Circulation Research

ORIGINAL RESEARCH

GATA4 Regulates Developing Endocardium Through Interaction With ETS1

Pingzhu Zhou,* Yan Zhang,* Isha Sethi, Lincai Ye, Michael A. Trembley, Yangpo Cao, Brynn N. Akerberg, Feng Xiao, Xiaoran Zhang, Kai Li, Blake D. Jardin, Neil Mazumdar, Qing Ma, Aibin He, Bin Zhou, William T. Pu

BACKGROUND: The pioneer transcription factor GATA4 is expressed in multiple cardiovascular lineages and is essential for heart development. GATA4 lineage-specific occupancy in the developing heart underlies its lineage specific activities. Here, we characterized GATA4 chromatin occupancy in cardiomyocyte and endocardial lineages, dissected mechanisms that control lineage specific occupancy, and analyzed GATA4 regulation of endocardial gene expression.

METHODS: We mapped GATA4 chromatin occupancy in cardiomyocyte and endocardial cells of embryonic day 12.5 (E12.5) mouse heart using lineage specific, Cre-activated biotinylation of GATA4. Regulation of GATA4 pioneering activity was studied in cell lines stably overexpressing GATA4. GATA4 regulation of endocardial gene expression was analyzed using single cell RNA sequencing and luciferase reporter assays.

RESULTS: Cardiomyocyte-selective and endothelial-selective GATA4 occupied genomic regions had features of lineage specific enhancers. Footprints within cardiomyocyte- and endothelial-selective GATA4 regions were enriched for NKX2-5 and ETS1 motifs, respectively, and both of these TFs interacted with GATA4 in co-immunoprecipitation assays. In stable NIH3T3 cell lines expressing GATA4 with or without NKX2-5 or ETS1, the partner TFs re-directed GATA4 pioneer binding and augmented its ability to open previously inaccessible regions, with ETS1 displaying greater potency as a pioneer partner than NKX2-5. Single-cell RNA sequencing of embryonic hearts with endothelial cell–specific *Gata4* inactivation identified *Gata4*-regulated endocardial genes, which were adjacent to GATA4-bound, endothelial regions enriched for both GATA4 and ETS1 motifs. In reporter assays, GATA4 and ETS1 cooperatively stimulated endothelial cell enhancer activity.

CONCLUSIONS: Lineage selective non-pioneer TFs NKX2-5 and ETS1 guide the activity of pioneer TF (transcription factor) GATA4 to bind and open chromatin and create active enhancers and mechanistically link ETS1 interaction to GATA4 regulation of endocardial development.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: animals ■ chromatin ■ gene expression ■ genomics ■ immunoprecipitation

Embryonic development and organ morphogenesis require elegantly choreographed, cell type specific regulation of gene transcription. General principles of transcriptional regulation during cell lineage specification and differentiation are well established. TFs (transcription factors) bind to cis-acting regulatory regions to modulate the expression of nearby genes. The 5-10 base pair binding motif of each TF is not sufficient to achieve the observed binding specificity. One important mechanism for increasing specificity is combinatorial binding and transcriptional activation by multiple TFs. A second is restricting TF-binding site accessibility through cell type specific DNA packaging into nucleosomes and higher order chromatin structures. A limited set of pioneering TFs have the ability to bind to nucleosomal DNA and locally reorganize chromatin to make it

Correspondence to: William T. Pu, Email william.pu@cardio.chboston.org
*P. Zhou and Y. Zhang are co-first authors.
Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCRESAHA.120.318102.
For Sources of Funding and Disclosures, see page XXX.

Circulation Research is available at www.ahajournals.org/journal/res

ORIGINAL RESEARCH

Novelty and Significance

What Is Known?

- Mutations in GATA4 and ETS1 are implicated in causing congenital heart disease.
- GATA4 is required in both cardiomyocyte and endocardial lineages for fetal heart development.

What New Information Does This Article Contribute?

- Regions selectively occupied by GATA4 in cardiomyocyte or endocardial lineages had features of cell-type specific enhancers.
- GATA4 interacts with ETS1, a transcription factor linked to congenital heart disease, to regulate endocardial development.
- GATA4 pioneer factor activity can be partitioned into the binding, chromatin opening, and enhancer activation steps.
- Non-pioneer partner transcription factors ETS1 and NKX2-5 differentially impact discrete aspects of GATA4 pioneer activity.

The transcription factor GATA4 is a critical regulator of heart development, and its mutation causes congenital heart disease. GATA4 belongs to a subset of transcription factors (pioneer factors) that bind to DNA wrapped around histones, making these regions accessible for binding by non-pioneer transcription factors. GATA4 is expressed in cardiomyocyte and endocardial lineages, and mutation of GATA4 in each lineage in mice disrupts heart development. Here, we used lineage-specific in vivo GATA4 biotinylation to map GATA4 chromatin occupancy in lineage. We show that lineage-selective GATA4 regions have features of cell-type-specific transcriptional enhancers. ETS1, a transcription factor enriched in endocardial cells whose mutation is associated with congenital heart disease, co-occupies endocardial GATA4 regions. ETS1 directly interacted with GATA4, and this interaction redirected GATA4 pioneer activity. Our study identifies GATA4-regulated endocardial genes and enhancers and links ETS1 interaction to GATA4 regulation of endocardial development.

Nonstandard Abbreviations and Acronyms

ATAC-seq	assay for transposase-accessible chromatin and sequencing
bioChIP-seq	biotin-mediated chromatin pull down and sequencing
DEG	differentially expressed gene
EC	endothelial cell
ECC	endocardial cell
scRNA-seq	single-cell RNA seq
TF	transcription factor

accessible for binding by other TFs.[1] Thus, pioneering TFs play an important role in cell lineage specification and differentiation.

While pioneer TFs bind nucleosomal DNA, they still occupy chromatin in a cell type specific manner,[2] which is central to their function in guiding cell lineage programs.[1] Determinants of cell type specific binding by pioneer TFs are beginning to be elucidated and include the underlying chromatin landscape,[2–4] cell type specific cofactors, and partner TFs.[2,5] However, the cell type–specific partner TFs that direct cell type specific occupancy have been studied for few pioneer TFs, and the mechanisms by which they modulate pioneer TF activity are incompletely characterized.

The heart is the earliest organ system to develop and function. Two of the major lineages that contribute to the developing heart are cardiomyocytes and endocardial cells (ECCs).[6] In each of these lineages, the pioneer TF GATA4[7] is essential for normal cardiac development[8,9] In cardiomyocytes, GATA4 regulates cardiac gene expression through collaborative interactions with other cardiac TFs, including NKX2-5, MEF2A, MEF2C, SRF, TBX5, and TEAD1.[10–12] Less is known about the factors that cooperate with GATA4 to regulate gene expression in non-cardiomyocyte lineages such as ECC. Whether and how these interacting TFs modulate GATA4 pioneer activity is also unknown.

An important step in dissecting TF function is to measure its chromatin occupancy, most commonly through chromatin immunoprecipitation followed by sequencing (ChIP-seq). ChIP-seq has traditionally been performed on whole tissues, yielding aggregate chromatin occupancy data for tissue's multiple cell types. These aggregate data limit insights into cell type specific regulatory mechanisms. Recently, we described a method for highly sensitive and reproducible ChIP-seq from a Cre-labeled lineage within a tissue.[13] Here, we deploy this method to determine GATA4 chromatin occupancy in cardiomyocytes and endothelial cells (ECs). GATA4 occupied distinct chromatin regions in these lineages. In ECs, ETS1 acted as a GATA4-interacting partner TF. Gata4 inactivation in ECs coupled with single cell RNA sequencing (scRNAseq) identified GATA4-regulated ECC genes, with adjacent GATA4- and ETS1-dependent regulatory elements. Dissection of GATA4 pioneering activity in

an heterologous expression system demonstrated that ETS1 modulates GATA4 pioneer activity at multiple distinct steps.

METHODS

Please refer to the Supplemental Material and Supplemental Methods for detailed Methods and the accompanying Major Resources Table. Data sources for this article are summarized in Table S1.

Data Availability

Data used in this article and the data sources are summarized in Table S1. Original data in this article have been deposited to gene expression omnibus: GSE156001 (GATA4 biotin-mediated chromatin pull down and sequencing [bioChIP-seq], assay for transposase-accessible chromatin and sequencing (ATAC-seq), and bulk RNA-seq in E12.5 mouse hearts), GSE155652 (ChIP-seq in NIH3T3 cells). and GSE208162 (single cell RNA seq of E12.5 mouse hearts).

RESULTS

Cre-Activated, Lineage-Specific GATA4 bioChIP-Seq in Developing Murine CMs and ECs

GATA4 immunostaining of E12.5 heart sections confirmed its expression in cardiomyocytes and ECs of the developing fetal heart (Figure S1A and S1B). To map GATA4 chromatin occupancy in these lineages, we used a Cre-activated GATA4 biotinylation strategy (Figure 1A) comprising 2 engineered alleles, *Gata4*^{fb} and *Rosa26*^{fsBirA}. *Gata4*^{fb} is a knockin allele that expresses GATA4 fused to a C-terminal peptide containing Flag and AviTag epitopes (35 amino acid residues).[11] *Gata4*$^{fb/fb}$ mice are normal and have normal cardiac structure and function.[11] After Cre-mediated recombination, *Rosa26*^{fsBirA} (fs= floxed transcriptional stop) expresses BirA,[13] an *E. coli* enzyme that specifically recognizes and biotinylates the AviTag.[14] Thus, cardiomyocyte or endothelial cell (EC)-specific expression of Cre will result in lineage specific GATA4 biotinylation. Chromatin crosslinking, fragmentation, and high-affinity pulldown of GATA4 with streptavidin beads followed by high-throughput sequencing (bioChIP-seq) then identifies GATA4-bound regions across the genome.

We focused on embryonic day 12.5 (E12.5), when the developing heart provided sufficient material to make bioChIP-seq technically feasible. Furthermore, conditional *Gata4* inactivation studies suggested that GATA4 activity within cardiomyocyte and EC lineages was required for heart development during this time period. Inactivation of a conditional *Gata4* allele by cardiomyocyte-restricted *Tnnt2-Cre*[15] yielded hearts with thin ventricular walls, impaired development of the interventricular septum, and reduced myocardial trabeculation (Figure S1C). In ECs, *Gata4*$^{fl/fl}$*; Cdh5-CreERT2* embryos treated with tamoxifen at E9.5 died at E14.5 with peripheral edema (Figure S2A), suggestive of heart failure. Immunostaining confirmed GATA4 inactivation in ECs (Figure S2B), and H&E stained transverse sections demonstrated thin ventricular walls and increased myocardial trabeculation (Figure S2C and S2D), associated with decreased cardiomyocyte proliferation (Figure S2E and S2F). This phenotype is distinct from the previously reported failure of valve development in *Gata4*$^{fl/fl}$*; Tie2-Cre* embryos,[9] as the timing of activation of Cdh5-CreERT2 spared the valve mesenchyme from GATA4 inactivation (Figure S2B). Thus, this model reveals a previously unrecognized requirement for GATA4 in ECs to support myocardial growth and differentiation.

We validated that *Rosa26*^{fsBirA} biotinylates GATA4^{fb} only in the presence of Cre (Figure 1B). We activated BirA in cardiomyocytes and ECs using *Tnnt2-Cre*[15] and *Tie2-Cre*,[16] respectively, and enables GATA4 pulldown on streptavidin beads (Figure 1C). We preferred *Tie2-Cre* over *Cdh5-CreERT2*[17] for this application because *Tie2-Cre* similarly labels ECs but obviates the need for tamoxifen. In E12.5 heart ventricles, there are more ECCs than vascular ECs (vECs), and *Tie2-Cre*-labeled mesenchymal derivatives within the endocardial cushions were largely excluded from the dissected ventricles (Figure 1D; Figure S1D). Furthermore, GATA4 is more robustly expressed in ECCs compared with vascular ECs (vECs; see single cell sequencing data in Figure 7C). Consequently, GATA4 bioChIP-seq signal in *Gata4*$^{flbio/+}$*; Rosa26*^{fsBirA}*; Tie2-Cre* samples should reflect chromatin occupancy within ECCs.

We next mapped GATA4 chromatin occupancy in cardiomyocytes and ECs by performing bioChIP-seq from *Gata4*$^{flbio/+}$*; Rosa26*^{fsBirA}*; Tnnt2-Cre* or *Gata4*$^{flbio/+}$*; Rosa26*^{fsBirA}*; Tie2-Cre* heart ventricles in biological triplicate. We compared these results to our previously reported[11] pan-lineage GATA4 bioChIP-seq from E12.5 *Gata4*$^{flbio/+}$*; Rosa26*^{BirA} ventricles, which widely express BirA (Figure 1D; Table S1; Data S1). We observed excellent within-group correlation between replicate samples and substantial intergroup differences (Figure 1E). Successful lineage-selective GATA4 bioChIP-seq was supported by visualizing the bioChIP-seq signal at individual genomic loci (Figure 1F and 1G; Figure S3). For example, EC-selective genes *Nfatc1*, *Notch1*, *Egfl7*, and *Ets1* neighbored GATA4 peaks enriched in the *Tie2-Cre* samples (Figure 1F; Figure S3), and cardiomyocyte-enriched GATA4 peaks neighboring cardiomyocyte-selective genes such as *Nkx2-5* and *Ankrd1* (Figure 1G; Figure S3). GATA4 regions adjacent to *Nfatc1*, *Egfl7*, and *Nkx2-5* overlapped regions that were previously demonstrated to have EC or cardiomyocyte enhancer activity[13,18,19] and identified new candidate regulatory regions (Figure 1F and 1G; Figure S3).

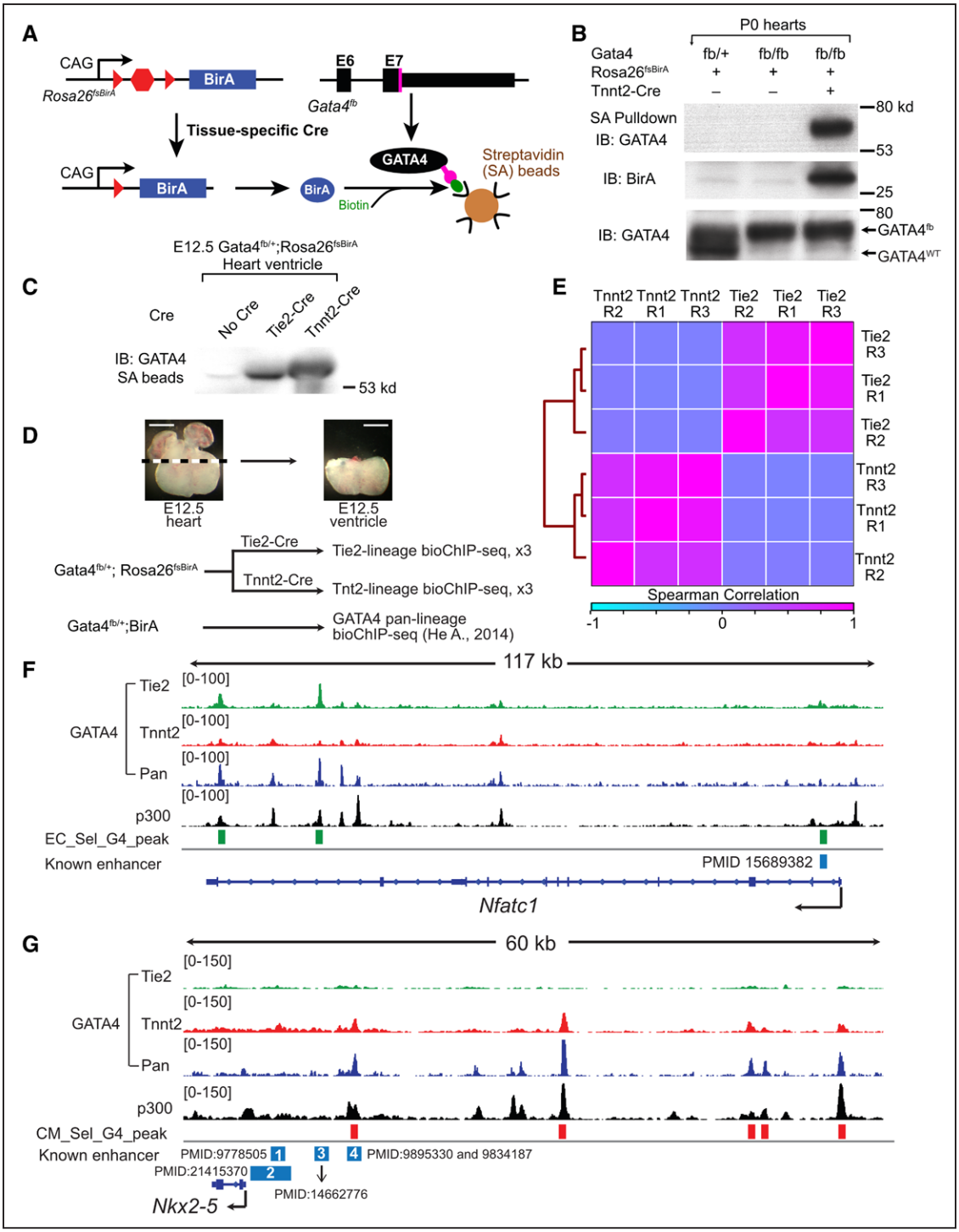

Figure 1. Lineage-selective GATA4 chromatin occupancy in fetal heart.

A, Schematic of Cre-directed in vivo GATA4 biotinylation. GATA4 with a C-terminal Flag-bio epitope tag is biotinylated by BirA, expressed in Cre-marked cells. **B**, Western blot demonstrating Cre-dependent GATA4 biotinylation and streptavidin (SA) pull down. P0 heart extract was incubated with SA beads. Input and SA-bound proteins were probed for GATA4 or BirA. **C**, GATA4 biotinylation in E12.5 CMs or ECs, directed by Tnnt2-Cre or Tie2-Cre. Heart lysate proteins bound to SA beads were probed with GATA4 antibody. **D.** Overview of GATA4 bioChIP-seq in cardiomyocyte (CM) and endothelial cell (EC) lineages. Pan-lineage GATA4 data was reported in study by He et al.[11] Representative images show dissected tissues used for GATA4 bioChIP-seq. Bar, 500 µm. **E**, Correlation of lineage selective GATA4 bio-ChIP-seq data. Clustered heatmap displays the Spearman correlation across the union of peak regions. **F** and **G**, Representative genome browser views of lineage-selective GATA4 bioChIP-seq data at an ECC (*Nfatc1*) and a CM (*Nkx2-5*) gene. Known enhancers were validated by transient transgenesis in the indicated manuscripts. **F**, The known enhancer is no longer active in ventricular ECCs by E12.5. **G**, Known enhancer 4, but not 1-3, are active in E12.5 heart. E12.5 heart ventricle p300 bioChIP-seq data was previously reported.[13]

Cell Type–Specific Chromatin Occupancy of GATA4 Identifies Regulatory Regions Related to Lineage-Specific Function

To systematically compare GATA4 bioChIP-seq signal in EC and cardiomyocyte lineages, we identified 44 464 regions with reproducible GATA4 binding (irreproducible discovery rate[20] <0.05). We then determined the GATA4 signal ratio in *Tie2-Cre* compared with *Tnnt2-Cre* samples (Figure 2A and Data S1). Based on this ratio, we divided the regions into sextiles and classified the top and bottom sextiles as EC-selective and cardiomyocyte-selective, respectively (each 7410 regions; Figure 2A). For comparison, we defined the middle 2 sextiles (14 820 regions) as shared. In all 3 region classes, GATA4 regions were predominantly distal, although among the shared regions relatively more were found within 2 kb of the transcriptional start site (Figure 2B). Subsequent analyses focused on the distal regions (>2 kb from transcriptional start site). We also performed a parallel analysis that identified GATA4 regions with statistically different signal strength between ECs and cardiomyocytes. This complementary analysis yielded very similar results (Figure S4). The sextile-based analysis was used for further studies.

Analysis of the top biological process gene ontology terms associated with each region class demonstrated that they are functionally distinct (Figure 2C). EC-selective regions were highly enriched for gene ontology terms related to vasculature development and blood vessel morphogenesis, whereas cardiomyocyte-selective regions related to cardiac muscle development. Interestingly, shared regions were not enriched for either vasculature or heart muscle-related gene ontology terms; rather, these regions were enriched for general biological process terms such as cell-substrate adhesion, regulation of protein ubiquitination, and TGF beta receptor signaling.

GATA4 interacts with the histone acetyltransferase EP300[21] to stimulate acetylation of H3K27ac,[11] the epigenetic mark written by EP300. Furthermore, EP300 and H3K27ac are validated markers of activated enhancers.[22–24] We recently reported EP300 and H3K27ac occupancy in fetal heart tissue, which presumably marks enhancers from both cardiomyocyte and EC lineages.[13] Consistent with the role of GATA4 in lineage-specific transcriptional regulation, most EC-selective and cardiomyocyte-selective GATA4 regions overlapped with EP300 and H3K27ac regions in the fetal heart (Figure 2D and 2E; Data S1). In contrast, EP300 and H3K27ac regions in the fetal forebrain had little significant overlap with the EC-selective or cardiomyocyte-selective regions, whereas there was greater overlap with the shared GATA4 regions (Figure 2D and 2E). These results were reinforced by analysis of H3K27ac and EP300 ChIP-seq signal in the lineage-specific regions, which showed significant signal from heart but not forebrain (Figure 2F). In contrast, there was significant signal from both heart and forebrain H3K27ac and EP300 ChIP-seq in shared regions (Figure 2F), consistent with many of these regions being general rather than tissue-specific enhancers. Consistent with GATA4 function as a pioneer factor, EC-selective or cardiomyocyte-selective GATA4 regions largely concurred with EC and cardiomyocyte-accessible chromatin, respectively, as assessed by ATAC-seq (Figure 2D and 2E; discussed further in Figure 4).

These results indicate that lineage-selective GATA4 bioChIP-seq identifies genomic regions with features of active enhancers that are associated with lineage relevant biological processes.

Association Between Cell Type–Specific Chromatin Occupancy and Gene Expression

To evaluate the relationship between lineage-selective GATA4-bound regions and genes selectively expressed in ECs or cardiomyocytes, we isolated ECs and cardiomyocytes from E12.5 mouse heart ventricles and measured gene expression by RNA sequencing (Figure 3A; Table S1; Data S2). PCA analysis showed large differences between EC and cardiomyocyte transcriptomes (Figure 3B). Known ECC genes such as *Nfatc1*, *Hapln1*, *Ets1*, *Pecam1*, and *Kdr* and cardiomyocyte genes such as *Pln*, *Ankrd1*, and *Nkx2-5* showed differential expression as expected, consistent with high quality expression data (Figure 3C).

We studied the relationship between genes with EC-selective or cardiomyocyte-selective GATA4 occupancy and lineage-biased gene expression. Genes neighboring EC-selective GATA4 regions were highly enriched for EC-biased gene expression and depleted for cardiomyocyte-biased gene expression, and reciprocally genes neighboring cardiomyocyte-selective GATA4 regions were highly enriched for cardiomyocyte-biased genes and depleted for EC-biased gene expression (Figure 3D). The association between GATA4 lineage-selective occupancy and lineage-biased gene expression was further supported by gene set enrichment analysis, in which genes nearest to EC- or cardiomyocyte-selective GATA4 regions were enriched among genes with EC- or cardiomyocyte-biased expression (Figure 3E). These analyses showed that genes neighboring regions with lineage-specific GATA4 occupancy are highly enriched for genes with lineage-selective expression.

Together, these data indicate that lineage-selective GATA4 regions are regulatory regions that promote lineage specific gene expression.

Lineage-Selective Chromatin Occupancy of GATA4 Correlates With Chromatin Accessibility

Different cell types have different chromatin accessibility patterns, which regulate TF chromatin occupancy. As a pioneer factor, GATA4 is expected to play

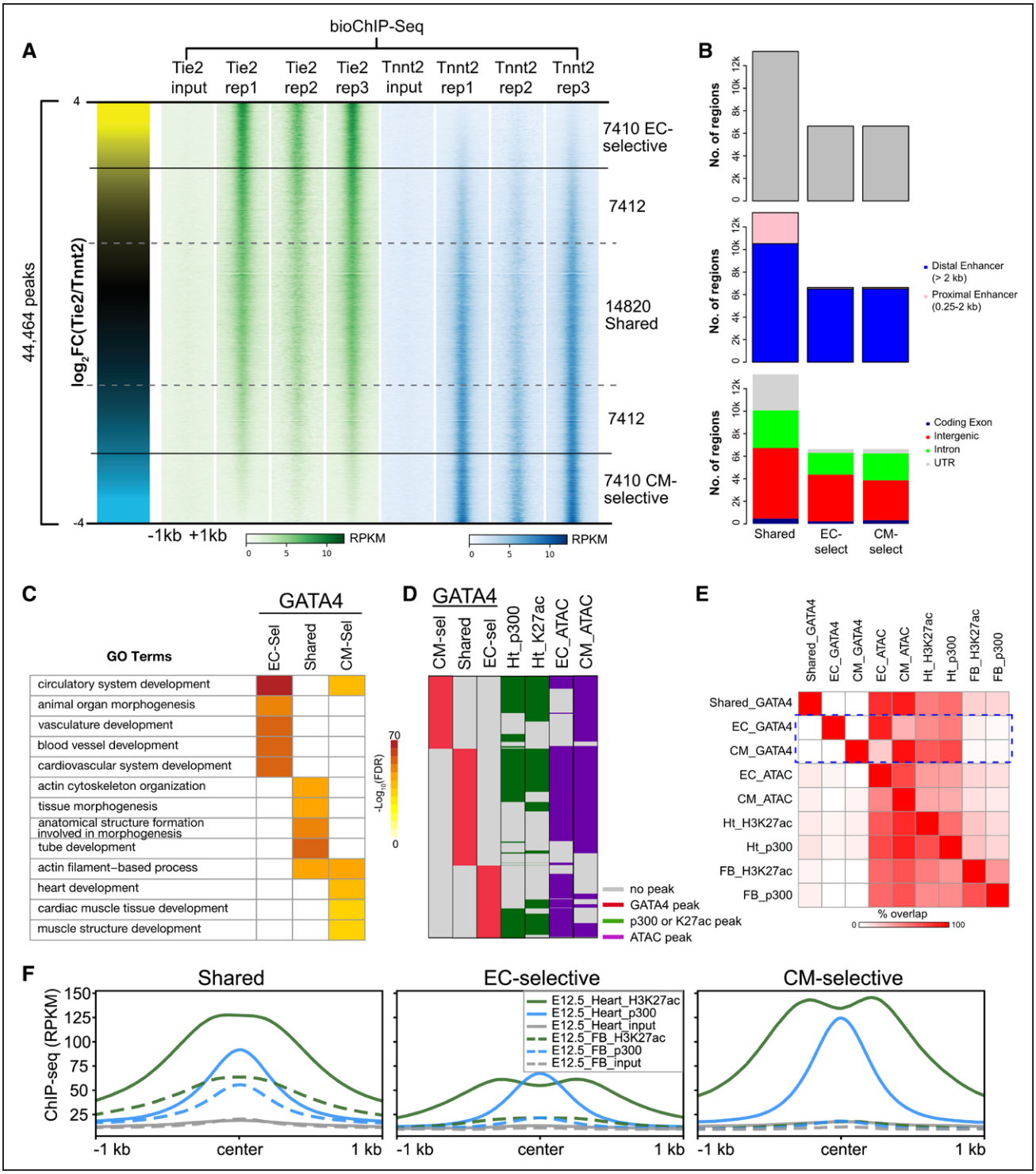

Figure 2. Features of lineage-selective GATA4 regions.

A, Lineage-selective GATA4 regions were defined by the ratio of GATA4 signal in reproducible Tie2- and Tnnt2-peaks. Each row represents a 2-kb region centered on a GATA4 peak, with the color intensity proportional to the GATA4 signal. **B**, Characteristics of GATA4 regions with respect to genome annotations. **C**, Gene Ontology terms enriched among shared, endothelial cell (EC)-selective, and cardiomyocyte (CM)-selective GATA4 regions. The top 5 GO terms for each GATA4 region class are shown. Regions were associated to genes using the basal plus extension rule of GREAT.[43] **D** and **E**, Overlap of chromatin features of active enhancers with GATA4 regions. D, GATA4 region overlap with heart p300 bioChIP-seq or H3K27ac ChIP-seq,[13] or EC- or CM- ATAC-seq. **E**, Pairwise overlap of region sets. Shared GATA4 regions (black rectangle) overlapped enhancer marks in both heart and forebrain, whereas lineage-selective regions (blue rectangle) were preferentially open in the concordant heart lineage and selectively overlapped heart but not forebrain p300 or H3K27ac. **F**, H3K27ac or p300 ChIP-seq signal at GATA4 peaks. GATA4 regions had significant p300 and H3K27ac signal in heart. Shared, but not lineage-selective, GATA4 regions also had significant p300 and H3K27ac signal in forebrain.

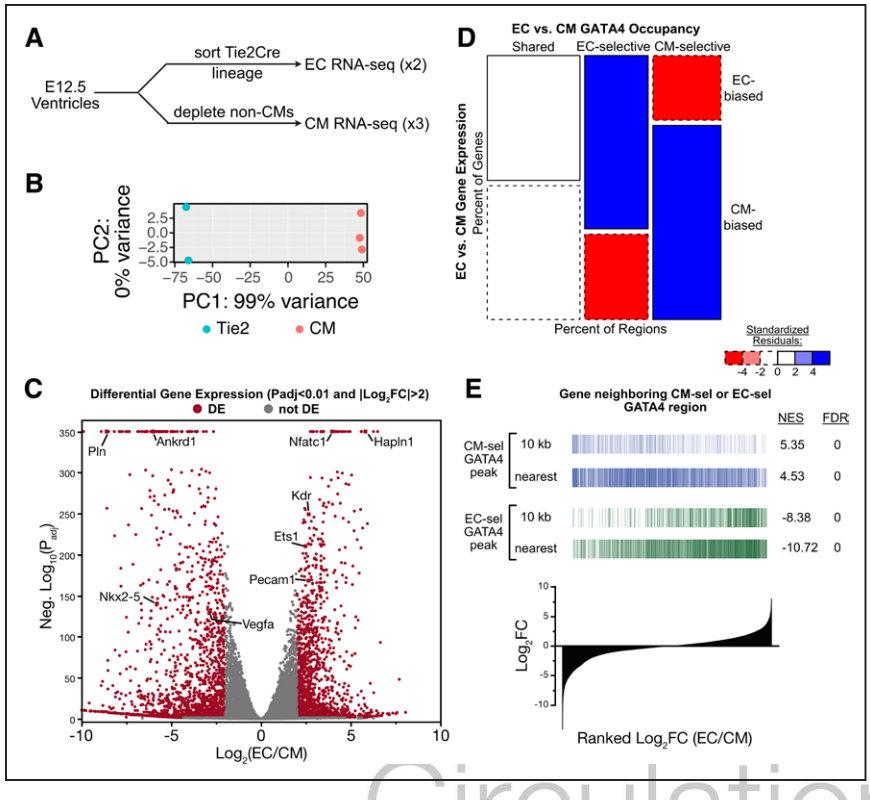

Figure 3. Lineage-selective GATA4 regions and lineage-biased gene expression.
A, Schematic for RNA-seq measurement of gene expression in endothelial cell (EC) and cardiomyocyte (CM) lineages. Hearts from 1 litter (6–10 embryos) were pooled and dissociated for each sample. **B**, PCA plot showing tight grouping of biological replicate EC and CM samples. All expressed genes were used. **C**, Volcano plot of EC or CM differential gene expression. Selected lineage-biased genes are labeled. **D** and **E**, Relationship of lineage-biased gene expression and lineage-selective GATA4 regions. **D**, Mosaic plot of EC- vs CM-selective gene expression versus GATA4 occupancy. Regions were associated with the nearest gene. **E**, Enrichment of genes associated to EC-selective or CM-selective GATA4 regions among EC- or CM-biased genes. Regions were associated to genes whose TSS was within 10 kb, or to the nearest gene. Genes were ranked by the ratio of their expression in ECs versus CMs. NES indicates normalized enrichment score (NES) calculated by GSEA.

an important role in establishing accessible chromatin patterns.[7] To investigate the relationship between lineage-specific GATA4 occupancy and chromatin accessibility in the developing heart, we isolated ECs and cardiomyocytes from E12.5 mouse heart ventricles and measured chromatin accessibility using ATAC-seq (Figure 4A; Table S1). Biological replicates were highly correlated (Figure 4A). ATAC-seq signal was concordant with GATA4 occupancy of EC-selective and cardiomyocyte-selective regions (Figures 2D, 2E, 4B, and 4C; Figure S3). For example, the previously validated intronic *Notch1* enhancer[13] overlapped an EC-selective GATA4 region with an ATAC-seq peak in ECs but not in cardiomyocytes (Figure 4D), whereas cardiomyocyte-selective GATA4 regions upstream of *Myl2* had a cardiomyocyte-specific ATAC-seq peak (Figure 4E). Interestingly, regions with shared occupancy had the strongest ATAC-seq signal, especially in cardiomyocytes (Figure 4B and 4C). These results demonstrate that regions with lineage-selective GATA4 binding are strongly associated with lineage-specific chromatin accessibility.

To determine whether GATA4 is necessary to maintain lineage-specific chromatin accessibility, we performed ATAC-seq on FACS-sorted endocardial cells from E12.5 *Gata4*$^{fl/fl}$*; Cdh5-CreERT2*; *Rosa26*^{mTmG} (EC-specific knockout) or *Gata4*$^{fl/+}$*; Cdh5-CreERT2; Rosa-26*^{mTmG} littermate (control) embryos that were treated with tamoxifen at E9.5 (Figure S5A and S5B). In *Gata4* knockout, regions with significant loss of ATAC-seq signal were rare, and there was no significant bias toward loss versus gain of chromatin accessibility (Figure S5C and S5D). This indicated that GATA4 is not required to maintain chromatin accessibility. However, these data do not exclude an essential role for GATA4 in initiating establishment of open chromatin at the EC-selective GATA4 regions.

Co-Motifs Enriched in Lineage-Selective GATA4-Binding Regions

GATA4 frequently occupies regions in collaboration with partner TFs.[12] TF binding with nucleosome-free chromatin protects the underlying DNA from cleavage by nucleases or Tn5 transposase, yielding a footprint in accessibility signal.[25,26] Motif analysis of footprinted regions within TF bound regions identifies candidate partner TFs.

To identify GATA4 partner TFs at EC-selective and cardiomyocyte-selective regions, we analyzed these regions for ATAC-seq footprints and then performed motif analysis on the footprinted regions (Figure 5A). As expected, GATA4 motifs were highly enriched in EC-selective, cardiomyocyte-selective, and shared GATA4 regions. In EC-selective GATA4 regions, footprints were highly enriched for the motifs of ETS, NFAT, and SOX TFs. Several ETS factors are highly expressed in cardiac ECs (Figure S6A). *Ets1* is highly expressed in ECs and its mutation has been implicated in the pathogenesis of hypoplastic left heart syndrome, a severe form of congenital heart disease.[27,28] *Nfatc1* is selectively expressed in ECCs (Figure S6A), where it regulates valve development,[29] and GATA4 has been

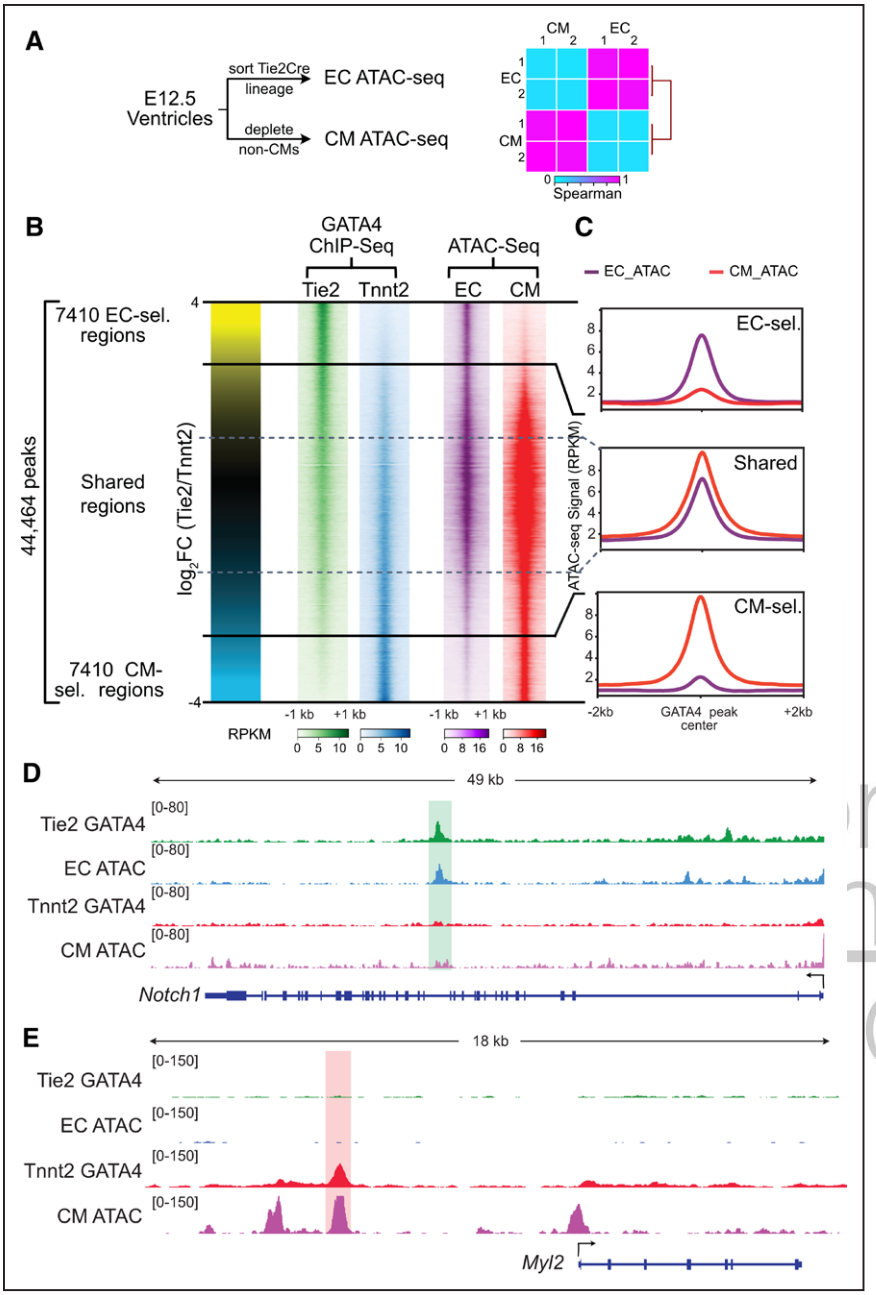

Figure 4. Lineage selective GATA4 regions and lineage selective chromatin accessibility.
A, Schematic for measuring lineage-selective chromatin accessibility using ATAC-seq. E12.5 endothelial cells (ECs) and cardiomyocytes (CMs) were isolated and chromatin accessibility was measured by ATAC-seq. Heatmap shows Spearman correlation between EC and CM ATAC-seq data across the union of identified peaks. **B**, Lineage-selective GATA4 regions and lineage-selective chromatin accessibility. Each row represents a GATA4 region, and the color intensity indicates ChIP-seq or ATAC-seq signal. **C**, ATAC-seq signal at GATA4 regions. EC- and CM-selective GATA4 regions were accessible in the concordant lineage. **D** and **E**, Representative browser views showing lineage selective GATA4 and ATAC-seq data.

shown to interact with NFAT family members in the heart.[30] SOX family TFs are also expressed in ECs and regulate vessel and cardiac valve development.[31] In cardiomyocyte-selective GATA4 regions, MEF2, T-box, and NKX2-5 motifs were highly enriched. GATA4 makes protein-protein interactions with these TFs, and they collaboratively occupy chromatin with GATA4 at active cardiomyocyte enhancers.[12] Consistent with this motif analysis, cardiac TFs including TBX5, NKX2-5, and MEF2C preferentially co-occupied cardiomyocyte-selective but not EC-selective GATA4 regions (Figure S6B).

To complement this analysis, we examined differential motif footprints in EC and cardiomyocyte ATAC-seq data.[26] ETS and FOX-ETS motifs were the top motifs differentially enriched in EC ATAC-seq at EC-selective GATA4 regions, and MEF2, TEAD, and NKX2-5 motifs were among the top motifs differentially enriched in cardiomyocyte ATAC-seq at cardiomyocyte-selective GATA4 regions (Figure 5B and 5C). Interestingly, the cardiomyocyte-enriched footprints included the motif of an ETS-basic helix-loop-helix heterodimer (ETV5-HES7), suggesting that some ETS family-containing complexes may also be active within cardiomyocytes.

In summary, motif analysis of ATAC-seq footprints within GATA4 regions identified likely lineage-restricted GATA4 partner TFs. Among these candidates were ETS family TFs in ECs and NKX2-5 in cardiomyocytes.

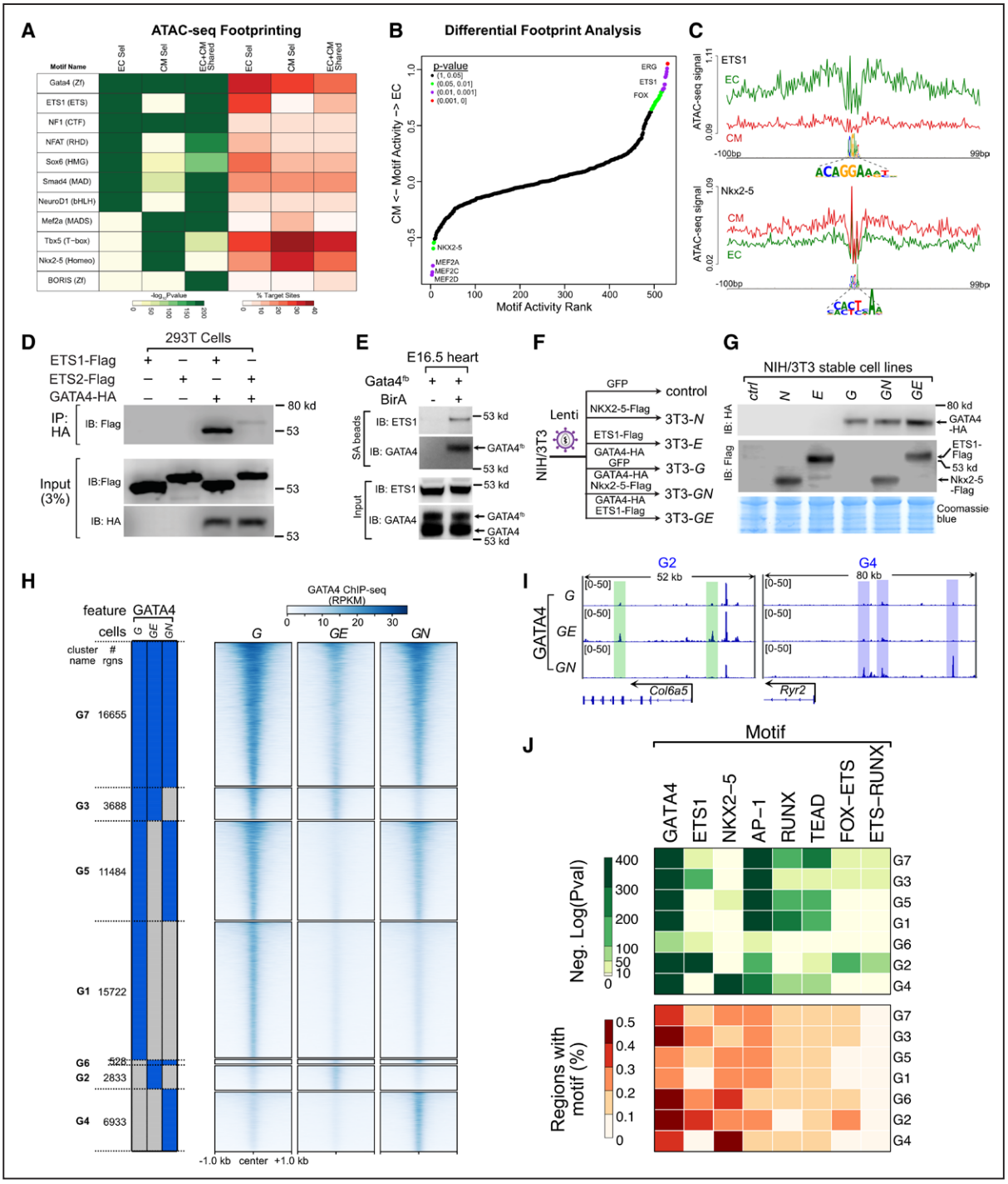

Figure 5. Effect of ETS1 and NKX2-5 on GATA4 chromatin occupancy.

A, Footprint analysis of endothelial cell (EC) and cardiomyocyte (CM) ATAC-seq regions identified enriched TF (transcription factor)-binding motifs. EC-selective and CM-selective GATA4 regions were enriched for the ETS1 and NKX2-5 motifs, respectively. **B**, Differential ATAC-seq footprint analysis. EC or CM ATAC-seq at EC-selective or CM-selective GATA4 regions were analyzed for differences in motif enrichment at footprints. Differences in motif activity between lineages were ranked and statistically scored by HINT-ATAC.[26] **C**, Lineage-selective footprints of the FOX-ETS and NKX2-5 motifs within EC or CM ATAC-seq signal at EC-selective and CM-selective GATA4 regions, respectively. **D**, GATA4 and ETS1 interaction by co-IP assay in transfected 293T cells. IP of HA-GATA4 co-precipitated ETS-FLAG. **E**, ETS1 co-precipitated with GATA4 in E16.5 heart ventricles. Biotinylated GATA4 from Gata4^{fb};BirA mouse embryonic heart lysate was pulled down by streptavidin (SA) beads. Precipitated proteins were probed with ETS1 antibody. Negative control, Gata4^{fb} hearts without BirA. **F**, Schematic of stable cell lines used to study ETS1 or NKX2-5 effect on GATA4 occupancy in NIH3T3 cells. Stable NIH3T3 cell lines expressing the indicated proteins were developed using lentivirus. **G**, Western blot validation of protein expression in the indicated stable cell lines. **H**, GATA4 occupancy in cells stably expressing GATA4, GATA4, and ETS1 (GE), or GATA4 and NKX2-5 (GN). Regions were clustered by their binary pattern of occupancy across these cell lines. **I**, Representative genome browser views of differential GATA4 binding in GE (green shading) or GN (blue shading) cells. **J**, Selected TF motifs enriched among the GATA4-binding clusters.

Cardiomyocyte-Selective VEGFA Enhancer Dependent Upon the Motifs of Multiple Cardiac TFs

To functionally test if the enriched motifs of cofactors are important for the activity of transcriptional enhancers, we studied a region around *Vegfa* with cardiomyocyte-selective GATA4 occupancy. VEGFA is a potent angiogenic factor with cardiomyocyte-selective cardiac expression by E12.5. In adult heart, GATA4 was reported to directly regulate *Vegfa* through a binding site near the transcriptional start site.[32] However, this site did not significantly bind GATA4 in fetal or adult heart, although it was occupied by p300 (Figure S7A). We identified a cardiomyocyte-selective GATA4 region 23 kb downstream of the transcriptional start site. H3K27ac HiChIP revealed chromatin loops between this candidate enhancer and the *Vegfa* promoter (Figure S7B), suggesting that it likely regulates *Vegfa* expression. This highly conserved 360 bp region contains GATA, MEF2, NKX2-5, and TBX5 motifs and is co-occupied by these factors and EP300 in E12.5 and adult ventricles (Figure S7A).

To measure the enhancer activity of this region in vivo in cardiomyocytes, we used AAV9, which in heart selectively and efficiently transduces cardiomyocytes (Figue S7C). We cloned the putative *Vegfa* enhancer into an AAV enhancer-reporter construct,[12] positioning it within the 3' UTR so that it drives its own transcription (Figure S7D). After delivery of AAV-mCherry-enh (*Vegfa*^{WT}) to newborn mice, we visualized mCherry at postnatal day 7 (P7; Figure S7D). The *Vegfa*^{WT} enhancer drove robust cardiac mCherry expression that was reduced by mutation of the GATA4 motif (Figure S7A, S7E). Mutation of the T-Box, MEF2, or NKX2-5 motifs also qualitatively reduced reporter expression. To more quantitatively assess relative enhancer activity, we measured the level of enhancer-driven RNA expression, normalized to AAV DNA to control for transduction efficiency (Figure S7E). This demonstrated that the *Vegfa*^{WT} enhancer was ~4-fold more active than control, mutation of single TF motifs partially reduced activity, and mutation of all 4 motifs abolished enhancer activity (Figure S7G).

To gain broader insight into the value of lineage-selective GATA4 regions to predict cardiac enhancer activity, we evaluated the GATA4 regions against the VISTA database of in vivo tested enhancers (Figure S7H). Pan-GATA4 had very high sensitivity (99%) and modest specificity (70%). Using the cardiomyocyte-selective regions boosted the specificity (98%) at the cost of sensitivity (29%). EC-selective regions did not substantially overlap with active heart enhancers (5%), potentially due to biased selection of tested regions toward cardiomyocyte enhancers and limited detection of endocardial activity in the whole mount transgenic assay.

These results identify a novel *Vegfa* enhancer that is collaboratively regulated in cardiomyocytes in vivo by multiple TFs including GATA4 and NKX2-5 and suggest that the cardiomyocyte-selective GATA4 regions are highly predictive of in vivo cardiac enhancer activity.

GATA4 Interacting Factors Modulate GATA4 Chromatin Occupancy

ATAC-seq footprinting showed that EC-selective GATA4 regions are highly enriched for the ETS motif. Among ETS family members, ETS1 and ETS2 were most highly and selectively expressed in ECC (Figure S6A). To determine if GATA4 can interact with ETS1 or ETS2, we expressed ETS1, ETS2, and GATA4 in 293T cells and performed co-immunoprecipitation (co-IP) assays. For comparison, we selected NKX2-5 as a cardiomyocyte-selective GATA4 partner, because NKX2-5 is selectively expressed in cardiomyocytes, whereas other candidate factors such as MEF2 or TBX5 are expressed in both cardiomyocytes and ECs (Figure S6A). ETS1 strongly co-precipitated with GATA4, whereas ETS2-GATA4 interaction was much weaker (Figure 5D). In side-by-side assays, GATA4-ETS1 interaction was comparable in strength to the well-established interaction between GATA4 and NKX2-5 (Figure S8A). Both NKX2-5-GATA4 and ETS1-GATA4 interactions were resistant to nuclease treatment, indicating that the interaction is mediated by protein-protein contacts rather than DNA co-binding (Figure S8B). GATA4-ETS1 interaction required the ETS DNA-binding domain (Figure S8C). To determine if GATA4 and ETS1 interact in developing hearts, we pulled down endogenous GATA4^{fb} from *Gata4*$^{fb/+}$; *Rosa26*$^{BirA/+}$ embryonic heart lysate and found that ETS1 co-precipitated (Figure 5E). Together, these data support GATA4-ETS1 interaction in endocardium.

Pioneer factors such as GATA4 contribute to establishing lineage specific chromatin landscapes by binding to nucleosome occluded sites and making them accessible to other TFs and chromatin modifying factors. We hypothesized that GATA4 interactions with distinct cofactors in cardiomyocytes and ECs modulates its pioneering activity and chromatin occupancy in each lineage. Addressing this hypothesis required studying the effect of TF interactions on GATA4 binding in cells with little or no endogenous expression of the factors. Therefore we turned to NIH3T3 cells, which do not endogenously express GATA4 or NKX2-5, and lowly express ETS1 (Figure S8D through S8F). We created stable NIH3T3 cell lines that stably expressed GFP (3T3-*control*), GATA4-HA (3T3-*G*), Flag-ETS1 (3T3-*E*), Flag-NKX2-5 (3T3-*N*), GATA4-HA and Flag-ETS1 (3T3-*GE*), or GATA4-HA and Flag- NKX2-5 (3T3-*GN*; Figure 5F). GATA4, NKX2-5, and ETS were comparably expressed between sets of GATA4-, NKX2-5-, and ETS1-programmed cells (Figure 5G). As expected, NKX2-5 and ETS1 both co-precipitated GATA4 in co-expressing cell lines (Figure S8G).

Next, we performed GATA4, ETS1, and NKX2-5 ChIP-seq in these cell lines (Table S1). Biological triplicates correlated well within each group (Figure S9A), indicating highly reproducibility. Between groups, the GATA4 signal correlated across 3T3-*G*, -*GE*, and -*GN* cell lines, and ETS1 and NKX2-5 signal correlated across 3T3-*E*, -*GE*, -*N*, and -*GN* cell lines, respectively. Co-expression of NKX2-5 and GATA4 shifted their occupancy so that both TFs were more correlated (Figure S9A, yellow dashed boxes), and similarly ETS1 and GATA4 co-expression resulted in higher correlation between these factors (Figure S9A, green dashed boxes). These data are consistent with collaborative occupancy by these TF pairs.

We defined GATA4 regions in each cell line as the ChIP-seq peaks called in at least 2 of 3 replicates (Figure 5H). In cells expressing only GATA4, GATA4 occupied 47 606 regions. Interestingly, co-expression of either ETS1 or NKX2-5 reduced the number of GATA4 regions to 23 711 and 35 608, respectively, as a result of lost occupancy of 27 257 and 19 460 regions and gained occupancy of 3362 and 7462 regions, respectively. The regions that gained occupancy in *GE* or *GN* cell lines were largely distinct (Figure 5H and 5I). We grouped GATA4 regions into clusters (G1-G7) based on their TF occupancy in different cell lines (Figure 5H; Figure S9B). These regions had limited overlap with GATA4 regions from embryo hearts, consistent with the differences in cell types (Figure S9C), with the greatest overlap occurring with the GATA4 regions shared between ECs and cardiomyocytes. EC-selective GATA4 regions overlapped most with ETS1-related clusters, and cardiomyocyte-selective heart GATA4 regions overlapped most NKX2-5-related clusters.

Motif analysis of these clusters showed that the GATA4 motif was present in a high percentage (20%–s50%) of regions and highly enriched (Figure 5J). We also detected high enrichment of AP-1, RUNX, and TEAD motifs across most clusters, suggesting potential GATA4 interactions with these TFs. The ETS motif was enriched in clusters in which GATA4 occupancy was contingent on ETS1 expression (G2, G3, G6) whereas it was less prevalent among regions not occupied by GATA4 in the presence of ETS1 (G1, G4, G5) or occupied in all 3 cell lines (G7). This was mirrored by enrichment of the NKX2-5 motif in clusters in which GATA4 occupancy was contingent on NKX2-5 expression (G2, G4, G5) and not in the other clusters. Together, the motif analysis suggests that gained GATA4 occupancy in cells co-expressing ETS1 or NKX2-5 is due to direct co-occupancy of regions by partner TFs, whereas its loss from regions in co-expressing cells occurs through less direct mechanisms, such as competition for a limiting amount of GATA4.

A similar analysis of ETS1 and NKX2-5 ChIP-seq data identified far fewer occupied regions (8678 and 17471, respectively; Figure S9B through S9D). Clusters in which a high percentage of GATA4 regions had motifs of partner TFs ETS1 or NKX2-5 (eg, G2, G3, and G6 or G4, G5, and G6, respectively), had a surprisingly low fraction of regions with ChIP-seq signal for these TFs. We interpreted these data to indicate that the Flag antibody was less sensitive for ChIP-seq, which limited comparative analysis of relationships between GATA4 and ETS1 or NKX2-5 occupancy.

Effect of ETS1 Versus NKX2-5 on GATA4 Pioneer Activity

Pioneer TF activity can be defined as binding to nucleosomal DNA (pioneer binding, Figure 6A, step 1), with its functional outcomes being increased accessibility of bound DNA (pioneer opening, Figure 6A, step 2) and establishment of an active enhancer (pioneer enhancer activation; Figure 6A, step 3). We assessed each of these outcomes in the stable cell lines. As detailed below, only a subset of regions with pioneer binding became accessible and decorated with active enhancer marks.

To assess GATA4 pioneer binding activity and how it is modified by ETS1 or NKX2-5, we measured chromatin accessibility in control and in the cells expressing GATA4, NKX2-5, and ETS1 using ATAC-seq (Figure 6B; Table S1). Due to its pioneer binding activity,[7] GATA4 can bind inaccessible DNA (ATAC-seq negative; Figure 6C). Of regions occupied by GATA4 without a partner TF (G1, G3, G5, G7, Figure S10A), 46% were inaccessible in 3T3-*control* (GATA4-naive) cells (Figure 6D; Figure S10B). In contrast, regions occupied by ETS1 and NKX2-5 without GATA4 were only 13% and 12% inaccessible in control cells (E1, E3, N1, N3, Figure S10B; Fisher P<2.2E-16 vs. GATA4 regions). These data are consistent with GATA4, but not ETS1 or NKX2-5, having substantial pioneer binding activity.

However, non-pioneer TFs ETS1 and NKX2-5 modified the pioneer binding activity of GATA4 (Figure 6A, step 1; Figure 6C and 6D). Regions bound by GATA4 only in the presence of either ETS1 or NKX2-5 (G2, G4, G6), were 85% closed in control cells, compared with 56% for regions bound by GATA4 without these partner TFs (G1, G3, G5, G7; Fisher exact test P<2.2E-16; Figure 6D). Compared with NKX2-5, ETS1 more potently enhanced GATA4 pioneer binding activity, since the inaccessible fraction was higher in ETS1-dependent compared with NKX2-5-dependent regions (G2, 95%, versus G4, 80%, Fisher P<2.2E-16; and G3, 87%, versus G5, 45%, Fisher $P<2.2\times10^{-16}$). These observations were supported by quantitative comparison of ATAC-seq signal in control cells at G1-G7 regions (Figure S10C). These data indicate that despite GATA4's pioneering binding activity, chromatin accessibility influences GATA4 occupancy, and the extent and sites of pioneering binding activity are determined by GATA4 interaction with partner TFs.

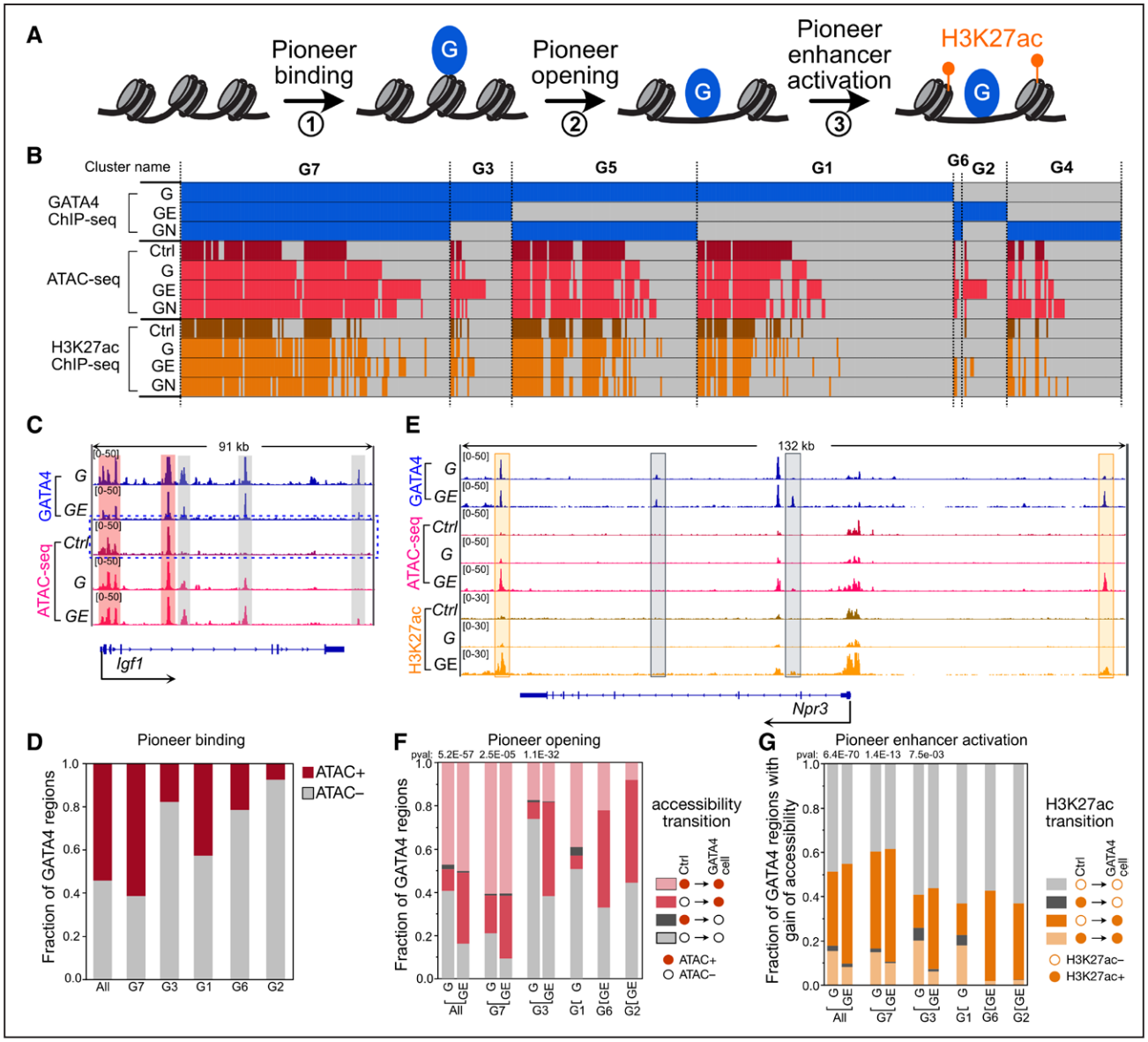

Figure 6. ETS1 and NKX2-5 enhance GATA4 pioneer activity.
A, Model for steps of pioneer TF (transcription factor) activity. **B**, Overview of chromatin accessibility and H3K27ac acetylation at GATA4 regions in cell lines expressing GATA4 (G), GATA4 plus ETS1 (GE), GATA4 plus NKX2-5 (GN), or GFP (Ctrl). G1-G7 denote regions grouped by their pattern of GATA4 occupancy across cell lines. **C**, Representative genome browser view of GATA4 pioneer binding at a region that was inaccessible in 3T3-Ctrl (gray shading), and additional regions that were accessible in 3T3-Ctrl (red shading). **D**, Fraction of GATA4 regions that were accessible or inaccessible in control cells. All refers to all GATA4 regions across all cell lines. Nearly half of GATA4 regions were not accessible in control cells, indicative of pioneer binding. **E**, Representative genome browser views of pioneer opening and enhancer activation by GATA4 in the presence of ETS1 (orange shading). **F**, Change in accessibility with GATA4 occupancy. The distribution of 4 possible chromatin accessibility transitions from control cells (GATA4-naive) is plotted for the GATA4 clusters relevant for each GATA-expressing cell line. The fraction of closed regions in control cells that became open in the GATA4 cells was compared using the Fisher exact test without multiple testing correction. **G.** The distribution of H3K27ac transitions for regions that gained accessibility is shown. GE cells had a consistently high fraction of open regions that gained H3K27ac, compared with G cells lacking ETS1. Nominal Fisher *P* indicates comparison of the proportion of GATA4 regions with gain of accessibility that undergo enhancer activation between cells expressing G vs GE.

We next evaluated GATA4's pioneer opening activity, i.e. its ability to make closed chromatin accessible (Figure 6A, step 2; Figure 6E and 6F). Surprisingly, GATA4 alone had low pioneer opening activity—only 7% of chromatin that was inaccessible in control cells became accessible upon GATA4 occupancy in 3T3-*G* cells (Figure 6F). Presence of ETS1 or NKX2-5 augmented GATA4's pioneer opening activity, such that 63% and 43% of GATA4 occupied regions that were inaccessible in control cells became accessible in 3T3-*GE* or 3T3-*GN*, respectively (Fisher P<2.2E-16 compared with 3T3-*G*; Figure 6F; Figure S10D). ETS1 consistently exhibited greater pioneer opening activity than NKX2-5. For example, at regions occupied by GATA4 in both

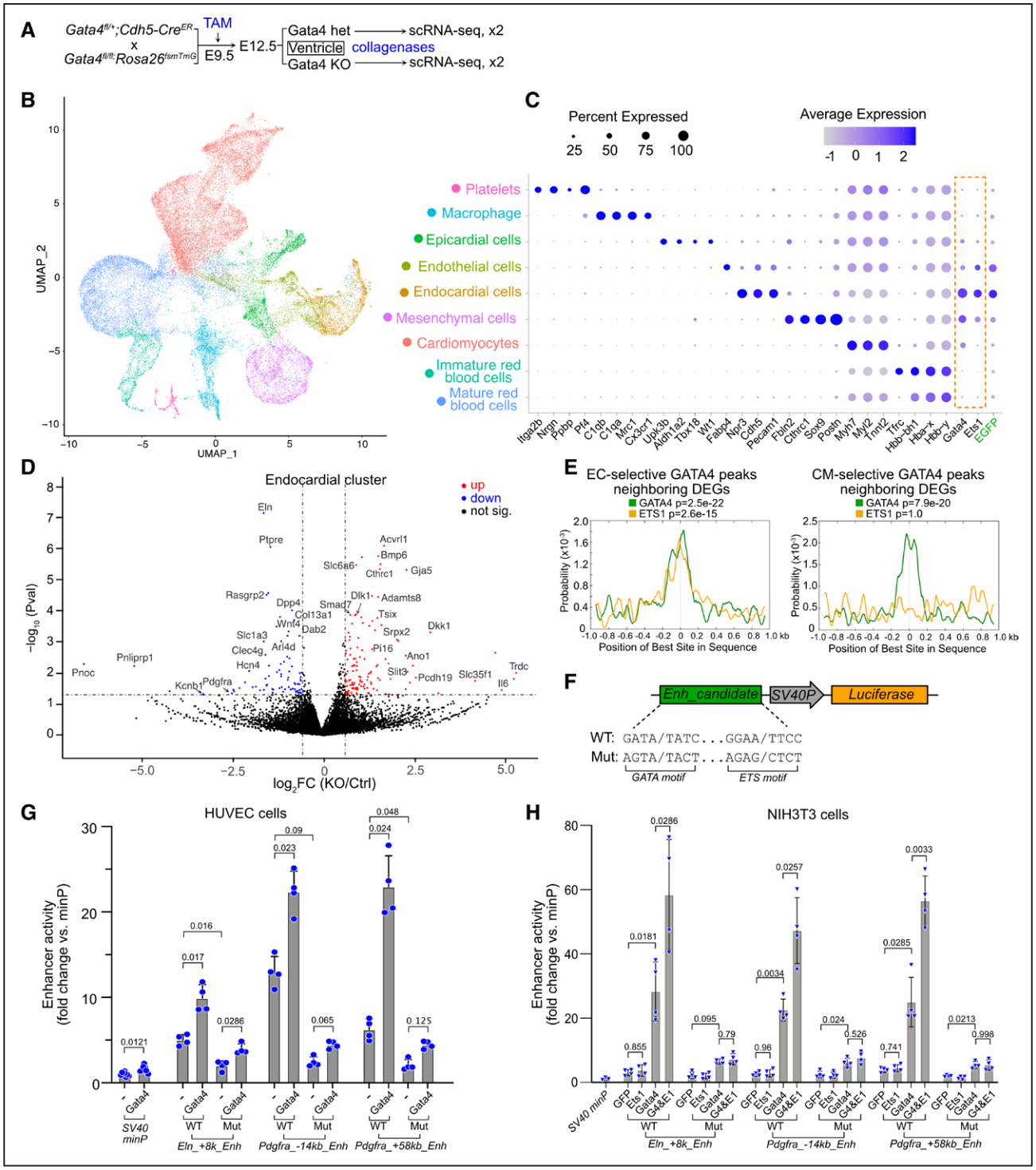

Figure 7. GATA4 and ETS1 synergistically regulate endocardial cell gene expression.
A, Schematic for scRNA-seq measurement of gene expression after Gata4 ablation by *Cdh5-CreERT2*. Four to five hearts were pooled for each of 2 replicates per group. **B**, UMAP visualization of cell clusters identified in E12.5 mouse ventricles. **C**, Expression of *Gata4*, *Ets1*, and marker genes of cell clusters. GFP marks cells exposed to *Cdh5-CreERT2* recombinase activity. **D**, Differentially expressed genes (DEGs) in *Gata4* KO cells (*P*adj<0.05 and absolute fold-change >1.5). **E**, Central enrichment of ETS1 and GATA4 motifs within GATA regions related to DEGs. One-tailed binomial test, adjusted for multiple testing. **F.** Diagram of reporter constructs used to test candidate enhancer activities. SV40P is the minimal promoter from SV40 (SV40 minP in **G** and **H**). **G** and **H**, GATA4 and ETS1 synergistically activate candidate endocardial enhancers through cognate binding motifs. Assays were performed in HUVEC (**G**) or NIH3T3 (**H**). ANOVA with Tukey post-hoc test within each enhancer tested.

3T3-*GE* and 3T3-*GN* (G6 and G7), the percentage of regions that underwent pioneer opening was 73% and 56% in 3T3-*GE*, compared with 54% and 34% in 3T3-*GN* (Fisher P<2.2E-16 and 7.6E-11, respectively; Figure S10D). Similarly, the fraction of regions that were opened by GATA4 plus ETS1 was 50% (G2 and G3), compared with 35% for GATA4 plus NKX2-5 (G4 and G5; Fisher P<2.2E-16; Figure S10D). Analysis of ATAC-seq signal changes between control and TF-expressing cells (Figure S10E and S10F) further supported stimulation of greater chromatin opening at regions with low ATAC signal in control cells in GATA4 plus ETS compared with GATA4 alone, with GATA4 plus NKX2-5 displaying intermediate chromatin opening. These data demonstrate that non-pioneering TFs augment and direct the chromatin opening activity of pioneering TFs.

We then analyzed the effectiveness of GATA4 to establish active enhancers by performing H3K27ac ChIP-seq in the control and stable TF-expressing cell lines (Figure 6A, step 3; Figure 6E; Figure S10H, Table S1). As expected, the large majority (92-97%) of H3K27ac+ regions were accessible in 3T3-*G*, 3T3-*GE*, and 3T3-*GN* (Figure S10G). Regions that gained accessibility through GATA4 pioneer activity were consistently less likely to gain H3K27ac than regions persistently accessible in both control and GATA4-expressing (Figure S10I), suggesting that GATA4 activates enhancers more efficiently at persistently open sites than at sites that require pioneering activity. Next, we analyzed how ETS1 or NKX2-5 influence enhancer activation at sites of GATA4 pioneering activity. Regions that gained accessibility became consistently activated at the highest frequency by GATA4 plus ETS1 (Figure 6G; Figure S10J). GATA4 plus NKX2-5 activated enhancers at comparable frequency to GATA4 alone.

The overall distribution of chromatin state transitions observed in 3T3-*G*, 3T3-*GE*, and 3T3-*GN* is summarized in Figure S10K. Collectively, our analyses indicate that ETS1 enhances GATA4's pioneer binding, opening, and enhancer activation. NKX2-5 less potentially stimulated GATA4 pioneer binding and opening, and did not increase enhancer activation at sites requiring pioneer activity.

GATA4 and ETS1 Synergistically Regulate Endocardial Enhancers

To investigate the functional requirement of GATA4 in cardiac ECs, we selectively inactivated *Gata4* in ECs using *Cdh5-CreERT2* and tamoxifen treatment at E9.5 (Figure 7A). At E12.5, heart ventricle apexes were analyzed by scRNAseq, in biological duplicate (Figure S11A through S11C). Forty-one thousand three hundred eighty-fives cells that passed quality control. Grouping these cells by their transcriptomes yielded nine cell clusters (Figure 7B). Duplicate samples agreed well (Figure S11A and S11B) and were merged for further analyses. Clusters were identified by their expression of known marker genes (Figure 7C and Figure S11E through S11H). Two EC clusters were marked by *Cdh5*, *Pecam1,* and GFP, here a Cre-activated lineage marker (Figure S11D and S11E). One EC cluster expressed ECC markers *Nfatc1*, *Npr3*, and *Hapln1*[28] (Figure S11F), while the other expressed vascular endothelial cell marker *Fabp4*[33] (Figure S11G). *Gata4* expression was the highest in ECCs and mesenchymal cells, intermediate in cardiomyocytes and epicardial cells, and lower in vECs (Figure 7C; Figure S11I). *Ets1* expression was the highest in ECCs and vECs, intermediate in mesenchymal cells, and low in cardiomyocytes and epicardial cells.

To identify ECC genes regulated by GATA4, we performed differential gene expression analysis between *Gata4* knockout and control (*Gata4*$^{fl/+}$*; Cdh5-CreERT2*) cells within the ECC cluster. There were 278 differentially expressed genes (DEGs; P_{adj}<0.05 and absolute fold-change >1.5; 129 upregulated and 76 downregulated; Data S3). These DEGs included genes with known function in heart development, such as *Pdgfra,*[34,35] *Wnt4*,[36] and *Bmp6*[37] Gene ontology analysis showed that the DEGs were enriched for functional terms involving HIF-1, Hippo, and PI3K-Akt signaling pathways (Figure S11J). The DEGs in ECCs had minimal overlap with DEGs in other cell types (Figure S11K). The 87 DEGs in ECCs were associated with EC-selective GATA4 regions, suggesting that these genes are the direct targets of GATA4 in developing endocardium (Figure S11L; Data S3), In comparison, DEGs in ECCs were less likely to be associated with cardiomyocyte-selective GATA4 regions (Figure S11L). The ETS1 motif co-localized with the GATA4 motif at the center of these EC-selective GATA4 regions, whereas cardiomyocyte-selective GATA4 regions neighboring DEGs were enriched for GATA4 but not ETS1 motifs (Figure 7E). These data suggest that ETS1 collaborates with GATA4 to regulate these DEGs in ECCs.

To further test if GATA4 and ETS1 co-regulate enhancers of these DEGs, we focused on 2 downregulated genes, *Eln* and *Pdgfra*. We cloned three neighboring EC-selective GATA4 regions (Figure S11M through S11O) upstream of a minimal promoter and firefly luciferase (Figure 7F; Figure S11P). We also generated mutant versions in which GATA4 and ETS1 motifs were ablated (Figure 7F; Table S2). We tested the activity of these enhancers by transfecting them into human umbilical vein endothelial cells, which endogenously express ETS1 and GATA4 (Figure S11Q). The enhancers demonstrated robust activity that was augmented by co-transfection of GATA4 and diminished by mutation of GATA4 and ETS1 motifs (Figure 7G). To circumvent the endogenous expression of GATA4 and ETS1 in human umbilical vein endothelial cells, we turned to NIH3T3 cells, which express no GATA4 and low ETS1 (Figure S11R). Overexpression of GATA4 greatly stimulated the activity of all 3 enhancers (Figure 7H). While ETS1 overexpression did

not substantially stimulate enhancer activity by itself, it significantly boosted enhancer activation by GATA4 (Figure 7H). Collectively, these results indicate that GATA4 and ETS1 synergistically regulate a subset of EC-selective GATA4 enhancers in endocardial development.

DISCUSSION

Organ development requires coordinated interactions between multiple different cell types, each with its own transcriptional regulatory program. Here, we used Cre-directed bioChIP-seq to map occupancy of the critical cardiac pioneer factor GATA4 in EC and cardiomyocyte lineages. We identified regions selectively occupied by GATA4 in these lineages and showed that these regions have properties of lineage-specific enhancers, including signature chromatin landscape features (tissue-specific ATAC, H3K27ac, and EP300 signals) and enrichment adjacent to lineage-selective genes. Lineage-selective GATA4 regions were associated with lineage specific functional terms and with cardiac specific enhancer marks. In contrast, GATA4 regions shared between cardiomyocyte and EC lineages were not enriched for lineage-selective functional annotations and shared enhancer features with the forebrain. This observation suggests that cell-restricted GATA4 regions are more likely to participate in cell-type specific gene regulation.

We functionally tested 1 lineage-selective enhancer associated with *Vegfa*, a key angiogenic factor. *Vegfa* is highly expressed in developing cardiomyocytes,[38] and heart development and embryo survival is highly dependent upon *Vegfa* dose.[39,40] GATA4 was previously identified as a regulator of *Vegfa* expression in adult heart, and targeted ChIP and reporter assays suggested direct GATA4 stimulation of the *Vegfa* promoter.[32] Our GATA4 ChIP-seq studies of fetal and adult heart did not identify GATA4 occupancy of the *Vegfa* promoter. Cell lineage-selective GATA4 ChIP-seq identified a cardiomyocyte-selective GATA4 region that directly contacts the *Vegfa* transcriptional start site 23 kb away. This region functioned as a cardiomyocyte enhancer and was dependent upon motifs of GATA4 and multiple other cardiac TFs. Given the importance of *Vegfa* expression for cardiac development and function,[39,40] it is likely that this is a functionally important *Vegfa* enhancer.

Footprints within EC-selective GATA4 regions were highly enriched for the ETS motif, and we demonstrated that GATA4 physically interacts with ETS1, one of the most highly expressed ETS family members in endocardium. Furthermore, through single-cell RNA-seq analysis of embryonic hearts with EC-selective *Gata4* inactivation and integration with EC-selective GATA4 chromatin occupancy, we identified endocardial genes likely to be directly regulated by GATA4. These GATA4-bound enhancers were centrally enriched for GATA4 and ETS1 motifs, and reporter assays demonstrated synergistic activation by GATA4 and ETS1. Together, these data indicate that GATA4-ETS1 interaction governs ECC development.

Considerable attention has been directed toward pioneer TFs because their ability to bind to nucleosomal DNA endows them with the capacity to direct the reshaping of the chromatin landscape, which is necessary to enable and guide lineage specification.[1] Here, we analyzed how interaction of non-pioneer partner TFs, ETS1 and NKX2-5, with a pioneer TF, GATA4, modulates its pioneer activity. We considered 3 separable aspects of pioneer activity: binding to nucleosomal DNA (pioneer binding), opening inaccessible DNA to make it accessible (pioneer opening), and establishing an active enhancer (pioneer enhancer activation). Both ETS1 and NKX2-5 increased GATA4's pioneer binding activity, with ETS1 being more potent than NKX2-5. Without its partner TFs, GATA4 exhibited weak pioneer opening activity. Both ETS1 and NKX2-5 augmented this pioneer opening activity, again with ETS1 being more potent than NKX2-5. GATA4 pioneer enhancer activation—establishment of active enhancer marks at regions that had undergone pioneer opening—was generally weaker than its stimulation of enhancer activation at already accessible chromatin. NKX2-5 did not substantially alter GATA4 pioneer enhancer activation, whereas it was stimulated by ETS1. Together these observations indicate that pioneer activity can be partitioned into distinct steps, and each step can be differentially modulated by specific partner, non-pioneer TFs. In future studies it will be interesting to systematically probe the interaction of pioneer TFs with partner TFs to determine if there are principles that will allow better prediction of the outcome of pioneer TF-partner TF interactions.

ETS1 is in the critical region for Jacobsen syndrome, a chromosomal deletion syndrome that includes hypoplastic left heart syndrome, a severe form of congenital heart disease,[41] and a patient with de novo *ETS1* frameshift mutation had complex heart disease including left ventricular hypoplasia.[42] *ETS1* was downregulated in hypoplastic left heart syndrome iPSC-derived endocardial cells, and *ETS1* knockdown in primary human ECCs shared abnormal gene expression and endothelium to mesenchymal transition signaling with hypoplastic left heart syndrome ECCs.[28] Our work suggests that ETS1 likely regulates ECC development by interacting with GATA4 to modulate its pioneering activity, resulting in both gain and loss of GATA4 bound regions, and increased pioneer binding, opening, and enhancer activation. Additionally, ETS1 and GATA4 interact at a subset of ECC enhancers, resulting in synergistic enhancer activation. Although GATA4 was not required to maintain chromatin accessibility in murine endocardium, we hypothesize that GATA4-ETS1 pioneer activity is required to establish the initial chromatin landscape of endocardium, and that

disruption of this process contributes to hypoplastic left heart syndrome pathogenesis.

Our experiments had some limitations. In the studies of GATA4 pioneering activity, we used NIH3T3 cell lines that stably overexpressed GATA4, ETS1, and/or NKX2-5. We cannot exclude that stable expression of these factors led to alterations of chromatin accessibility or TF binding. Our single-cell RNA sequencing study of GATA4 mutant endocardium used GATA4 inactivation in ECs at E9.5. This was necessary to circumvent embryonic lethality at or before E12.5 caused by earlier GATA4 inactivation in ECs, for example, by Tie2Cre[9] This relatively late inactivation likely missed essential functions of GATA4 for ECC specification and sculpting of the accessible chromatin landscape. Further studies of GATA4- and ETS1-dependent functions in endocardium at earlier endpoints are warranted to further understand the contribution of these TFs to heart development and congenital heart disease.

ARTICLE INFORMATION

Received August 26, 2020; revision received September 29, 2022; accepted October 7, 2022.

Affiliations

Department of Cardiology, Boston Children's Hospital, MA (P.Z., Y.Z., I.S., M.A.T., Y.C., B.N.A., F.X., X.Z., K.L., B.D.L., N.M., Q.M., W.T.P.). Department of Thoracic and Cardiovascular Surgery, Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, China (L.Y.). State Key Laboratory of Membrane Biology, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Institute of Molecular Medicine, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China (A.H.). State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, China (B.Z.). Harvard Stem Cell Institute, Cambridge, MA (W.T.P.).

Sources of Funding

This project was supported by NIH/NHLBI (UM1 HL098166 and R01HL146634) and by charitable donations from the Department of Cardiology, Boston Children's Hospital.

Acknowledgments

P. Zhou conceived of the project, designed and performed experiments, analyzed data, and co-wrote the article. Y. Zhang and I. Sethi performed bioinformatic analyses. Y. Cao contributed H3K27ac HiChIP data of neonatal cardiomyocytes. B.N. Akerberg contributed ATAC-seq data of E12.5 cardiomyocytes. F. Xiao, M.A. Trembley, A. He, and B. Zhou performed experiments. B.D. Jardin, N. Mazumdar, and Q. Ma performed animal husbandry and histology. W.T. Pu oversaw the project, contributed to the analyses, and co-wrote the article.

Disclosures

The authors declare no competing interests.

Supplemental Materials

Expanded Materials and Methods
Tables S1 and S2
Figures S1–S11
Data S1–S3
References[45–63]

REFERENCES

1. Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev.* 2016;37:76–81. doi: 10.1016/j.gde.2015.12.003
2. Donaghey J, Thakurela S, Charlton J, Chen JS, Smith ZD, Gu H, Pop R, Clement K, Stamenova EK, Karnik R, et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet.* 2018;50:250–258. doi: 10.1038/s41588-017-0034-3
3. Meers MP, Janssens DH, Henikoff S. Pioneer factor-nucleosome binding events during differentiation are motif encoded. *Mol Cell.* 2019; Available from: http://dx.doi.org/10.1016/j.molcel.2019.05.025
4. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell.* 2015;161:555–568. doi: 10.1016/j.cell.2015.03.017
5. Chen J, Zhang Z, Li L, Chen B-C, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E, et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell.* 2014;156:1274–1285. doi: 10.1016/j.cell.2014.01.062
6. Lescroart F, Meilhac SM. Cell lineages, growth and repair of the mouse heart. *Results Probl Cell Differ.* 2012;55:263–289. doi: 10.1007/978-3-642-30406-4_15
7. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell.* 2002;9:279–289. doi: 10.1016/s1097-2765(02)00459-8
8. Zeisberg EM, Ma Q, Juraszek AL, Moses K, Schwartz RJ, Izumo S, Pu WT. Morphogenesis of the right ventricle requires myocardial expression of Gata4. *J Clin Invest.* 2005;115:1522–1531. doi: 10.1172/JCI23769
9. Rivera-Feliciano J, Lee K-H, Kong SW, Rajagopal S, Ma Q, Springer Z, Izumo S, Tabin CJ, Pu WT. Development of heart valves requires Gata4 expression in endothelial-derived cells. *Development.* 2006;133:3607–3618. doi: 10.1242/dev.02519
10. He A, Kong SW, Ma Q, Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci USA.* 2011;108:5632–5637. doi: 10.1073/pnas.1016959108
11. He A, Gu F, Hu Y, Ma Q, Ye LY, Akiyama JA, Visel A, Pennacchio LA, Pu WT. Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat Commun.* 2014;5:4907. doi: 10.1038/ncomms5907
12. Akerberg BN, Gu F, VanDusen NJ, Zhang X, Dong R, Li K, Zhang B, Zhou B, Sethi I, Ma Q, et al. A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nat Commun.* 2019;10:4907. doi: 10.1038/s41467-019-12812-3
13. Zhou P, Gu F, Zhang L, Akerberg BN, Ma Q, Li K, He A, Lin Z, Stevens SM, Zhou B, et al. Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *Elife.* 2017;6: Available from: http://dx.doi.org/10.7554/eLife.22039
14. Beckett D, Kovaleva E, Schatz PJ. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* 1999;8:921–929. doi: 10.1110/ps.8.4.921
15. Jiao K, Kulessa H, Tompkins K, Zhou Y, Batts L, Baldwin HS, Hogan BLM. An essential role of Bmp4 in the atrioventricular septation of the mouse heart. *Genes Dev.* 2003;17:2362–2367. doi: 10.1101/gad.1124803
16. Kisanuki YY, Hammer RE, Miyazaki J-I, Williams SCC, Richardson JA, Yanagisawa M. Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis in vivo. *Dev Biol.* 2001;230:230–242.
17. Sörensen I, Adams RH, Gossler A. DLL1-mediated Notch activation regulates endothelial identity in mouse fetal arteries. *Blood.* 2009;113:5680–5688. doi: 10.1182/blood-2008-08-174508
18. Zhou B, Wu B, Tompkins KL, Boyer KL, Grindley JC, Baldwin HS. Characterization of Nfatc1 regulation identifies an enhancer required for gene expression that is specific to pro-valve endocardial cells in the developing heart. *Development.* 2005;132:1137–1146. doi: 10.1242/dev.01640
19. Schwartz RJ, Olson EN. Building the heart piece by piece: modularity of cis-elements regulating Nkx2-5 transcription. *Development.* 1999;126:4187–4192. doi: 10.1242/dev.126.19.4187
20. Kundaje A, Rozowsky J, Harmanci A, Wilder S, Gerstein M, Dunham I, Birney E, Batzoglou S, Others. ENCODE: TF ChIP-seq peak calling using the Irreproducibility Discovery Rate (IDR) framework. 2014;
21. Dai YS, Markham BE. p300 Functions as a coactivator of transcription factor gata-4. *J Biol Chem.* 2001;276:37178–37185. doi: 10.1074/jbc.M103731200
22. Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet.* 2010;42:806–810. doi: 10.1038/ng.650
23. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 2009;457:854–858. doi: 10.1038/nature07730

24. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, et al. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*. 2013;155:1521–1531. doi: 10.1016/j.cell.2013.11.033
25. Piper J, Elze MC, Cauchy P, Cockerill PN, Bonifer C, Ott S. Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res*. 2013;41:e201. doi: 10.1093/nar/gkt850
26. Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol*. 2019; 20:45. doi: 10.1186/s13059-019-1642-2
27. Ye M, Coldren C, Liang X, Mattina T, Goldmuntz E, Benson DW, Ivy D, Perryman MB, Garrett-Sinha LA, Grossfeld P. Deletion of ETS-1, a gene in the Jacobsen syndrome critical region, causes ventricular septal defects and abnormal ventricular morphology in mice. *Hum Mol Genet*. 2010;19:648–656. doi: 10.1093/hmg/ddp532
28. Intrinsic endocardial defects contribute to hypoplastic left heart syndrome. *Cell Stem Cell*. 2020;27:574–589.e8. doi: 10.1016/j.stem.2020.07.015
29. Wu B, Wang Y, Lui W, Langworthy M, Tompkins KL, Hatzopoulos AK, Baldwin HS, Zhou B. Nfatc1 coordinates valve endocardial cell lineage development required for heart valve formation. *Circ Res*. 2011;109:183–192. doi: 10.1161/CIRCRESAHA.111.245035
30. Molkentin JD, Lu JR, Antos CL, Markham B, Richardson J, Robbins J, Grant SR, Olson EN. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell*. 1998;93:215–228. doi: 10.1016/s0092-8674(00)81573-1
31. Robinson AS, Materna SC, Barnes RM, De Val S, Xu S-MM, Black BL. An arterial-specific enhancer of the human endothelin converting enzyme 1 (ECE1) gene is synergistically activated by Sox17, FoxC2, and Etv2. *Dev Biol*. 2014;395:379–389. doi: 10.1016/j.ydbio.2014.08.027
32. Heineke J, Auger-Messier M, Xu J, Oka T, Sargent MA, York A, Klevitsky R, Vaikunth S, Duncan SA, Aronow BJ, et al. Cardiomyocyte GATA4 functions as a stress-responsive regulator of angiogenesis in the murine heart. *J Clin Invest*. 2007;117:3198–3210. doi: 10.1172/JCI32573
33. Zhang H, Pu W, Li G, Huang X, He L, Tian X, Liu Q, Zhang L, Wu SM, Sucov HM, et al. Endocardium minimally contributes to coronary endothelium in the embryonic ventricular free walls. *Circ Res*. 2016;118:1880–1893. doi: 10.1161/CIRCRESAHA.116.308749
34. El-Rass S, Eisa-Beygi S, Khong E, Brand-Arzamendi K, Mauro A, Zhang H, Clark KJ, Ekker SC, Wen X-Y. Disruption of alters endocardial and myocardial fusion during zebrafish cardiac assembly. *Biol Open*. 2017;6:348–357. doi: 10.1242/bio.021212
35. Moore K, Fulmer D, Guo L, Koren N, Glover J, Moore R, Gensemer C, Beck T, Morningstar J, Stairley R, et al. PDGFRα: expression and function during mitral valve morphogenesis. *J Cardiovasc Dev Dis*. 2021;8: Available from: http://dx.doi.org/10.3390/jcdd8030028
36. Alfieri CM, Cheek J, Chakraborty S, Yutzey KE. Wnt signaling in heart valve development and osteogenic gene induction. *Dev Biol*. 2010;338:127–135. doi: 10.1016/j.ydbio.2009.11.030
37. Kim RY, Robertson EJ, Solloway MJ. Bmp6 and Bmp7 are required for cushion formation and septation in the developing mouse heart. *Dev Biol*. 2001;235:449–466. doi: 10.1006/dbio.2001.0284
38. Miquerol L, Gertsenstein M, Harpal K, Rossant J, Nagy A. Multiple developmental roles of VEGF suggested by a LacZ-tagged allele. *Dev Biol*. 1999;212:307–322. doi: 10.1006/dbio.1999.9355
39. Ferrara N, Carver-Moore K, Chen H, Dowd M, Lu L, O'Shea KS, Powell-Braxton L, Hillan KJ, Moore MW. Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature*. 1996;380:439–442. doi: 10.1038/380439a0
40. Carmeliet P, Ferreira V, Breier G, Pollefeyt S, Kieckens L, Gertsenstein M, Fahrig M, Vandenhoeck A, Harpal K, Eberhardt C, et al. Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature*. 1996;380:435–439. doi: 10.1038/380435a0
41. Favier R, Akshoomoff N, Mattson S, Grossfeld P. Jacobsen syndrome: advances in our knowledge of phenotype and genotype. *Am J Med Genet C Semin Med Genet*. 2015;169:239–250. doi: 10.1002/ajmg.c.31448
42. Tootleman E, Malamut B, Akshoomoff N, Mattson SN, Hoffman HM, Jones MC, Printz B, Shiryaev SA, Grossfeld P. Partial Jacobsen syndrome phenotype in a patient with a de novo frameshift mutation in the ETS1 transcription factor. *Cold Spring Harb Mol Case Stud*. 2019;5:a004010. doi: 10.1101/mcs.a004010. Available from: http://dx.doi.org/10.1101/mcs.a004010
43. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28:495–501. doi: 10.1038/nbt.1630
44. Watt AJ, Battle MA, Li J, Duncan SA. GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proc Natl Acad Sci USA*. 2004;101:12573–12578. doi: 10.1073/pnas.0400752101
45. Zhou P, Zhang Y, Ma Q, Gu F, Day DS, He A, Zhou B, Li J, Stevens SM, Romo D, et al. Interrogating translational efficiency and lineage-specific transcriptomes using ribosome affinity purification. *Proc Natl Acad Sci USA*. 2013;110:15395–15400. doi: 10.1073/pnas.1304124110
46. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L. A global double-fluorescent Cre reporter mouse. *Genesis*. 2007;45:593–605. doi: 10.1002/dvg.20335
47. Cardiff RD, Miller CH, Munn RJ. Manual hematoxylin and eosin staining of mouse tissue sections. cold spring harbor protocols. 2014;2014:db.prot073411. Available from: http://dx.doi.org/10.1101/pdb.prot073411
48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. doi: 10.1186/s13059-014-0550-8
49. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–589. doi: 10.1016/j.molcel.2010.05.004
50. Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*. 2015;527:384–388. doi: 10.1038/nature15518
51. Mahony S, Auron PE, Benos PV. DNA familial binding profiles made easy: comparison of various motif alignment and clustering strategies. *PLoS Comput Biol*. 2007;3:e61. doi: 10.1371/journal.pcbi.0030061
52. Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res*. 2012;40:e128. doi: 10.1093/nar/gks433
53. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14:959–962. doi: 10.1038/nmeth.4396
54. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–359. doi: 10.1038/nmeth.1923
55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome project data processing subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079. doi: 10.1093/bioinformatics/btp352
56. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137. doi: 10.1186/gb-2008-9-9-r137
57. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21. doi: 10.1093/bioinformatics/bts635
58. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–930. doi: 10.1093/bioinformatics/btt656
59. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–15550. doi: 10.1073/pnas.0506580102
60. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8:14049. doi: 10.1038/ncomms14049
61. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. 2019;8:329–337.e4. doi: 10.1016/j.cels.2019.03.003
62. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–3587.e29. doi: 10.1016/j.cell.2021.04.048
63. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–26. doi: 10.1038/nbt.1754
64. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44:W160–W165. doi: 10.1093/nar/gkw257