peak list was generated using all peak lists. The barcoded fragments files were used as input data for analysis using 'FeatureMatrix' and 'CreateChromatinAssay' functions in Signac. The different samples were pooled together. Quality-control metrics were calculated and cells were filtered as follows: number of fragments, 1,000–30,000; blacklist_fraction <0.05; nucleosome_signal <4; and transcription start site enrichment score >2. In addition to Signac metrics, ArchR was also used to perform quality control and identify doublets. The matrix was classified into Dox^- and Dox^+ with a total 59,785 cells retained. A list of the number of cells in each sample is provided in Supplementary Table 16. Standard iterative latent semantic indexing dimension reduction and clustering was performed as follows using default parameters: RunTFIDF, FindTopFeatures, RunSVD, RunUMAP, FindNeighbors and FindClusters. UMAP was used to visualize cell embedding for all cells.

Integrative analysis of scATAC-seq and scRNA-seq. Seurat's canonical correlation analysis was used to align scRNA-seq and scATAC-seq datasets by comparing scATAC-seq gene activity with the scRNA-seq gene expression matrix. For this purpose, gene activity in the cells profiled by scATAC-seq was calculated using the 'GeneActivity' function in Signac followed by Normalize Data with the LogNormalize method. The highly variable genes from the scRNA-seq dataset were used as input to 'FindTransferAnchors' function in Seurat with reduction method 'cca' to identify integration anchors. We then assigned each of the scATAC-seq cells a cell-type subcluster identity from the matching scRNA-seq data and an associated label prediction score using the 'TransferData' functions in Seurat with default parameters. To ensure the prediction accuracy, only the cells with a prediction score higher than 0.5 were retained for further analysis. Finally, the annotation of scATAC-seq clusters was defined based on the identity of its cells.

Downstream analysis was performed using the ArchR package, unless otherwise noted. Gene scores were calculated for each Arrow file and saved in GeneScoreMatrix using the default gene score model in ArchR[74]. The GeneScoreMatrix and GeneIntegrationMatrix in scATAC-seq were created after the Seurat label transfer procedure. To identify the marker genes in scATAC-seq, gene expression in each cell type was calculated using Wilcoxon test with $|\log_2$ fold change$| > 1$ and false discovery rate (FDR) < 0.05. Peak-to-gene correlation analysis for the feature genes was computed to identify putative regulatory relationships by correlating peak accessibility to GeneIntegrationMatrix across scATAC-seq metacells. This procedure was invoked using the 'addPeak2GeneLinks' function. Only links with correlation >0.45 and FDR $< 1 \times 10^{-4}$ were visualized (Supplementary Table 3). Genome coverage tracks displaying the pseudo-bulk coverage patterns were generated using 'plotBrowserTrack' function. We used the chromVAR pipeline[24] to calculate enrichment of chromatin accessibility over all active motif instances of each TF at single-cell resolution with 'addDeviationsMatrix'. Trajectory analysis was performed using the ArchR wrapper functions. The heatmaps were plotted using plotTrajectory-Heatmap function.

Identification of differentially accessible peaks. scATAC-seq data were used to generate pseudo-bulk replicates based on cluster identities. Pseudo-bulk peak calling was performed within each cluster using MACS2 (ref. 66) using the 'addReproduciblePeakSet'. ArchR's default iterative overlap peak merging procedure was then applied to merge all peak calls and generate a reproducible peak set. Differentially accessible marker peaks in each cluster were identified by comparing each cluster cell using Wilcoxon test with $|\log_2$ fold change$| > 1$ and FDR < 0.01. For pairwise comparisons, the differentially accessible peaks were calculated using the Wilcoxon test with with $|\log_2$ fold change$| > 1$ and FDR < 0.05. After we identified cell-type-specific marker peaks, we then computed the enrichment of active motif instances of all TFs expressed in marker peaks using the 'peakAnnoEnrichment' function. GO analysis of regions was performed using GREAT[75] with its basal plus extension option and default parameters.

Statistics

Statistical analyses were performed in R or Prism v.10.0.2 (GraphPad). Results are presented as the mean ± s.e.m. or mean ± s.d. (error bars). Statistical tests and the numbers of independent biological replicates for each experiment are described in the figure legends. Statistical *P* values were calculated and reported on graphs. $P < 0.05$ was considered statistically significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing and related processed data have been deposited in the Gene Expression Omnibus under accession number GSE267565. Source Data are provided with this publication. Other information is available from the corresponding author upon reasonable request.

Code availability

No custom code was used in this study. Analysis was carried out with publicly available pipelines using approaches described in Methods.

References

1. Wang, K. et al. Robust differentiation of human pluripotent stem cells into endothelial cells via temporal modulation of ETV2 with modified mRNA. *Sci. Adv.* https://doi.org/10.1126/sciadv.aba7606 (2020).
2. Lai, W. K. M. & Pugh, B. F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* **18**, 548–562 (2017).
3. Barral, A. & Zaret, K. S. Pioneer factors: roles and their regulation in development. *Trends Genet.* **40**, 134–148 (2024).
4. Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).
5. Zhou, P. et al. GATA4 regulates developing endocardium through interaction with ETS1. *Circ. Res.* https://doi.org/10.1161/CIRCRESAHA.120.318102 (2022).
6. Swinstead, E. E. et al. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165**, 593–605 (2016).
7. Liu, Z. & Kraus, W. L. Catalytic-independent functions of PARP-1 determine Sox2 pioneer activity at intractable genomic loci. *Mol. Cell* **65**, 589–603.e9 (2017).
8. Sunkel, B. D. & Stanton, B. Z. Pioneer factors in development and cancer. *iScience* **24**, 103132 (2021).
9. Oh, S. Y., Kim, J. Y. & Park, C. The ETS factor, ETV2: a master regulator for vascular endothelial cell development. *Mol. Cell* **38**, 1029–1036 (2015).
10. Liu, F. et al. Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO Rep.* **16**, 654–669 (2015).
11. Gong, W. et al. ETV2 functions as a pioneer factor to regulate and reprogram the endothelial lineage. *Nat. Cell Biol.* **24**, 672–684 (2022).
12. Koyano-Nakagawa, N. & Garry, D. J. Etv2 as an essential regulator of mesodermal lineage development. *Cardiovasc. Res.* **113**, 1294–1306 (2017).
13. Lee, D. et al. ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. *Cell Stem Cell* **2**, 497–507 (2008).
14. Ferdous, A. et al. Nkx2-5 transactivates the Ets-related protein 71 gene and specifies an endothelial/endocardial fate in the developing embryo. *Proc. Natl Acad. Sci. USA* **106**, 814–819 (2009).
15. Lee, S. et al. Direct reprogramming of human dermal fibroblasts into endothelial cells using ER71/ETV2. *Circ. Res.* **120**, 848–861 (2017).
16. Ng, A. H. M. et al. A comprehensive library of human transcription factors for cell fate engineering. *Nat. Biotechnol.* **39**, 510–519 (2021).
17. Abedin, M. J. et al. Fli1 acts downstream of Etv2 to govern cell survival and vascular homeostasis via positive autoregulation. *Circ. Res.* **114**, 1690–1699 (2014).
18. Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.* **10**, 81 (2010).

19. Lorenzi, L. et al. The RNA Atlas expands the catalog of human non-coding RNAs. *Nat. Biotechnol.* **39**, 1453–1465 (2021).
20. Ginsberg, M. et al. Efficient direct reprogramming of mature amniotic cells into endothelial cells by ETS factors and TGFβ suppression. *Cell* **151**, 559–575 (2012).
21. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
22. Chen, J. et al. Molecular and spatial signatures of mouse embryonic endothelial cells at single-cell resolution. *Circ. Res.* **134**, 529–546 (2024).
23. You, L.-R. et al. Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity. *Nature* **435**, 98–104 (2005).
24. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
25. Corada, M. et al. Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nat. Commun.* **4**, 2609 (2013).
26. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
27. Galceran, J., Hsu, S. C. & Grosschedl, R. Rescue of a Wnt mutation by an activated form of LEF-1: regulation of maintenance but not initiation of Brachyury expression. *Proc. Natl Acad. Sci. USA* **98**, 8668–8673 (2001).
28. Russ, A. P. et al. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95–99 (2000).
29. Li-Villarreal, N. et al. FOXO1 represses sprouty 2 and sprouty 4 expression to promote arterial specification and vascular remodeling in the mouse yolk sac. *Development* https://doi.org/10.1242/dev.200131 (2022).
30. Szklarczyk, D. et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
31. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *eLife* https://doi.org/10.7554/elife.46314 (2019).
32. Jolma, A. et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384–388 (2015).
33. Zhou, P. et al. Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *eLife* https://doi.org/10.7554/elife.22039 (2017).
34. De Val, S. et al. Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. *Cell* **135**, 1053–1064 (2008).
35. Usuba, R., Pauty, J., Soncin, F. & Matsunaga, Y. T. EGFL7 regulates sprouting angiogenesis and endothelial integrity in a human blood vessel model. *Biomaterials* **197**, 305–316 (2019).
36. Lee, H. K., Chauhan, S. K., Kay, E. & Dana, R. Flt-1 regulates vascular endothelial cell migration via a protein tyrosine kinase-7-dependent pathway. *Blood* **117**, 5762–5771 (2011).
37. Zhu, S. et al. Guide RNAs with embedded barcodes boost CRISPR-pooled screens. *Genome Biol.* **20**, 20 (2019).
38. Naruse, Y., Aoki, T., Kojima, T. & Mori, N. Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proc. Natl Acad. Sci. USA* **96**, 13691–13696 (1999).
39. Huang, Y., Myers, S. J. & Dingledine, R. Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nat. Neurosci.* **2**, 867–872 (1999).
40. May, D. et al. Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* **44**, 89–93 (2011).
41. Ameen, M. et al. Integrative single-cell analysis of cardiogenesis identifies developmental trajectories and non-coding mutations in congenital heart disease. *Cell* **185**, 4937–4953.e23 (2022).
42. Kataoka, H. et al. Etv2/ER71 induces vascular mesoderm from Flk1^{+}PDGFRα^{+} primitive mesoderm. *Blood* **118**, 6975–6986 (2011).
43. Resendes, K. K. & Rosmarin, A. G. GA-binding protein and p300 are essential components of a retinoic acid-induced enhanceosome in myeloid cells. *Mol. Cell. Biol.* **26**, 3060–3070 (2006).
44. Choi, J. & Lee, E. A. Analysis of REST binding sites with canonical and non-canonical motifs in human cell lines. *BMC Med. Genomics* **17**, 92 (2024).
45. Paik, D. T. et al. Large-scale single-cell RNA-seq reveals molecular signatures of heterogeneous populations of human induced pluripotent stem cell-derived endothelial cells. *Circ. Res.* **123**, 443–450 (2018).
46. Rasmussen, T. L. et al. ER71 directs mesodermal fate decisions during embryogenesis. *Development* **138**, 4801–4812 (2011).
47. Liu, F. et al. ER71 specifies Flk-1^{+} hemangiogenic mesoderm by inhibiting cardiac mesoderm and Wnt signaling. *Blood* **119**, 3295–3305 (2012).
48. Matsui, S. et al. Pioneer and PRDM transcription factors coordinate bivalent epigenetic states to safeguard cell fate. *Mol. Cell* **84**, 476–489.e10 (2024).
49. Chong, D. C., Koo, Y., Xu, K., Fu, S. & Cleaver, O. Stepwise arteriovenous fate acquisition during mammalian vasculogenesis. *Dev. Dyn.* **240**, 2153–2165 (2011).
50. Fish, J. E. & Wythe, J. D. The molecular regulation of arteriovenous specification and maintenance. *Dev. Dyn.* **244**, 391–409 (2015).
51. Samuel, R., Duda, D. G., Fukumura, D. & Jain, R. K. Vascular diseases await translation of blood vessels engineered from stem cells. *Sci. Transl. Med.* https://doi.org/10.1126/scitranslmed.aaa1805 (2015).
52. Morita, R. et al. ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. *Proc. Natl Acad. Sci. USA* **112**, 160–165 (2015).
53. Luo, A. C. et al. A streamlined method to generate endothelial cells from human pluripotent stem cells via transient doxycycline-inducible ETV2 activation. *Angiogenesis* https://doi.org/10.1007/s10456-024-09937-5 (2024).
54. Chen, D. Y. et al. Endothelium-derived semaphorin 3G attenuates ischemic retinopathy by coordinating β-catenin-dependent vascular remodeling. *J. Clin. Invest.* https://doi.org/10.1172/jci135296 (2021).
55. Chen, D. Y. et al. GPR124 facilitates pericyte polarization and migration by regulating the formation of filopodia during ischemic injury. *Theranostics* **9**, 5937–5955 (2019).
56. Joung, J. et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
57. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
58. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
59. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).
61. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
62. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

63. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
64. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).
65. Nordin, A., Zambanini, G., Pagella, P. & Cantù, C. The CUT&RUN suspect list of problematic regions of the genome. *Genome Biol.* **24**, 185 (2023).
66. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol*. https://doi.org/10.1186/gb-2008-9-9-r137 (2008).
67. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
68. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
69. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
70. Robinson, J. T. et al. Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24 (2011).
71. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
72. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
73. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
74. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
75. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

Acknowledgements

This work was supported by the National Institutes of Health (R01HL151450 to W.T.P. and J.M.M.-M.) and American Heart Association (23POST1019621 to D.C.).

Author contributions

D.C. and W.T.P. conceived of and designed the experiments, interpreted results, and wrote the paper with contributions from all authors. D.C. performed the majority of the experiments, library preparation and bioinformatic analyses. X.F. assisted with cell culture and FACS experiments. N.S. assisted with CRISPR screening and library preparation. K.W., L.G. and J.M.M.-M. provided reagents, resources and interpreted results.

Competing interests

K.W. and J.M.M.-M. are inventors on a pending patent related to this work, filed by the Children's Medical Center Corporation (application no. 16/885,999, filed 28 May 2020). The patent covers the transient activation of ETV2 in mesodermal intermediates as a strategy to generate functionally competent ECs from iPS cells. The other authors declare no competing interests.

Additional information

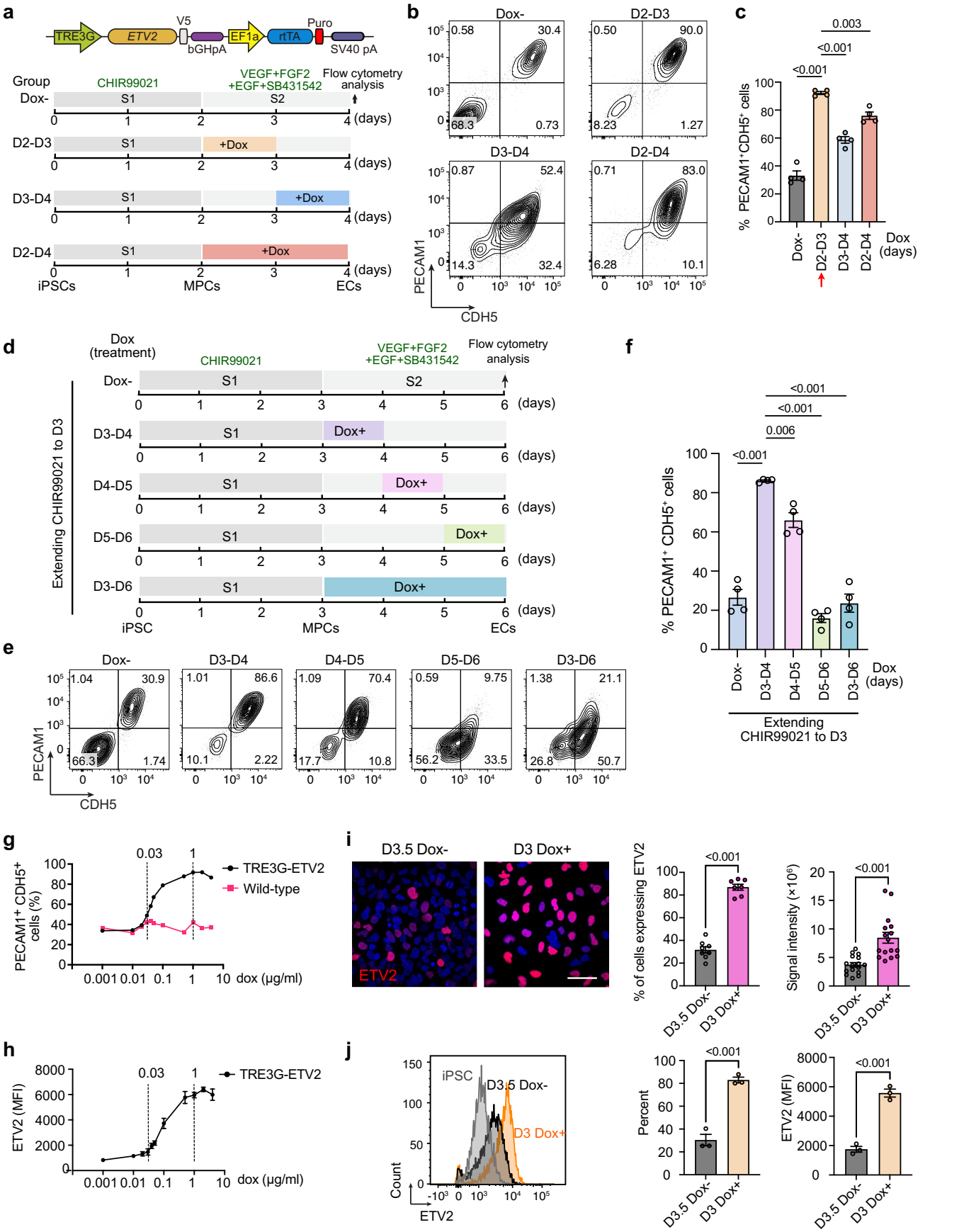
Extended data is available for this paper at https://doi.org/10.1038/s44161-025-00660-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44161-025-00660-y.

Correspondence and requests for materials should be addressed to William T. Pu.

Peer review information *Nature Cardiovascular Research* thanks Claudio Cantù and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s44161-025-00660-y

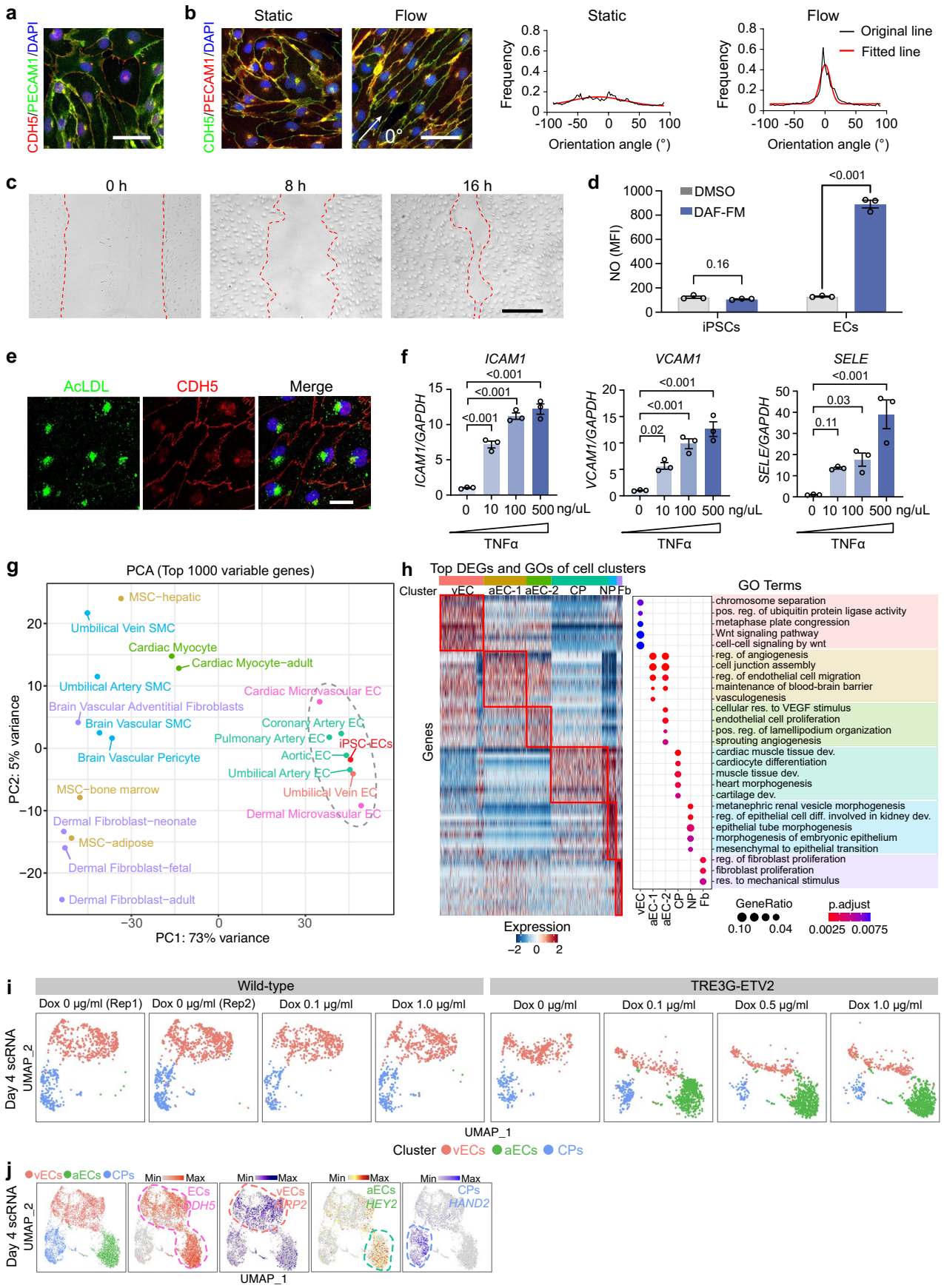


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s44161-025-00660-y

Extended Data Fig. 1 | Highly efficient and rapid EC differentiation from mesodermal progenitors by ETV2 induction. a, Overview of the TRE3G-ETV2 system (upper panel) and timeline of 4 day EC differentiation protocol. Cells were treated without Dox (Dox−) or treated with Dox (1 μg/ml) for the indicated length of time. **b,c**, Flow cytometry analysis (**b**) and quantification (**c**) for PECAM1 and CDH5 expression at day 4 of differentiation. n = 4 independent biological replicates. One-way ANOVA with Tukey's multiple comparison test. **d**, Scheme of 6 day EC differentiation protocol showing timing of Dox addition. Cells were untreated or treated with Dox (1 μg/ml) for indicated length of time. **e,f**, Flow cytometry analysis (**e**) and quantification (**f**) for PECAM1 and CDH5 expression at Day 6 of differentiation. n = 4 independent biological replicates. One-way ANOVA with Tukey's multiple comparison test. **g**, TRE3G-ETV2 iPS cells and wild-type parental iPS cells treated with different dose of Dox at day 2 for 24 hours were analyzed for flow cytometry of PECAM1$^+$ CDH5$^+$ ECs at Day 4. **h**, TRE3G-ETV2 iPS cells treated with different dose of Dox at day 2 for 24 hours were analyzed for flow cytometry of ETV2 expression at Day 3. n = 3 independent biological replicates. **i**, Representative immunofluorescence for ETV2 in control conditions at day 3.5 of differentiation (D3.5 Dox−) or 24 h after doxycycline induction at day 3 of differentiation (D3 Dox +). Scale bars, 50 μm. For quantification, n = 8 (D3.5 Dox−) or 7 (D3 Dox +) independent biological replicates. Level of ETV2 expression in each condition as measured by the immunofluorescence signal intensity. n = 16 representative images per condition. Unpaired two-tailed Student's t-test. **j**, Left panel, histogram of ETV2 expression levels of iPS cells, D3.5 Dox− cells, and D3 Dox+ cells. Right panel, quantification of ETV2$^+$ cells in D3.5 Dox− and D3 Dox+ condition. n = 3 independent biological replicates. Unpaired two-tailed Student's t-test. MFI, mean fluorescence intensity. In panels **c,f,h,i,j**, data are shown as mean ± s.e.m.

https://doi.org/10.1038/s44161-025-00660-y

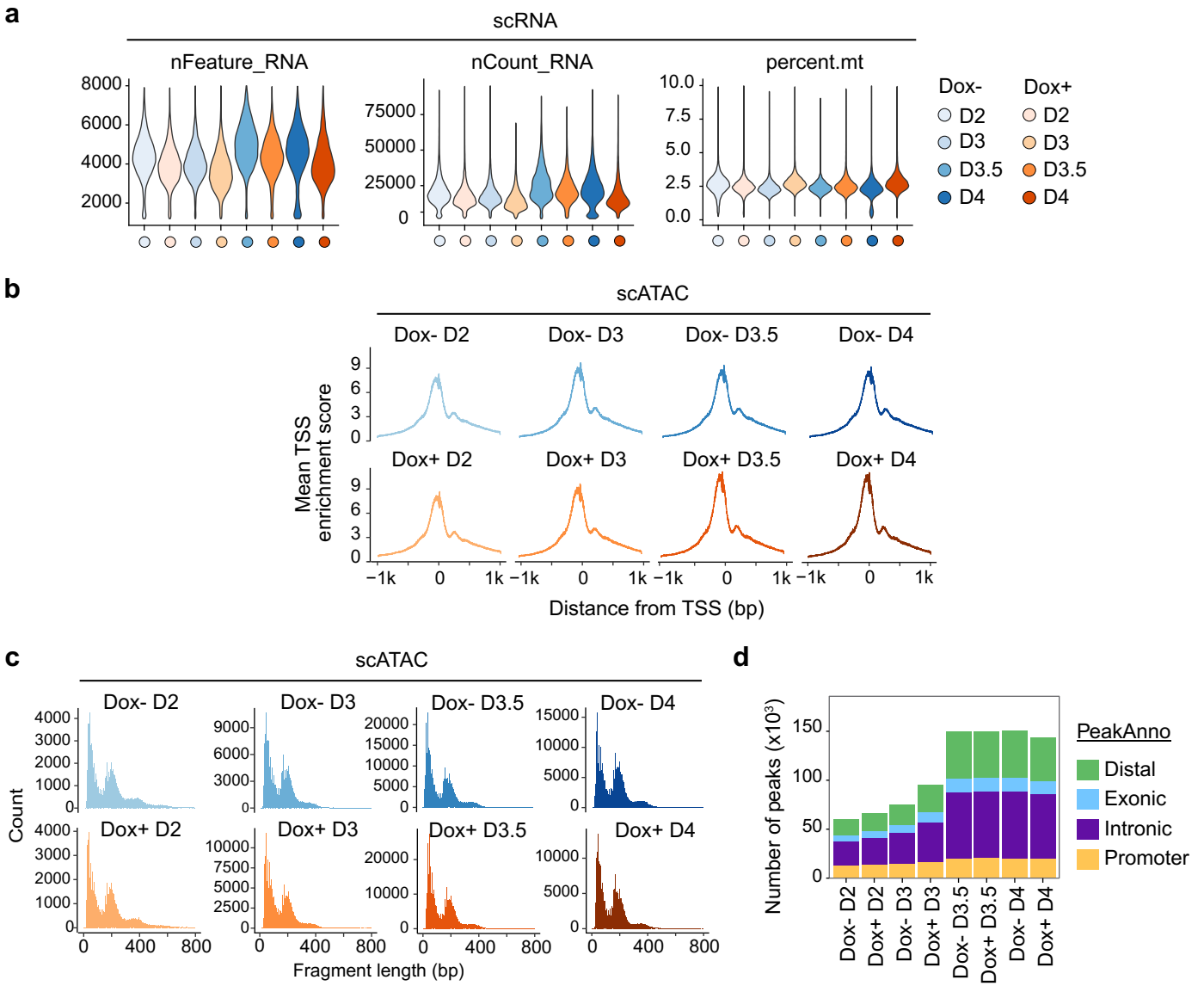


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s44161-025-00660-y

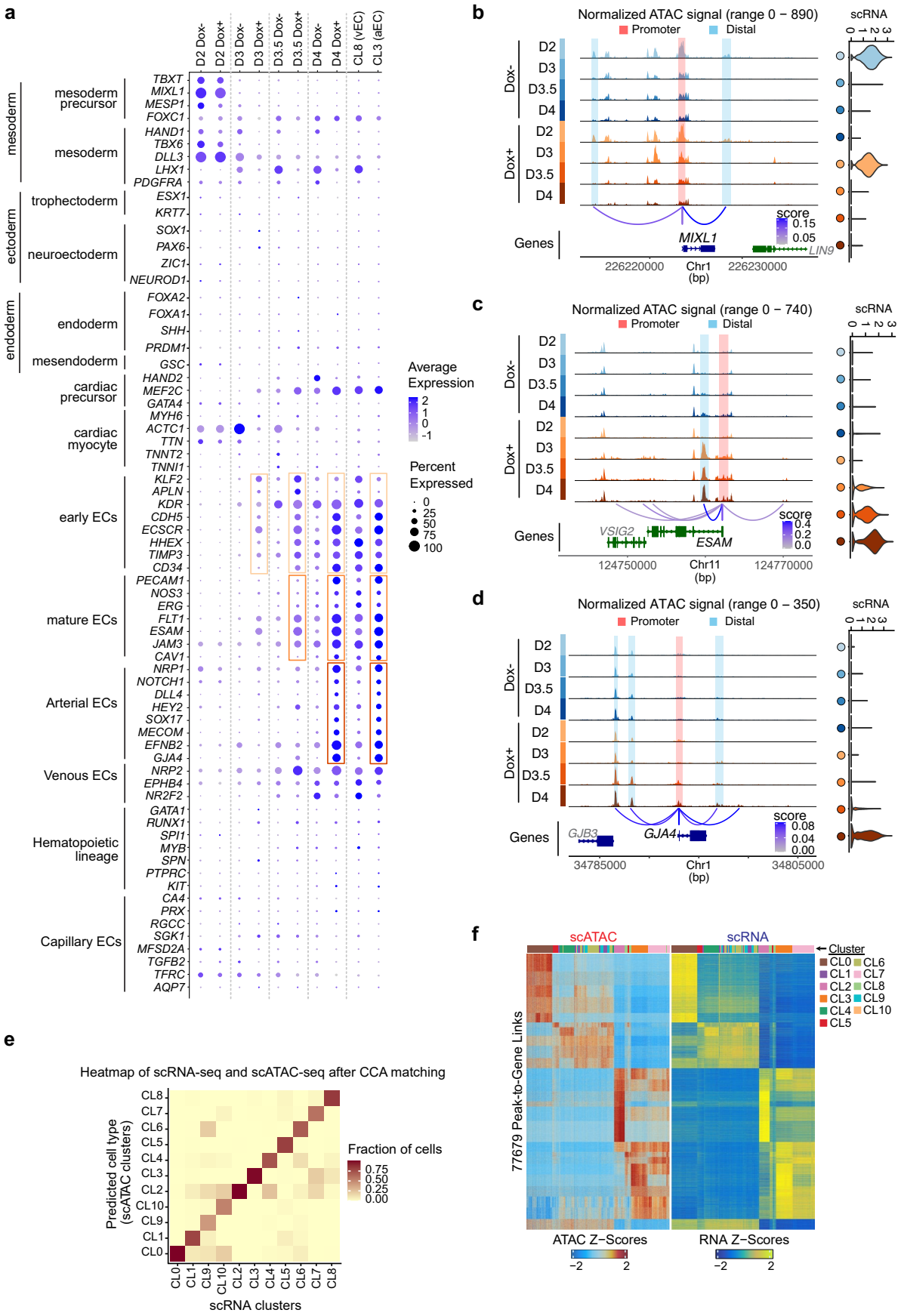
Extended Data Fig. 2 | Functional characterization of iPS cell-derived ECs with ETV2 overexpression. a, iPS cell-derived ECs immunostained for CDH5 (red), PECAM1 (green), and DAPI (blue). Scale bar, 50 µm. **b**, Alignment of iPS cell-derived ECs under static and flow conditions. 0° represents the direction of flow. Scale bar, 50 µm. **c**, Migration of iPS cell-derived ECs in the wound scratch assay. Scale bar, 300 µm. **d**, Nitric oxide (NO) production of iPS cell-derived ECs measured by flow cytometry upon exposure to DAF-FM. MFI, mean fluorescence intensity. n = 3 independent biological replicates. Unpaired two-tailed Student's t-test. **e**, AcLDL uptake by iPS cell-derived ECs. Scale bar, 25 µm. Data in **a**-**c**,**e** are representative of n = 3 biologically independent experiments. **f**, Expression of cell surface adhesion molecules *ICAM1*, *VCAM1*, and *SELE* in iPS cell-derived ECs when stimulated with 10, 100, or 500 ng/µL TNF. n = 3 independent biological replicates. One-way ANOVA with Dunnett's multiple comparison test. Data in **d,f** are shown as mean ± s.e.m. **g**, PCA plot based on the top 1,000 most variable genes across iPS cell-derived ECs and indicated human cell types (from GSE138734). **h**, Heatmap of the top DEGs in each cluster and their top enriched GO terms (Fisher's exact test, corrected for multiple testing using the Benjamini–Hochberg method). reg, regulation; dev, development; pos, positive. **i,j**. Wild-type parental iPS cells and TRE3G-ETV2 iPS cells were differentiated with different concentrations of Dox and analyzed at Day 4 by scRNA-seq. **i**, UMAP of each genotype and Dox concentration, colored by clusters. **j**, Integrated UMAPs colored by cell clusters (left) and by expression of selected marker genes.

https://doi.org/10.1038/s44161-025-00660-y



Extended Data Fig. 3 | Quality control of scRNA-seq and scATAC-seq data. a, Quality control metrics of scRNA-seq data for cell samples from Dox− and Dox+ groups. Number of genes (nFeature_RNA), number of molecular (nCount_RNA), and percentage of mitochondrial counts (percent.mt) is plotted for each sample. **b**, Quality control metrics of scATAC-seq data for cell samples from Dox− and Dox+ groups. TSS enrichment score for the cells in each sample. **c**, Fragment length periodicity for all the cells in each sample. **d**, The number of peaks for each time point. Bars are colored by peak location relative to genomic annotations.

https://doi.org/10.1038/s44161-025-00660-y

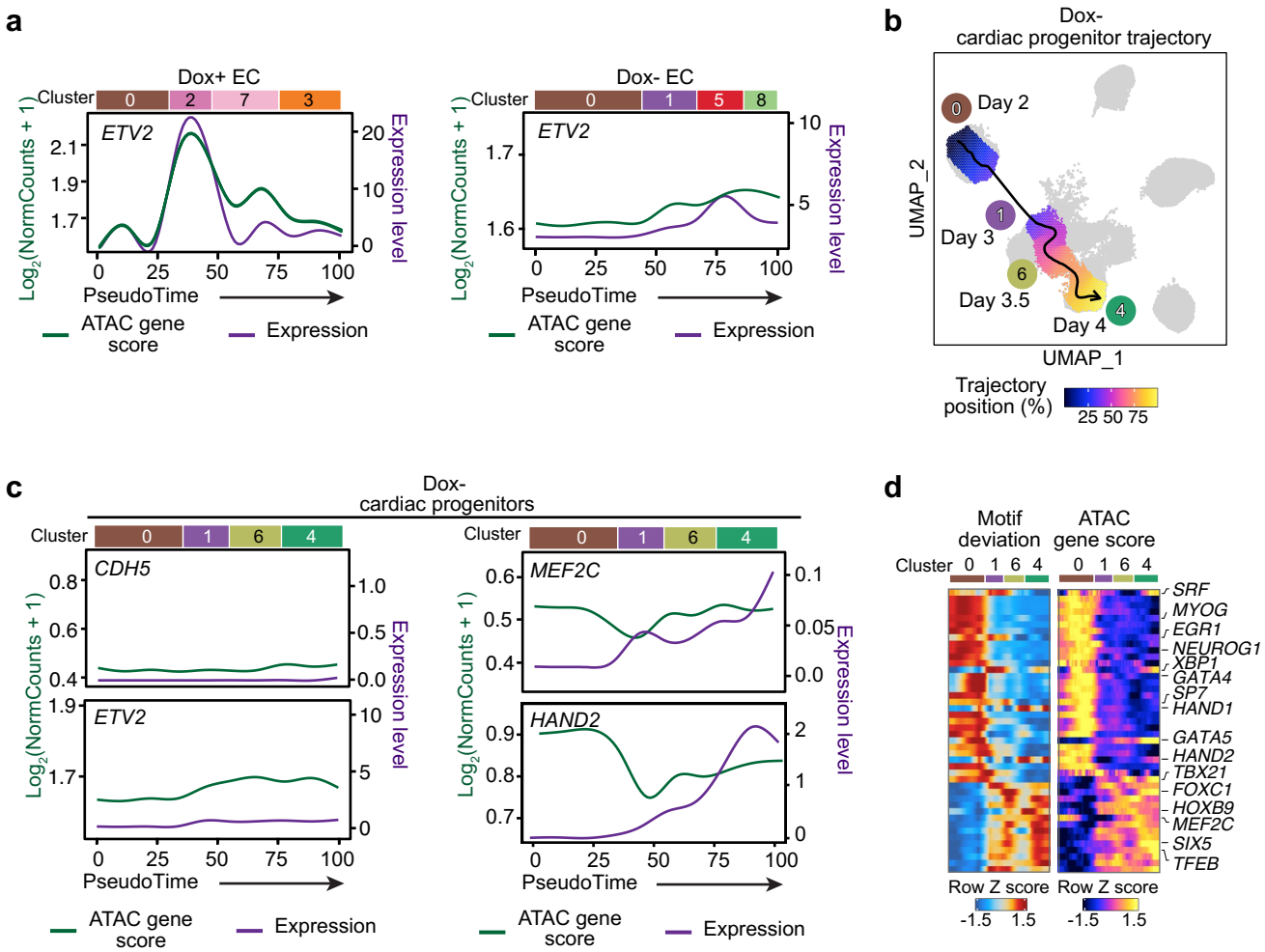


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s44161-025-00660-y

Extended Data Fig. 4 | Gene expression of representative cell type markers and integration of scRNA-seq and scATAC-seq data. a, Dot plots of quantitative bulk changes in gene expression between Dox− and Dox+ cells for markers of mesoderm, ectoderm, endoderm, cardiac precursors, early ECs, mature ECs, arterial EC, venous EC, genes involved in hematopoietic lineage development, and capillary ECs. **b**–**d**, Genome tracks of time point-resolved aggregate scATAC-seq data around the *MIXL1* (**b**), *ESAM* (**c**) and *GJA4* (**d**) gene loci. **e**, Heatmap showing the cluster-cluster mapping between scRNA-seq and scATAC-seq clusters after canonical correlation analysis (CCA) matching. **f**, Heatmap of 77,679 statistically significant peak-to-gene links identified across the dataset with ArchR. Each row represents expression of a gene (scRNA, right) correlated to accessibility of a ATAC-seq peak (scATAC, left).

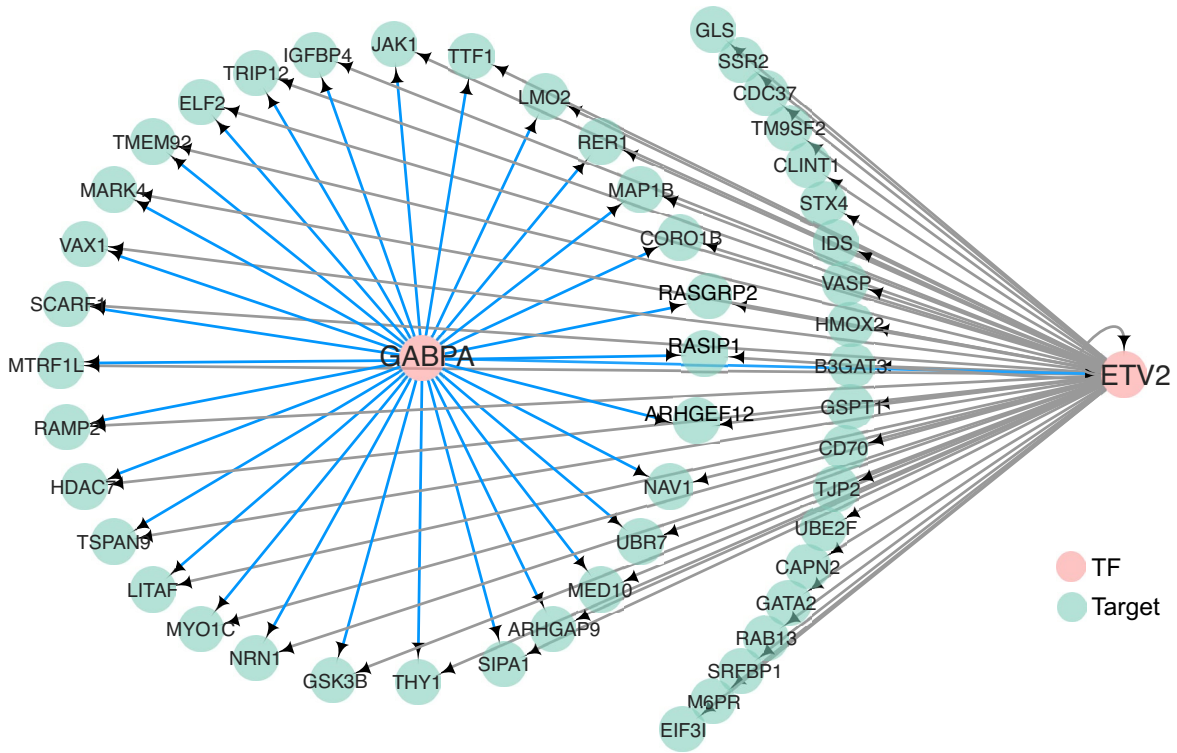
https://doi.org/10.1038/s44161-025-00660-y



Extended Data Fig. 5 | Differentiation trajectory terminating on cardiac progenitor cells. a, Gene expression and ATAC gene score dynamics of *ETV2* in Dox+ and Dox− EC trajectories across pseudotime. **b**, UMAP visualization of the cardiac progenitor trajectory of Dox− cells. Each cell is colored by its pseudotime and each cluster is annotated by sample collection time. The smoothed arrow represents a visualization of the interpreted trajectory in the UMAP embedding. **c**, Gene expression and ATAC gene score dynamics of the *CDH5*, *ETV2*, and cardiac progentior genes *MEF2C* and *HAND2*. **d**, Heatmap of TF regulators for which ATAC gene score is positively correlated with chromVAR TF deviation (motif deviation) across the Dox− cardiac progenitor trajectory, ordered by pseudotime.

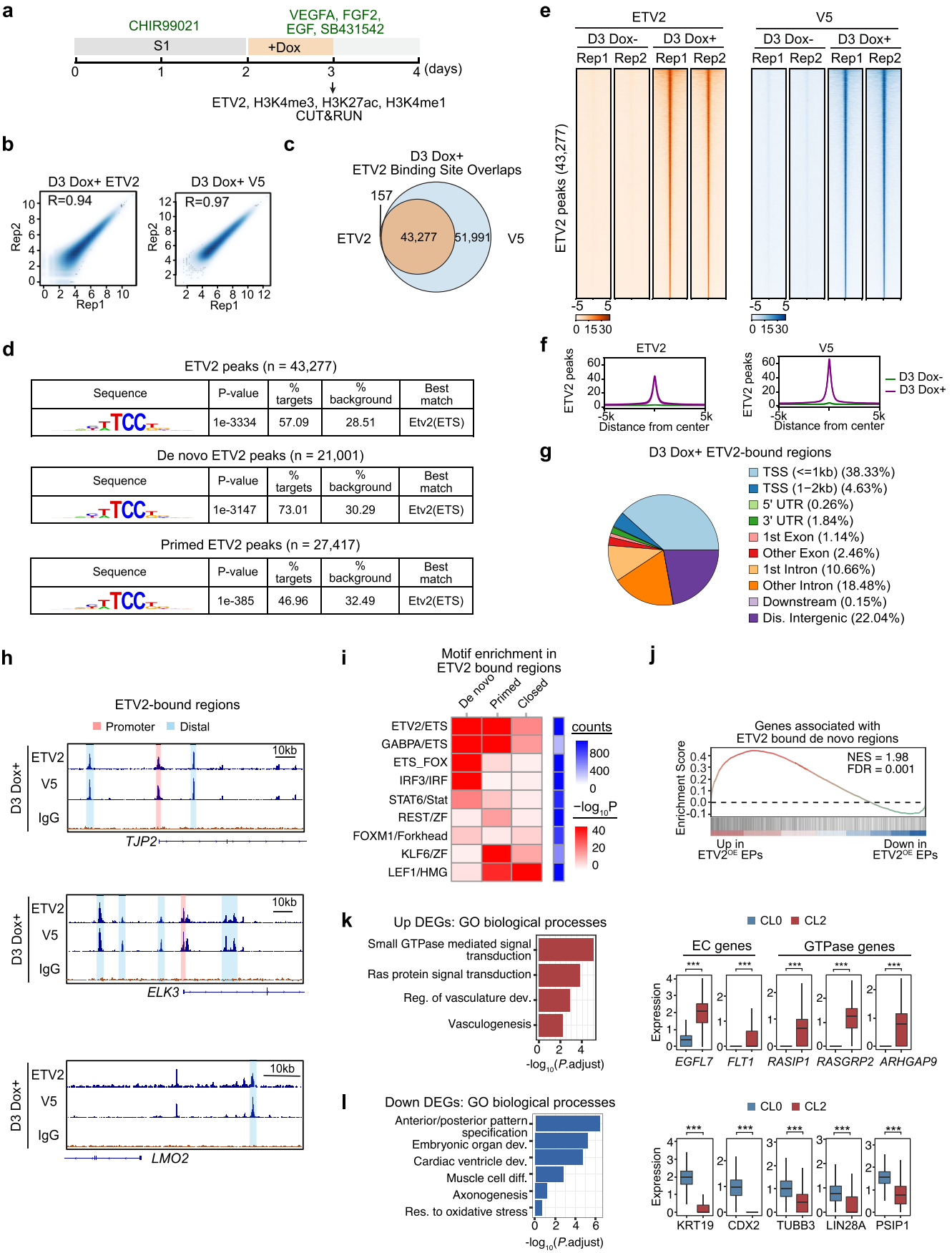
https://doi.org/10.1038/s44161-025-00660-y

a



Extended Data Fig. 6 | ETV2 drives endothelial specification in collaboration with GABPA in endothelial progenitors. a, Visualization of TF-target regulatory networks formed by *ETV2* and *GABPA*. The pink nodes represent the TFs. The green nodes represent the targets. The arrows represent the connections between each of TFs and their predicted target genes (GABPA, blue line; ETV2, gray line).

https://doi.org/10.1038/s44161-025-00660-y

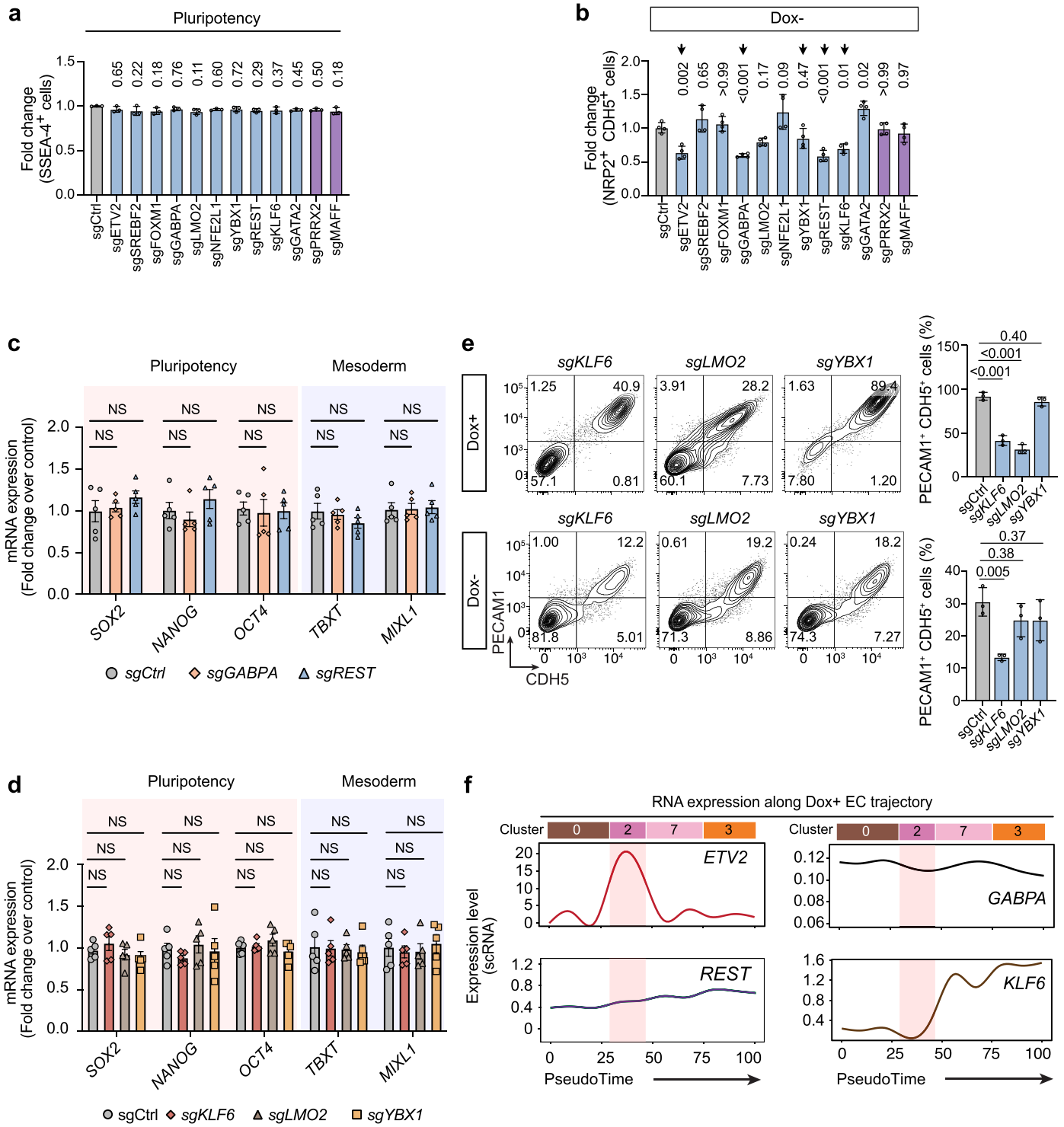


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s44161-025-00660-y

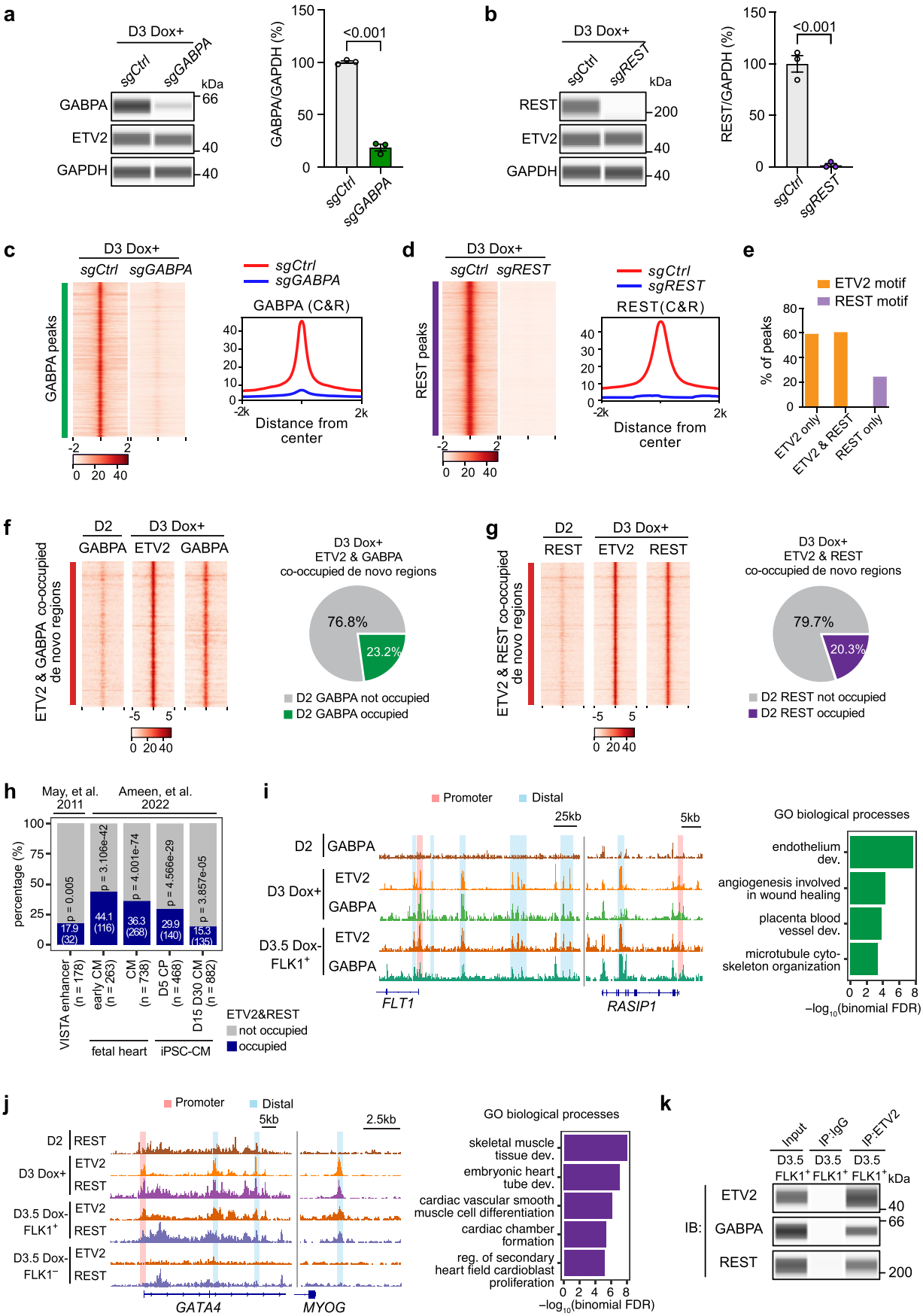
Extended Data Fig. 7 | ETV2 directly binds promoters and enhancers of target genes. a, Timing of the ETV2, H3K4me3, H3K27ac and H3K4me1 CUT&RUN analyses. ETV2 tagged with V5 was induced at day 2 by addition of Dox, and CUT&RUN was performed at day 3. **b**, Scatter plots comparing the D3 Dox+ ETV2 and V5 CUT&RUN signals during EC differentiation between two biological replicates. Pearson correlation coefficients are shown. **c**, Venn diagram showing overlap between D3 Dox+ ETV2 and V5 CUT&RUN peaks. Overlapping peaks were defined as high confidence ETV2 regions. **d**, The top enriched motifs in all ETV2 regions, de novo or primed ETV2 regions in D3 Dox+ cells. Motif enrichment *P* values were calculated using a one-sided binomial test in HOMER. **e**, ETV2 (orange) and V5 (blue) CUT&RUN signal at high confidence ETV2 regions. Each row represents a peak region (peak center ± 5 kb). **f**, Aggregation plots of ETV2 and V5 CUT&RUN signal at high confidence ETV2 regions in D3 Dox− and D3 Dox+ cells. **g**, Genomic distribution of high confidence ETV2 regions in D3 Dox+ cells. The number in brackets is the percentage of ETV2 peaks. **h**, Genome tracks of D3 Dox+ ETV2, V5 and IgG CUT&RUN data around the *TJP2*, *ELK3*, and *LMO2* gene loci. **i**, Motifs enriched in ETV2-bound de novo, primed, and closed regions (presented in Fig. 4b, one-sided binomial test). **j**, GSEA of genes linked to ETV2-bound de novo accessible regions. These genes were enriched among differentially expressed genes in D3 Dox+ $ETV2^{OE}$ EPs (CL2 versus CL0). Normalized enrichment score (NES) and FDR are calculated using GSEA. **k,l**, GO term analysis of upregulated genes (**k**) or downregulated genes (**l**) in D3 Dox+ $ETV2^{OE}$ EPs (CL2 versus CL0, Fisher's exact test, corrected for multiple testing using the Benjamini–Hochberg method). Box plots showing representative gene expression in CL0 (n = 17,707 cells) and CL2 (n = 10,905 cells). Two-sided LR test, $^{***}P < 2.22e\text{-}16$. Box plots visualize the median, first and third quartiles, and whiskers extend to 1.5× the interquartile range (IQR).

https://doi.org/10.1038/s44161-025-00660-y



Extended Data Fig. 8 | Validation of top candidates in EC differentiation. **a**, Flow cytometry for SSEA-4$^+$ in TRE3G-ETV2 iPS cells that were transfected with sgCtrl, or transfected with sgRNAs targeting to the indicated gene. n = 3 independent biological replicates. One-way ANOVA with Dunnett's multiple comparison test to sgCtrl. **b**, Validation of screening hits. TRE3G-ETV2 iPS cells were transfected with sgCtrl or sgRNAs targeting to the indicated gene. Differentiation efficiency to NRP2$^+$ CDH5$^+$ ECs under Dox− conditions was measured by flow cytometry. n = 4 independent biological replicates. One-way ANOVA with Dunnett's multiple comparison test to sgCtrl. **c**,**d**, RT–qPCR analysis of pluripotency markers *SOX2, NANOG, OCT4* mRNA expression in iPS cells across groups, and mesoderm markers *TBXT* and *MIXL1* mRNA expression in MPCs across groups. n = 5 independent biological replicates. One-way ANOVA with Dunnett's multiple comparison test to sgCtrl. **e**, Flow cytometry analysis and quantification for PECAM1$^+$ CDH5$^+$ in sg*KLF6*, sg*LMO2* and sg*YBX1* group under Dox+ and Dox− condition on day 4. n = 3 independent biological replicates. One-way ANOVA with Dunnett's multiple comparison test to sgCtrl. **f**, RNA expression dynamics of *ETV2*, *GABPA*, *REST* and *KLF6* across the Dox+ EC trajectory (presented in Fig. 2e). The red shaded area indicates the ETV2 expression window. Graphs in **a**,**b**,**e** show mean ± s.d. Graphs in **c** and **d** show mean ± s.e.m.

https://doi.org/10.1038/s44161-025-00660-y

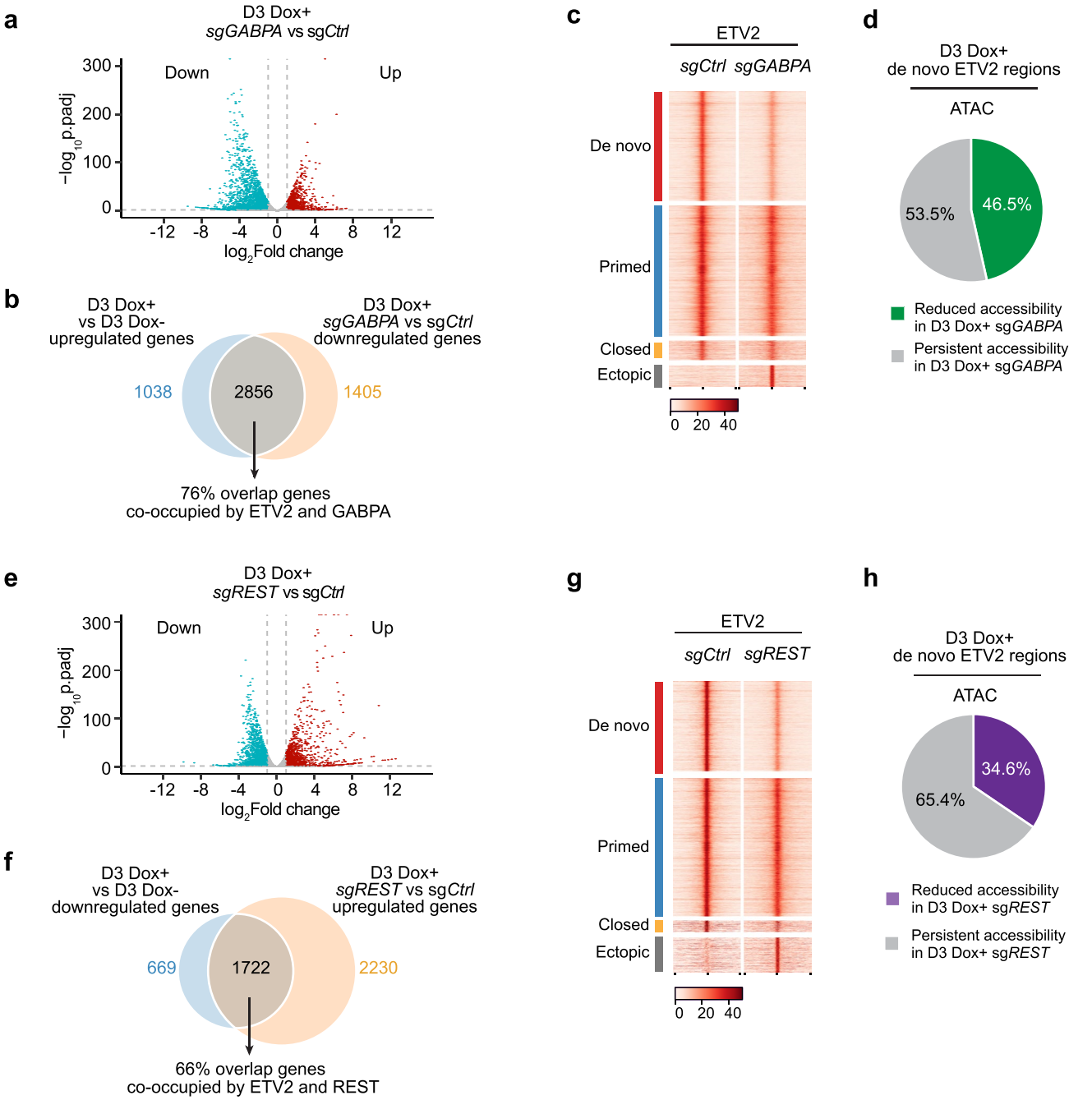


Extended Data Fig. 9 | See next page for caption.

https://doi.org/10.1038/s44161-025-00660-y

Extended Data Fig. 9 | Knockout of GABPA and REST in TRE3G-ETV2 iPS cells. a, GABPA and ETV2 proteins level in D3 Dox+ sg*Ctrl* and sg*GABPA* cells. Representative capillary westerns and quantitative analyses of GABPA normalized to GAPDH. n = 3 independent biological replicates. Unpaired two-tailed Student's t-test. **b**, REST and ETV2 proteins level in D3 Dox+ sg*Ctrl* and sg*REST* cells. Representative capillary westerns and quantitative analyses of REST normalized to GAPDH. n = 3 independent biological replicates. Unpaired two-tailed Student's t-test. **c**, GABPA binding signals in D3 Dox+ sg*Ctrl* and sg*GABPA* cells. Aggregation plots indicate the mean CUT&RUN signal at GABPA sites in each group. **d**, REST binding signals in D3 Dox+ sg*Ctrl* and sg*REST* cells. Aggregation plots indicate the mean CUT&RUN signal at REST sites in each group. **e**, Percentage of ETV2-only peaks, ETV2 & REST co-occupied peaks, REST-only peaks with the indicated motif. **f**, Heatmaps and pie charts showing D2 GABPA binding signals at D3 Dox+ ETV2 & GABPA co-occupied de novo peaks. **g**, Heatmaps and pie charts showing D2 REST binding signals at D3 Dox+ ETV2 & REST co-occupied de novo peaks. **h**, ETV2 & REST binding at a subset of VISTA-validated cardiac enhancers (column 1), accessible chromatin regions in fetal heart (column 2,3) or iPS cell-derived CM (column 4,5). VISTA enhancer coordinates were downloaded from the VISTA Enhancer Browser and filtered for human sequences and positive hits with activity in heart regions. One-sided hypergeometric test. CP, cardiac progenitor. CM, cardiomyocyte. **i**, Left panel, genome browser views showing GABPA binding signals in D2 cells, ETV2 and GABPA binding signals in D3 Dox+ cells, and D3.5 Dox− FLK1$^+$ cells at representative genes. Right panel, ETV2 and GABPA co-occupied regions enriched GO biological processes terms (GREAT; $-\log_{10}$(binomial FDR), one-sided binomial test) in D3.5 Dox− FLK1$^+$ cells. **j**, Left panel, genome browser views showing REST binding signals in D2 cells, ETV2 and REST binding signals in D3 Dox+ cells, D3.5 Dox− FLK1$^+$ cells, and D3.5 Dox− FLK1$^-$ cells at representative genes. Right panel, ETV2 and REST co-occupied regions enriched GO biological processes terms (GREAT; $-\log_{10}$(binomial FDR), one-sided binomial test) in D3.5 Dox− FLK1$^+$ cells. **k**, Co-immunoprecipitation (Co-IP) using ETV2 or IgG control antibodies, followed by capillary immunoblotting (IB) for ETV2, GABPA or REST in D3.5 Dox− FLK1$^+$ cells. Data is representative of n = 3 biologically independent experiments. Graphs in **a**,**b** show mean ± s.e.m.

https://doi.org/10.1038/s44161-025-00660-y



Extended Data Fig. 10 | GABPA and REST are required for chromatin opening and transcription activity. a, Volcano plot showing DEGs identified in D3 Dox+ sg*GABPA* cells versus D3 Dox+ sg*Ctrl* cells. P values were calculated using two-sided Wald test with Benjamini–Hochberg correction. Significantly upregulated (red) or downregulated genes (teal) are colored, respectively. **b**, Overlap between genes upregulated in D3 Dox+ cells (versus D3 Dox−) and downregulated in D3 Dox+ sg*GABPA* cells (versus sg*Ctrl*). **c**, ETV2 binding signals in D3 Dox+ sg*Ctrl* and D3 Dox+ sg*GABPA* cells. Regions are group by chromatin occupancy class (de novo, primed, closed and ectopic binding sites). **d**, Effect of *GABPA* knockout on chromatin accessibility. A subset of ETV2-bound de novo regions lost accessibility in D3 Dox+ sg*GABPA* cells. **e**, Volcano plot showing DEGs identified in D3 Dox+ sg*REST* cells versus D3 Dox+ sgCtrl cells. P values were calculated using two-sided Wald test with Benjamini–Hochberg correction. Significantly upregulated (red) or downregulated genes (teal) are colored, respectively. **f**, Overlap between genes downregulated in D3 Dox+ cells (versus D3 Dox−) and upregulated in D3 Dox+ sg*REST* cells (versus sg*Ctrl*). **g**, ETV2 binding signals in D3 Dox+ sg*Ctrl* and D3 Dox+ sg*REST* cells. Regions are group by chromatin occupancy class (de novo, primed, closed and ectopic binding sites). **h**, Effect of *REST* knockout on chromatin accessibility. A subset of ETV2-bound de novo regions lost accessibility in D3 Dox+ sg*REST* cells.

nature portfolio

Corresponding author(s): William Pu

Last updated by author(s): Mar 25, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Bio-Rad CFX96 Touch Deep Well Real-Time PCR, BD LSRFortessa, BD FACSMelody Cell Sorter, Olympus FV3000R confocal microscope, Keyence automated epifluorescent microscope, automated WES capillary based electrophoresis system (Wes; ProteinSimple), NovaSeq6000.
Data analysis	All packages used are available online and the versions and programs are described in the methods section. For more details, we have used FlowJo software (v10.7.1), ImageJ (v2.1.0), Compass for SW (v6.1.0), MAGeCK (v0.5.9), FastQC (v0.11.9), TrimGalore (v0.6.7), Hisat2 (v2.1.0), FeatureCounts (v2.0.1), Bowtie2 (v2.3.4.3), samtools (v1.13), Picard tools MarkDuplicates (v 2.27.4), MACS2 (v2.1.1), Deeptools (v3.5.2), BEDTools (v 2.30.0), IGV software (v2.12.3), 10× Genomics CellRanger pipeline (v6.1.2), 10× Genomics CellRanger ATAC pipeline (v2.1.0), R environment (v4.1), Seurat R package (v4.1.0), Signac (1.9.0)79 and ArchR (1.0.2). Custom code was not generated outside of the use of the predesigned packages.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw sequencing and related processed data have been deposited in GEO under the accession number (GSE267565).
All data aligned to human reference genome (hg38).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for each experiment is indicated in the figures or corresponding figure legends. Sequencing experiments were performed in duplicate which is the standard for the filed. For all other studies, experiments were performed with 3-4 biological replicates which is standard for the field and our previous works. Hota SK, et al. Nature. 2022; Sweat ME, et al. Nat Cardiovasc Res. 2023; Prondzynski M, et al. Nat Commun. 2024;
Data exclusions	No data excluded.
Replication	Reproducibility of findings were assessed via statistical test, varying on sample size and experiment type. Differentiation of iPSCs to endothelial lineage were reliable and reproducible. Other standard assays were performed at least 3 independent times and all replications were successful. Details on the number of replications performed are described in figure legends.
Randomization	Randomization is not relevant to this study, where groups are defined they are determined by the genotype.
Blinding	Investigators were blinded to group allocation during data collection and analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

The following primary antibodies were used:
"Antibodies" "Clone" "Company" "Dilution" "Cat#"
CD31 (PECAM1) Antibody (WM59)-APC, WM59, Biolegend, 5 ul/test, Cat#303116
CD31 (WM59) Antibody-Brilliant Violet 421, WM59, Biolegend, 5 ul/test, Cat#303123
CD144 (VE-cadherin; CDH5) Antibody (16B1)-PE, 16B1, Invitrogen, 5 ul/test, Cat#12-1449-82
CD144 (VE-cadherin; CDH5) Antibody (16B1)-PE-Cyanine7, 16B1, Invitrogen, 5 ul/test, Cat#25-1449-42
Neuropilin-2 conjugated Antibody(# 257103)-APC, 257103, R&D, 5 ul/test, Cat#FAB22151A
Delta-like protein 4 (DLL4) Antibody(MHD4-46)-APC, MHD4-46, Biolegend, 5 ul/test, Cat#346507
ETV2 antibody [EPR5229(3)]- APC, EPR5229(3), abcam, 5 ul/test, Cat#ab319250
SSEA-4 Antibody-FITC, REA101, Miltenyi Biotec, 2 ul/test, Cat#130-122-918
a-Actinin (Sarcomeric) Antibody-FITC, REA402, Miltenyi Biotec, 2 ul/test, Cat#130-119-766
Cardiac Troponin T Antibody-FITC, REA400, Miltenyi Biotec ,2 ul/test, Cat#130-119-575
CUTANA Rabbit IgG Antibody, Polyclonal, Epicypher, 1 ug/test, Cat#SKU: 13-0042
Anti-ETV2 antibody, EPR5229(3), Abcam, 1 ug/test, 1:400 Wes, Cat#ab181847
V5 tag antibody, Polyclonal, Abcam, 1 ug/test, Cat#ab15828
H3K4me3 Antibody, Polyclonal, Diagenode, 1 ug/test, Cat#C15410003-10
H3K4me1 Antibody, Polyclonal, Diagenode, 1 ug/test, Cat#C15410194-10
H3K27ac antibody, Polyclonal, Active Motif, 1 ug/test, Cat#39133
GABPA antibody, Polyclonal, Proteintech, 1 ug/test, 1:200 Wes, Cat#21542-1-AP
REST antibody, Polyclonal, millipore, 1 ug/test, 1:100 Wes, Cat#17-641
GAPDH, Polyclonal, Thermo, 1:50 Wes, Cat#PA1-16777
CD31 (PECAM1), JC70A, Agilent/DAKO, 1:200 , Cat#M082329-2
VE-Cadherin (CDH5) antibody, Polyclonal, Abcam, 1:200, Cat# ab33168
Anti-a-Actinin (Sarcomeric)(ACTN2) antibody, EA-53, Sigma, 1:200, Cat# A7811
Donkey anti-Mouse IgG Secondary Antibody(Alexa Fluor 488), Polyclonal, Invitrogen, 1:300, Cat# A21202
Donkey anti-Rabbit IgG Secondary Antibody(Alexa Fluor 555), Polyclonal, Invitrogen, 1:300, Cat# A31572

The information on all primary antibodies is also provided in the Supplementary Table.

Validation

All antibodies used were purchased from commercial vendors and were selected because they had been well validated by their manufacturers and in different publications for use in the species we used them for. Validation details and relevant publications are provided on the manufacturer's website.
CD31 (PECAM1) Antibody (WM59), APC https://www.biolegend.com/en-us/products/apc-anti-human-cd31-antibody-6123?GroupID=BLG5721
CD31 (WM59) Antibody, Brilliant Violet 421 https://www.biolegend.com/en-us/products/brilliant-violet-421-anti-human-cd31-antibody-8588
CD144 (VE-cadherin; CDH5) Antibody (16B1), PE https://www.thermofisher.com/antibody/product/CD144-VE-cadherin-Antibody-clone-16B1-Monoclonal/12-1449-82
CD144 (VE-cadherin; CDH5) Antibody (16B1), PE-Cyanine7 https://www.thermofisher.com/antibody/product/CD144-VE-cadherin-Antibody-clone-16B1-Monoclonal/25-1449-42
Neuropilin-2 conjugated Antibody(# 257103), APC https://www.rndsystems.com/products/human-mouse-neuropilin-2-apc-conjugated-antibody-257103_fab22151a
Delta-like protein 4 (DLL4) Antibody(MHD4-46), APC https://www.biolegend.com/en-us/products/apc-anti-human-delta-like-protein-4-dll4-antibody-7835
ETV2 antibody [EPR5229(3)], APC https://www.abcam.com/en-us/products/primary-antibodies/apc-etv2-antibody-epr52293-ab319250
SSEA-4 Antibody, FITC https://www.miltenyibiotec.com/US-en/products/ssea-4-antibody-anti-human-reafinity-rea101.html#conjugate=fitc:size=100-tests-in-200-ul
a-Actinin (Sarcomeric) Antibody, FITC https://www.miltenyibiotec.com/US-en/products/a-actinin-sarcomeric-antibody-anti-human-mouse-rat-reafinity-rea402.html#Conjugate=FITC:size=100-tests-in-200-uL
Cardiac Troponin T Antibody, FITC https://www.miltenyibiotec.com/US-en/products/cardiac-troponin-t-antibody-anti-human-mouse-rat-reafinity-rea400.html#Conjugate=FITC:size=100-tests-in-200-uL
CUTANA Rabbit IgG Antibody https://www.epicypher.com/products/nucleosomes/snap-cutana-spike-in-controls/cutana-igg-negative-control-antibody-for-cut-run-and-cut-tag
Anti-ETV2 antibody https://www.abcam.com/en-us/products/primary-antibodies/etv2-antibody-epr52293-ab181847
V5 tag antibody https://www.abcam.com/en-us/products/primary-antibodies/v5-tag-antibody-ab15828
H3K4me3 Antibody https://www.diagenode.com/en/p/h3k4me3-polyclonal-antibody-premium-sample-size-10-ug
H3K4me1 Antibody https://www.diagenode.com/en/p/h3k4me1-polyclonal-antibody-premium-sample-size-10-ug

H3K27ac antibody https://www.activemotif.com/catalog/details/39133/histone-h3-acetyl-lys27-antibody-pab
GABPA antibody https://www.ptglab.com/products/NRF2-Antibody-21542-1-AP.htm?srsltid=AfmBOopAJc9wkRUOvsy91TWO9sktd2dkeD_aS3SMbfc2iMUPVftRATqG
REST antibody https://www.sigmaaldrich.com/US/en/product/mm/17641
GAPDH https://www.thermofisher.com/antibody/product/GAPDH-Antibody-Polyclonal/PA1-16777
CD31 (PECAM1) https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/cd31-endothelial-cell-%28concentrate%29-76539
VE-Cadherin (CDH5) antibody https://www.abcam.com/en-us/products/primary-antibodies/ve-cadherin-antibody-intercellular-junction-marker-ab33168
Anti-a-Actinin (Sarcomeric)(ACTN2) antibody https://www.sigmaaldrich.com/US/en/product/sigma/a7811

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	The human male BJ273 iPSC line was sourced from Boston Children's Hospital's Stem Cell Core. TRE3G-ETV2 iPSCs were engineered from this parental line transfecting them with piggyBac expression plasmid (SBI#PB210PA-1) and the piggyBac transposon vector depicted in Fig. 1A. HEK293T cells were purchased directly from ATCC (catologue # CRL-3216).
Authentication	Flow cytometry and molecular analysis such as qPCR were performed to characterize the iPSC line. HEK293T cells were not authenticated as they were purchased directly from ATCC.
Mycoplasma contamination	All cell lines tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Data was deposited under accession number GSE267565.
Files in database submission	Both unprocessed FASTQ files and downstream analyses (.bw files) were deposited in the GEO submission.
Genome browser session (e.g. UCSC)	The genome tracks of peaks were visualized by the IGV software.

Methodology

Replicates	We have validated the reproducibility of CUT&RUN by analyzing at least 2 replicates.
Sequencing depth	Reads were paired-end. These data are summarized in Supplementary Table.
Antibodies	ETV2 antibody (Abcam, cat# ab181847) V5 tag antibody (Abcam, cat# ab15828) H3K4me3 Antibody (Diagenode, cat# C15410003-10) H3K4me1 Antibody (Diagenode, cat# C15410194-10) H3K27ac antibody (Active Motif, cat# 39133) GABPA antibody (Proteintech, Cat#21542-1-AP) REST antibody (Millipore, Cat#17-641) CUTANA™ Rabbit IgG CUT&RUN Negative Control Antibody (Epicypher, cat# SKU: 13-0042)

Peak calling parameters	Peaks were identified using MACS2 (v2.1.1) and using the IgG control sequencing data as background, with the 'narrow peak' option for TFs and H3K4me3 and the 'broad peak' option for H3K27Ac and H3K4me1.
Data quality	Quality control was performed using FastQC (v0.11.9).
Software	Bowtie2 (v2.3.4.3), samtools (v1.13), Picard tools MarkDuplicates (v 2.27.4), MACS2 (v2.1.1)

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were dissociated using TrypLE Select, washed and resuspended in FACS buffer (PBS with 0.5% BSA). For surface antigens, live cells were stained with commercially available antibodies.
Instrument	BD LSRFortessa
Software	FlowJo v10.7.1
Cell population abundance	A minimum of 10,000 cells per sample were anaylzed.
Gating strategy	All samples were initially gated using FSC-A and SSC-A gating to identify the live cell population. Single cells were identified and doublets excluded by gating forward scatter height (FSC-H) versus area (FSC-A). The positively fluorescent cells was gated based on the fluorescent intensity of positive control cells.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.