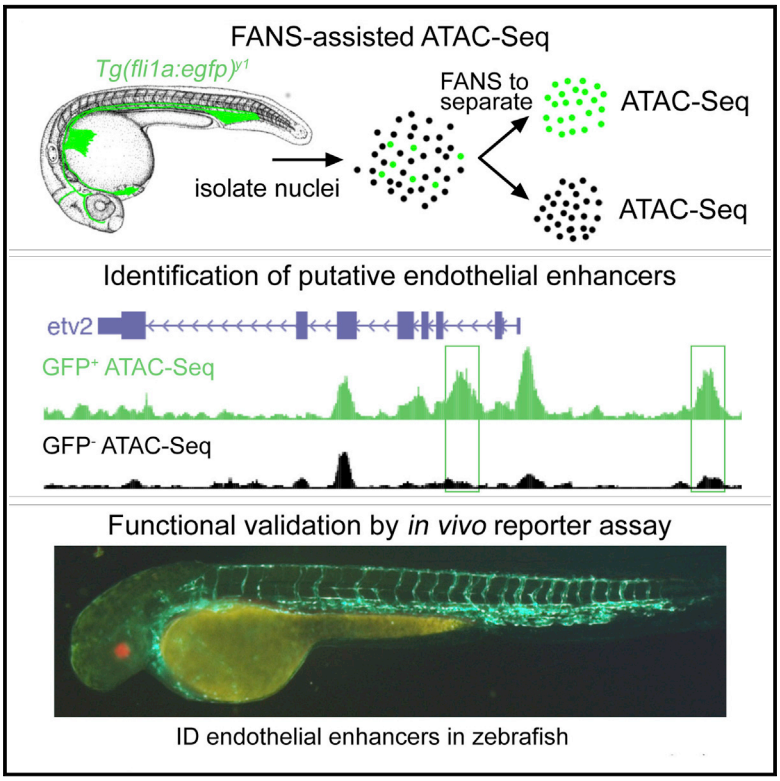
Resource

Cell Reports

Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq

Graphical Abstract



Authors

Aurelie Quillien, Mary Abdalla, Jun Yu, Jianhong Ou, Lihua Julie Zhu, Nathan D. Lawson

Correspondence

nathan.lawson@umassmed.edu

In Brief

Quillien et al. apply ATAC-seq to nuclei isolated from transgenic zebrafish embryos to successfully identify a compendium of active endothelial-specific enhancer elements.

Highlights

- Application of ATAC-seq to nuclei sorted from transgenic zebrafish embryos
- FANS-assisted ATAC-seq permits robust identification of putative enhancers
- FANS-assisted ATAC-seq applied to identify endothelial-specific enhancers
- These datasets and reagents are a resource to investigate vascular development

Accession Numbers

GSE97257

Quillien et al., 2017, Cell Reports *20*, 709–720
July 18, 2017
http://dx.doi.org/10.1016/j.celrep.2017.06.070

Cell Reports
Resource

Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq

Aurelie Quillien,[1,3,4] Mary Abdalla,[1,3] Jun Yu,[1] Jianhong Ou,[1] Lihua Julie Zhu,[1,2] and Nathan D. Lawson[1,5,*]
[1]Department of Molecular, Cell, and Cancer Biology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA
[2]Program in Bioinformatics and Integrative Biology, Department of Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA
[3]These authors contributed equally
[4]Present address: Centre de Biologie du Développement (CBD, UMR5547), Université de Toulouse, 31013 Toulouse, France
[5]Lead Contact
*Correspondence: nathan.lawson@umassmed.edu
http://dx.doi.org/10.1016/j.celrep.2017.06.070

SUMMARY

Identification of tissue-specific and developmentally active enhancers provides insights into mechanisms that control gene expression during embryogenesis. However, robust detection of these regulatory elements remains challenging, especially in vertebrate genomes. Here, we apply fluorescent-activated nuclei sorting (FANS) followed by Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) to identify developmentally active endothelial enhancers in the zebrafish genome. ATAC-seq of nuclei from *Tg(fli1a:egfp)*^{y1} transgenic embryos revealed expected patterns of nucleosomal positioning at transcriptional start sites throughout the genome and association with active histone modifications. Comparison of ATAC-seq from GFP-positive and -negative nuclei identified more than 5,000 open elements specific to endothelial cells. These elements flanked genes functionally important for vascular development and that displayed endothelial-specific gene expression. Importantly, a majority of tested elements drove endothelial gene expression in zebrafish embryos. Thus, FANS-assisted ATAC-seq using transgenic zebrafish embryos provides a robust approach for genome-wide identification of active tissue-specific enhancer elements.

INTRODUCTION

The zebrafish is an excellent model for dissecting transcriptional regulatory programs that govern development (Ferg et al., 2014). Much of this work derived from the well-known benefits of the zebrafish, including its rapid external development, transparent embryos, and its utility for genetic screens. Indeed, forward screening efforts have revealed previously unknown transcription factors required for the development of a number of different cell lineages (for examples, see Dickmeis et al., 2001; Kikuchi et al., 2001; Pham et al., 2007; Reischauer et al., 2016). However, a broader characterization of developmentally active *cis*-regulatory elements has lagged behind these genetic studies. Detailed analysis of transcriptional networks in human cells and model systems has benefitted from collaborative large-scale genomic efforts to define regulatory elements (Gerstein et al., 2010, 2012; modENCODE Consortium et al., 2010). However, the zebrafish was not included in this work. Thus, there is a need for robust approaches to detect developmentally active cell-type-specific *cis*-regulatory elements in the zebrafish genome.

Most efforts to identify regulatory elements in the zebrafish genome have focused on cell-type-specific enhancers. Early studies took advantage of the ability to easily detect transient mosaic transgene expression in zebrafish embryos following DNA injection, which allowed rapid in vivo reporter assays. Prior to the sequencing of the zebrafish genome, these efforts relied on traditional deletion approaches using moderate-sized fragments flanking a promoter of interest (Meng et al., 1997; Müller et al., 1999). Advances in the manipulation of bacterial artificial chromosomes (BACs) containing large genomic inserts and their use in zebrafish transgenesis permitted much larger fragments to be functionally assayed (Jessen et al., 1998). However, identifying the location of functional elements in flanking genomic sequence remained challenging. Subsequent availability of fish genome sequences, along with phylogenetic footprinting (Wasserman et al., 2000), enabled identification of small conserved non-coding elements (CNEs), which often possessed enhancer activity (Göttgens et al., 2002; Komisarczuk et al., 2009). However, unlike *Drosophila* and many mammalian species, for which closely related genomes are available for accurate comparison, the zebrafish is considered to be phylogenetically isolated regarding available genome sequences (Hiller et al., 2013). As a consequence, annotation of CNEs in the zebrafish genome is less comprehensive compared to other model systems. Thus, while conservation with distantly related fish genomes has aided in the identification of regulatory elements in zebrafish, these approaches likely only reveal a small subset of developmentally regulated enhancers.

Recent approaches have leveraged chromatin immunoprecipitation sequencing (ChIP-seq) to identify regions in the zebrafish genome associated with histone modifications that mark active promoters or enhancers (Aday et al., 2011; Bogdanovic et al., 2012). However, since this work was limited to chromatin isolated from whole embryos, it is not possible to predict where an enhancer may drive expression based simply on an active chromatin mark. Moreover, it can be difficult to identify enhancers that are active in small cell populations in the context of the whole embryo and would therefore only contribute negligible signal in a ChIP-seq analysis. Unfortunately, ChIP-seq requires large numbers of cells, making it difficult to apply on small populations of specific cell types isolated from zebrafish embryos. While methods have emerged that enable ChIP-seq on small numbers of cells (Adli and Bernstein, 2011), these techniques can be challenging and have not been implemented in the zebrafish.

Among recent advances in profiling open regions of chromatin is the Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq; Buenrostro et al., 2013). This entails treatment of unfixed nuclei with a hyperactive form of the yeast Tn5 transposase that is loaded with sequencing adapters. During incubation, Tn5 inserts the adapters into accessible areas of the genome, which are predominantly devoid of histones and largely represent active regulatory elements. Deep sequencing of genomic DNA isolated from Tn5-treated nuclei allows straightforward identification of open regions in a genome of interest. While similar to DNase-seq in identifying open regulatory regions, ATAC-seq is less technically demanding. Importantly for developmental studies, ATAC-seq can be applied to small numbers of cells (Buenrostro et al., 2013; Wu et al., 2016). For these reasons, we sought to apply ATAC-seq in an effort to identify developmentally active enhancers in the zebrafish genome. To aid identification of cell-type-specific enhancers, we incorporated fluorescence-activated nuclei sorting (FANS) from a zebrafish transgenic line in which GFP (Egfp) labels endothelial cells. Subsequent comparison of ATAC-seq reads from GFP-positive and GFP-negative nuclei provided a robust and technically straightforward means to identify a compendium of putative lineage-specific regulatory elements in the zebrafish. Moreover, the resulting datasets and plasmid collection from these efforts provide a valuable resource for investigators interested in transcriptional control of vascular development and endothelial differentiation.

RESULTS AND DISCUSSION

To identify cell-type-specific enhancers throughout the zebrafish genome, we applied ATAC-seq to nuclei isolated from transgenic zebrafish embryos. Given our interests in vascular development, we chose to focus on endothelial cells. For this purpose, we relied on the *Tg(fli1a:egfp)*^{y1} line in which GFP is expressed predominantly in endothelial cells at 24 hr post-fertilization (hpf; Lawson and Weinstein, 2002). We isolated nuclei from *Tg(fli1a:egfp)*^{y1} embryos rather than cells for two major reasons. First, conditions used for embryo dissociation and fluorescence-activated cell sorting (FACS) can lead to a significant loss of cells, possibly due to increased cell death (data not shown). In these cases, apoptosis would likely increase the accessibility of the Tn5 to fragmented DNA in nuclei, resulting in high background signal and uninterpretable results. Second, existing ATAC-seq protocols, which utilize a crude cell lysate, can yield a high proportion of mitochondrial reads (Wu et al., 2016). We reasoned that isolation and sorting of nuclei would reduce mitochondrial contamination. For these reasons, we applied FANS to isolate endothelial nuclei from *Tg(fli1a:egfp)*^{y1} embryos (Figure 1A). Despite the lack of a nuclear localization signal on the GFP, nuclei from *Tg(fli1a:egfp)*^{y1} embryos retained sufficient fluorescence to discern them from non-endothelial cells by microscopy and by FACS (Figures S1A–S1D). By gating on high fluorescence and low forward scatter, we obtained enriched populations of GFP-positive nuclei while also collecting GFP-negative nuclei from the same embryos for comparison (Figures S1D–S1F). Following FANS, we performed ATAC sequencing (ATAC-seq) on 20,000 GFP-positive and GFP-negative nuclei. This was repeated two additional times to give biological triplicates.

We first assessed the proportion of reads that mapped to the zebrafish mitochondrial genome, given reports of mitochondrial reads being a significant source of contamination. We consistently observed that mitochondrial reads typically made up less than 5% of all mapped reads from our libraries (Table S1). Thus, more stringent isolation of nuclei can prevent contamination of ATAC-seq libraries from insertions into mitochondrial DNA. To further assess the quality of our ATAC-seq data, we performed several genome-wide analyses. Tn5 can insert into open regions of chromatin as well as linker DNA between nucleosomes at active elements (Buenrostro et al., 2013). Therefore, by separately analyzing the pattern of mapped ATAC-seq fragments for small (<100 bp) or large (180–247 bp) DNA fragments, it is possible to observe nucleosome-free and nucleosome-bound regions, respectively, associated with regulatory elements. Indeed, fragments less than 100 bp in length clustered immediately upstream of transcriptional start sites (TSSs) throughout the zebrafish genome in both GFP-positive and GFP-negative nuclei (Figures 1B–1D). This pattern is consistent with nucleosome depletion observed in regions upstream of active TSSs in eukaryotic genomes (Radman-Livaja and Rando, 2010), as has been detected previously by ATAC-seq (Buenrostro et al., 2013). Similar patterns were observed for the other biological replicates (Figure S2A). By contrast, 180- to 247-bp fragments, corresponding to nucleosome footprints, were depleted from TSSs throughout the genome (Figure 1B) but displayed periodic peaks of mapping immediately downstream of the TSS (Figures 1C and 1D). This pattern is consistent with the strong positioning of the first several nucleosomes downstream of the TSS, starting at position +1 in eukaryotic cells (Radman-Livaja and Rando, 2010). Thus, we observe expected patterns of nucleosome occupancy and depletion at the TSSs across the genome in ATAC-seq data from isolated zebrafish nuclei.

Regions of open chromatin identified by ATAC-seq include promoters and enhancers (Buenrostro et al., 2013). Therefore, ATAC-seq reads should map to known histone modifications associated with these features. To determine whether this was the case in our samples, we compared ATAC-seq data from GFP-positive and GFP-negative nuclei to histone modifications previously demonstrated to mark active regulatory elements in

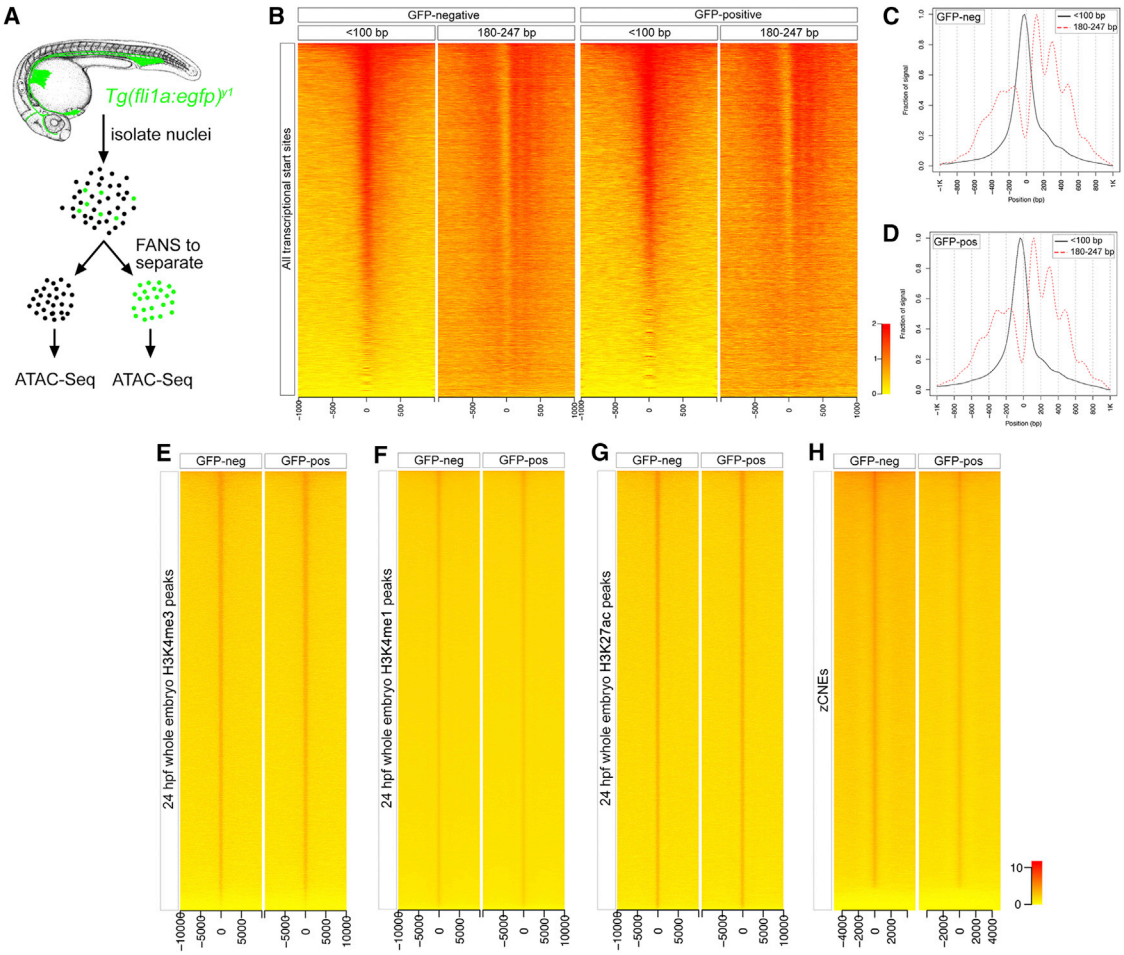


Figure 1. ATAC-Seq Data from Isolated Zebrafish Nuclei

(A) Schematic for isolation of nuclei from *Tg(fli1a:egfp)*^{y1} zebrafish embryos.

(B) Heatmaps showing density of mapped ATAC-seq reads from single biological replicates 1 kb up and downstream of ENSEMBL-annotated transcriptional start sites (TSS) in Zv9. Separate heatmaps for fragments less than 100 bp and between 180 and 247 bp are shown.

(C and D) Density plots flanking Zv9 TSS from (C) GFP-negative and (D) GFP-positive nuclei for mapped reads shown in (B).

(E–H) Heatmaps showing localization of ATAC-seq reads with (E) H3K4me3, (F) H3K4me1, (G) H3K27ac, or (H) zCNE. Intervals flanking indicated feature are shown in kilobases.

whole zebrafish embryos at 24 hpf (Bogdanovic et al., 2012). These modifications included histone 3, trimethylated at lysine 4 (H3K4me3), a hallmark of active promoters (Bernstein et al., 2005), as well as histone 3, acetylated at lysine 27 (H3K27ac), and monomethylated H3K4 (H3K4me1), which are associated with both promoter and enhancer elements (Heintzman et al., 2009). Heatmaps of ATAC-seq reads from GFP-positive and GFP-negative populations show strong enrichment at loci marked by each of these histone modifications (Figures 1E–1G). In addition, we observed mapping of ATAC-seq reads at a collection of annotated CNEs in the zebrafish genome (zCNEs; Figure 1H). These zCNEs were identified through local alignments of multiple fish and tetrapod genomes to detect remote homology between zebrafish and at least two other species (Hiller et al., 2013). In all cases, concordance between ATAC-seq reads and histone marks or zCNEs were obtained for the replicate libraries (Figures S1B–S1E). Thus, regions identified by FANS-assisted ATAC-seq exhibited hallmarks of active regulatory elements. These observations indicate that it is feasible to obtain high-quality ATAC-seq data from nuclei isolated from transgenic zebrafish embryos following fluorescence-based sorting.

Our goal for these studies was to apply ATAC-seq to identify developmentally active endothelial-specific enhancers. Therefore, we next examined ATAC-seq read density at loci where active endothelial enhancers have been characterized. For example, at the locus encoding the endothelial-specific transcription factor, Etv2, we observed much higher read density at the TSS and at an element approximately 2 kb upstream when comparing GFP-positive and GFP-negative nuclei (Figure 2A). This upstream element, which is conserved in multiple fish species (Figure 2A), is capable of driving endothelial-specific expression in transgenic zebrafish embryos (Veldman and Lin, 2012). Similarly, a conserved endothelial-specific enhancer

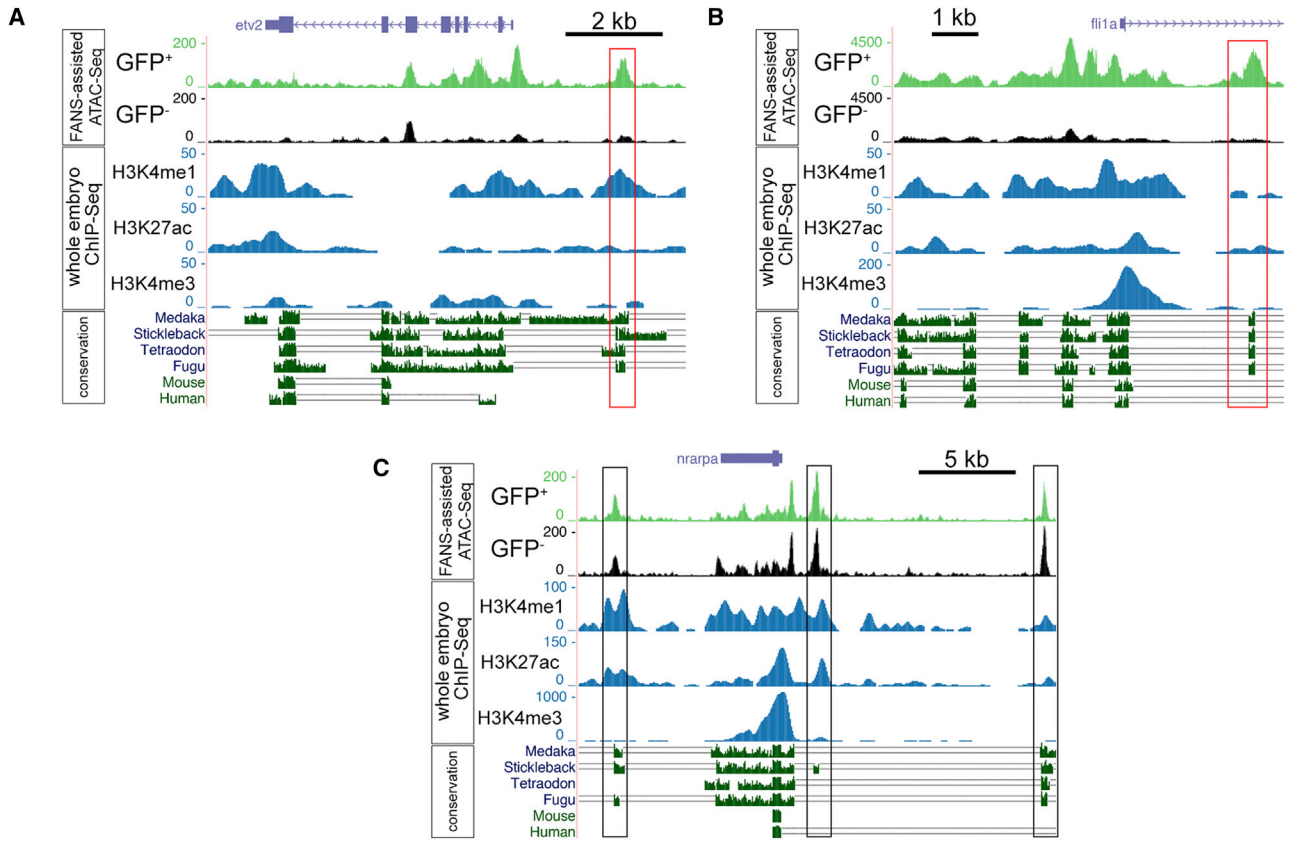


Figure 2. Mapping of ATAC-Seq Reads at Known Endothelial Enhancers

(A–C) Visualization of ATAC-seq from a single biological replicate, whole-embryo ChIP-seq mapped reads for H3K4me1, H3K4me3, and H3K27ac, and conservation from the UCSC Genome Browser.

(A) *etv2* locus. Red box is known active endothelial enhancer.

(B) *fli1a* locus. Red box indicates known active endothelial enhancer.

(C) *nrarpa* locus. Black boxes are open regions in both GFP-positive and -negative nuclei.

that we have previously identified at the *fli1a* locus (Villefranc et al., 2007) also exhibited specific enrichment of ATAC-seq reads in GFP-positive, but not GFP-negative nuclei (Figure 2B; note that the ATAC-seq read counts at the *fli1a* locus are elevated compared to other loci due to the high copy number of the *fli1a:egfp* transgene, which encompasses this sequence; Lawson and Weinstein, 2002). Although H3K4me1 could be detected at the *etv2* enhancer, we failed to observe consistent occupancy of H3K27ac or H3K4me1 at these elements using publicly available ChIP-seq data from whole embryos at 24 hpf (Bogdanovic et al., 2012; Figures 2A and 2B). This is likely because endothelial cells comprise less than 5% of all cells in embryos at 24 hpf, and therefore endothelial-specific enhancers may not be detectable in the context of whole-embryo ChIP-seq. In contrast to *etv2* and *fli1a*, we observed relatively equal mapped read density at conserved elements flanking the *nrarpa* locus (Figure 2C), which is expressed more broadly in both neural and endothelial cells in zebrafish embryos (Phng et al., 2009). Moreover, regions of open chromatin flanking *nrarpa* identified by ATAC-seq also showed enrichment of H3K27ac and H3K4me1 (Figure 2C). The enrichment of ATAC-seq in both GFP-positive and GFP-negative nuclei, as well as the robust H3K27ac and H3K4me1 signal in whole-embryo ChIP-seq studies, are likely explained by the broader expression pattern of this gene compared to *etv2* and *fli1a*. These observations suggested that ATAC-seq analysis of sorted nuclei from *Tg(fli1a:egfp)*^{y1} embryos could reveal known endothelial enhancers that would otherwise not be easily identified by whole-embryo ChIP-seq for active histone marks.

To more broadly identify putative endothelial-specific enhancers in the zebrafish genome, we compared local mapped read densities from ATAC-seq libraries in triplicate samples from GFP-positive and GFP-negative nuclei isolated from *Tg(fli1a:egfp)*^{y1} embryos. This analysis revealed 5,291 elements that exhibited a significantly higher density of mapped reads in GFP-positive nuclei when compared with GFP-negative nuclei ($\log_2$ fold change $\geq$ 1, false discovery rate [FDR] $\leq$0.05; Figure 3A; Table S2; Data S1). Conversely, 4,531 regions displayed higher read density in the GFP-negative population (log2 fold change ≤ -1, FDR $\leq$ 0.05; Table S2, Data S1), while 48,593 regions were similar in both populations (hereafter referred to as "common" elements; see Experimental Procedures, Table S2,

and Data S1). The differential enrichment of ATAC-seq reads in GFP-positive and GFP-negative nuclei suggested that these elements were selectively open in each of these populations and therefore represented putative cell-type-specific regulatory elements. Consistent with this possibility, regions of ATAC-seq enrichment in both GFP-positive and GFP-negative populations correlated with H3K4me1 and H3K27ac localization, although association of H3K4me3 was much lower (Figures 3B and 3C). By contrast, common ATAC-seq elements were equally associated with all three histone marks (Figure 3D). Accordingly, concordance analysis revealed that approximately one-third of common elements associated with TSSs, while less than 10% of ATAC-seq elements specific to GFP-positive or GFP-negative nuclei did so (Figure 3E). Similar results were seen with concordance of H3K4me3 (Figure 3E), which preferentially localizes to promoters (Bernstein et al., 2005). These observations suggest that most GFP-positive and GFP-negative ATAC-seq elements represented enhancers rather than promoters.

At the *etv2* and *fli1a* loci, known endothelial enhancers displayed selective mapping of ATAC-seq reads in GFP-positive compared to GFP-negative nuclei but were not consistently associated with active histone marks (e.g., H3K27ac) identified in whole-embryo ChIP-seq (Figures 2A and 2B). As noted above, this is likely due to the fact that GFP-positive cells comprise a small proportion of the cell population in *Tg (fli1a:egfp)*^{y1} embryos (Covassin et al., 2006). However, many GFP-negative ATAC-seq elements are likely those that are active in a large proportion of all remaining non-endothelial cells and would be more reliably detected in whole-embryo ChIP-seq for active chromatin marks. To expand this observation across the genome, we determined concordance of H3K27ac, as assessed by ChIP-seq in whole embryos, with each of the three classes of ATAC-seq elements identified above. For both common ATAC-seq elements and those from GFP-negative nuclei, we observed that more than one-third of these regions also displayed H3K27ac occupancy based on whole-embryo ChIP-seq (Figure 3F; Table S3). By contrast, only 16% of GFP-positive elements were found to also bind H3K27ac (Figure 3F; Table S3). We would note that this proportion of co-localized elements is likely sufficient to reveal mapping based on density plots shown above (Figure 3B), despite the low level of concordance. Since conservation is often used to identify putative tissue-specific enhancers in zebrafish, we also assessed the concordance of zCNEs with ATAC-seq elements. As with whole H3K27ac ChIP-seq concordance, less than 15% of GFP-positive ATAC-seq elements harbored a zCNE, similar to common elements (Figure 3F; Table S3). Interestingly, a much larger proportion of GFP-negative elements possessed zCNEs (Figure 3F), which may be due to the prevalence of putative enhancers flanking essential neural genes in this set (see below).

In general, promoter and enhancer activity correlates with levels of gene expression. Therefore, we would expect that the specific accessibility of active promoter and enhancer elements in a particular cell type would reflect the level of expression for respective transcripts in the same cell type. To determine whether this was the case, we first assessed the expression levels of transcripts where the TSS was specifically open in GFP-positive or GFP-negative nuclei. For genes where the TSS was specifically open in GFP-positive nuclei, we observed significantly higher levels of expression for respective transcripts, as assessed by RNA sequencing (RNA-seq) when comparing GFP-positive versus -negative cells (Figure 3G). The converse was true for TSSs open specifically in GFP-negative nuclei (Figure 3H). To determine whether non-TSS ATAC-seq elements were similarly correlated with expression, we compared the number of cell-type-specific elements flanking a TSS (±250 kb) with the degree of enrichment of the respective transcript in GFP-positive or GFP-negative cells. For genes that exhibited significant enrichment in GFP-positive cells compared with GFP-negative ($\log_2$ fold change ≥ 1, FDR ≤ 0.05), we observed a significant correlation with the number of GFP-positive elements flanking their TSS (Figure 3I). A similar degree of correlation between the number of GFP-negative elements and enrichment of expression in GFP-negative cells was also noted (Figure 3J). By contrast, we did not observe any significant correlation between the number of common elements and enrichment in GFP-positive or GFP-negative cells (Figure 3K). Together, these observations are consistent with a role for these elements in controlling cell-type-specific gene expression.

To further investigate the relationship between cell-type-specific elements and adjacent genes, we applied the Genomic Regions Enrichment of Annotations Tool (GREAT), which identifies enriched functional gene features associated with an input set of *cis*-regulatory elements (McLean et al., 2010). GREAT analysis of GFP-positive elements revealed an over-representation of nearby genes with Gene Ontology terms associated with artery development and vasculogenesis, as well as endothelial-specific expression patterns (top three hits shown in Table 1; all results are in Table S4). Accordingly, GFP-positive elements were enriched for binding sites of transcription factors with known roles in endothelial gene expression, including ETS, Forkhead, and GATA proteins (De Val and Black, 2009; Table S5). By contrast, GFP-negative elements were found to flank genes largely associated with neural development and that exhibited neural expression patterns (Table 1; Table S6). This is likely due to the high proportion of neurons in GFP-negative cells compared to other cell types at 24 hpf. Taken together, these observations strongly suggest that a high proportion of GFP-positive elements promote endothelial gene expression.

To more definitively assess whether GFP-positive elements identified by ATAC-seq are active endothelial enhancers, we performed transient reporter assays in live zebrafish embryos. Given our interest in understanding mechanisms of endothelial gene expression, we chose 12 GFP-positive elements flanking 11 different endothelial-expressed genes (Figure 4A; these are a subset of the 5291 GFP-positive elements identified above, Table S2). Although these chromatin regions were preferentially open in GFP-positive cells, they did not consistently display H3K27ac binding in whole embryos or bear zCNEs (Table 2). Thus, several of these elements would not otherwise be identifiable based on previous approaches to detect enhancers. To assay the activity of these elements, we used a construct in which an enhancer of interest was placed upstream of a basal promoter to drive expression of enhanced green fluorescent protein (Egfp; Figure 4B). This construct also bears the *crystallin, alpha a* (*cryaa*) promoter upstream of the red fluorescent *mcherry*

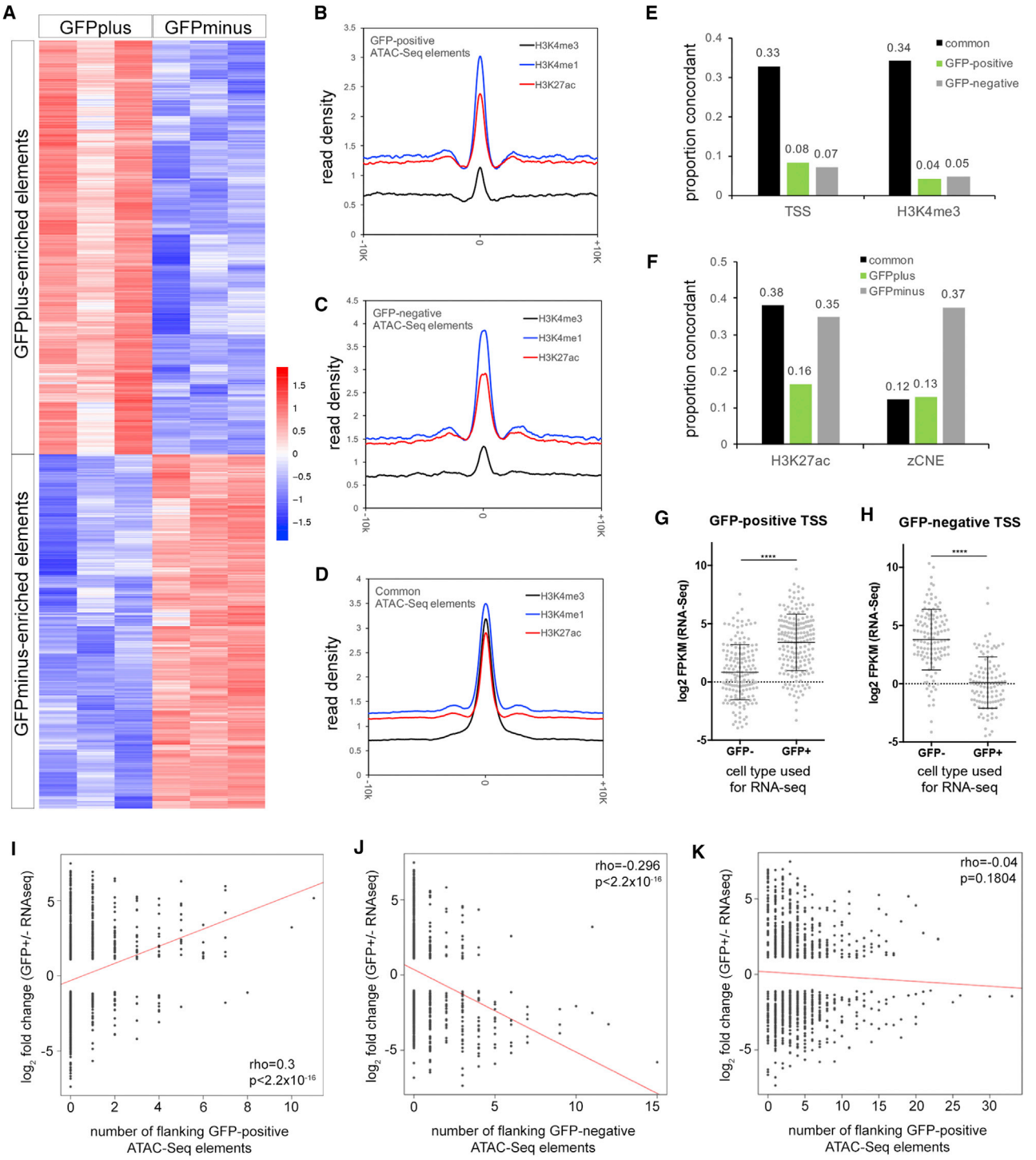


Figure 3. Identification and Characterization of Cell-Type-Specific Open Chromatin

(A) Heatmap of $\log_2$ fold change in GFP$^+$/GFP$^-$ ATAC-seq read density (FDR $\leq$ 0.05; biological triplicates).

(B–D) Density plots of mapped reads from H3K4me1, H3K4me3, and H3K27ac whole-embryo ChIP-seq against ATAC-seq elements specific to (B) GFP-positive, (C) GFP-negative, or (D) common elements.

(E) Proportion of elements of indicated class that associate with TSS or H3K4me3.

(F) Proportion of elements concordant with H3K27ac or zCNEs.

(legend continued on next page)

Table 1. GREAT Analysis of GFP-Positive and GFP-Negative ATAC-Seq Elements

Ontology	Term Name	Hyper FDR Q-Val	Hyper Fold Enrichment
GFP-Positive Elements			
GO biological process	artery development	1.19e-13	2.4939
	vasculogenesis	2.1e-12	2.2486
	VEGF receptor signaling pathway	4.7e-9	2.5761
Zebrafish wild-type expression	Pharyngula:Prim-5 24–30 hr; vasculature	3.5e-92	2.4591
	Pharyngula:Prim-5 24–30 hr; blood vessel	2.5e-92	2.5378
	Pharyngula:Prim-5 24–30 hr; artery	3.2e-85	2.7309
GFP-Negative Elements			
GO biological process	CNS development	5.4e-98	2.2269
	brain development	1.1e-78	2.2508
	hindbrain development	7.9e-49	2.6317
Zebrafish wild-type expression	Pharyngula:Prim-5 24–30 hr; spinal cord	4.5e-211	2.1568
	Pharyngula:Prim-5 24–30 hr; hindbrain	7.8e-209	2.0491
	Segmentation: 14–19 somites 16–19 hr; hindbrain	4.0e-181	2.4839

protein, which drives lens expression and provides a constitutive marker to control for injection (Figures 4B and 4C; Berger and Currie, 2013). The transgene cassette is flanked by Tol2 direct repeats to enable high-efficiency integration into the zebrafish genome (Kawakami et al., 2004), thereby facilitating robust transient reporter activity. Reporter assays with putative enhancers upstream of the basal promoter:Egfp transgene revealed that most drove endothelial expression at 55 hpf (Table 2), while the basal promoter alone did not (Figure 4C). For example, an element downstream of the *mafbb* gene drove Egfp expression in caudal vein endothelial cells (Figures 4D and 4E). Similarly, elements upstream of *nrp1b* and *tmem88a* could drive endothelial expression throughout the trunk blood vessels (Figures 4F–4H). However, the proportion of injected embryos with endothelial expression was low, despite robust expression of the *cryaa:mcherry* lens marker indicating successful injection of the reporter (Figures 4E, 4G, 4I and S3A). This low penetrance is likely because endothelial expression was much weaker than we have previously noted for other endothelial-specific enhancer elements (e.g., the *fli1a* enhancer; Villefranc et al., 2007). Nonetheless, eight out of twelve of the candidate elements were able to drive endothelial expression from the basal promoter:Egfp transgene in zebrafish embryos (Table 2; data not shown).

Enhancer-driven expression can be improved by including the likely cognate promoter elements from the flanking gene instead of a basal promoter (Gehrig et al., 2009). Therefore, we assessed the effect of replacing the basal promoter with the putative cognate promoter sequence for each enhancer. In six of 11 cases, the promoter alone drove low endothelial Egfp expression in a small proportion of *cryaa:mcherry*-positive embryos (Table 2; Figure S3A). For example, weak endothelial expression could be seen in only five out of 30 *cryaa:mcherry*-positive embryos bearing a short *lmo2* promoter sequence upstream of Egfp (Figures 5A and 5B). For the remaining promoters, we observed either ectopic or no expression (Table 2 and see following). Strikingly, inclusion of endothelial enhancers with their putative cognate promoter in the same construct increased the proportion of *cryaa:mcherry*-positive embryos that exhibit endothelial Egfp expression in many cases (Figure S3A). At the same time, endothelial expression driven by these chimeric elements appeared qualitatively stronger. For example, as noted above, the *lmo2* promoter alone drove endothelial expression only in a small proportion of embryos and at relatively low levels, similar to a putative *lmo2* enhancer (Figures 5B and 5C). By contrast, a chimeric *lmo2* enhancer/promoter element drove endothelial expression in all *cryaa:mcherry*-positive embryos, and this expression appeared much stronger than with either element alone (Figure 5D). Similar results were observed with enhancers flanking *clec14a*, *dll4*, *dusp5*, *nrp1b*, and *tmem88a* (Figures 5E–5L and S3). For six out of 12 of the enhancers, we observed that placement upstream of their cognate promoter increased the proportion of *cryaa:mcherry*-positive embryos that displayed endothelial expression, and, in these cases, expression levels appeared more robust (Table 2; Figure S3A). In one case where the enhancer failed to drive endothelial expression with a basal promoter (*she*), coupling the enhancer with its cognate promoter led to endothelial expression of the reporter, albeit at low levels and in a small proportion of embryos (Table 2; Figure S3A; data not shown). For three flanking elements (*cldn5b*, *snx5*, and *yrk*), we failed to detect an increase in endothelial enhancer activity when placed upstream of its cognate promoter. Thus, nine out of 12 of the ATAC-seq elements that were selectively open in GFP-positive cells from *Tg(fli1a:egfp)*^{y1} embryos were capable of driving endothelial reporter expression. Importantly, several of these active elements did not exhibit concordance with a zCNE or H3K27ac binding in whole embryos (Table 2). Thus,

(G and H) Expression level (by RNA-seq) in $\log_2$ fragments per kilobase per million reads (FPKM) in GFP-positive or GFP-negative cells from *Tg(fli1a:egfp)*^{y1} for genes with TSS open specifically in (G) GFP-positive or (H) GFP-negative nuclei; ****$p < 0.0001$. Error bars represent ± SEM.
(I–K) Plots comparing $\log_2$ fold change in expression (by RNA-seq), between GFP-positive and -negative cells from *Tg(fli1a:egfp)*^{y1} embryos to number of flanking (I) GFP-positive, (J) GFP-negative, or (K) common elements. Only genes with GFP^+/GFP^- $\log_2$ fold change ≥ 1, or ≤ -1 (FDR ≤ 0.05) were included.

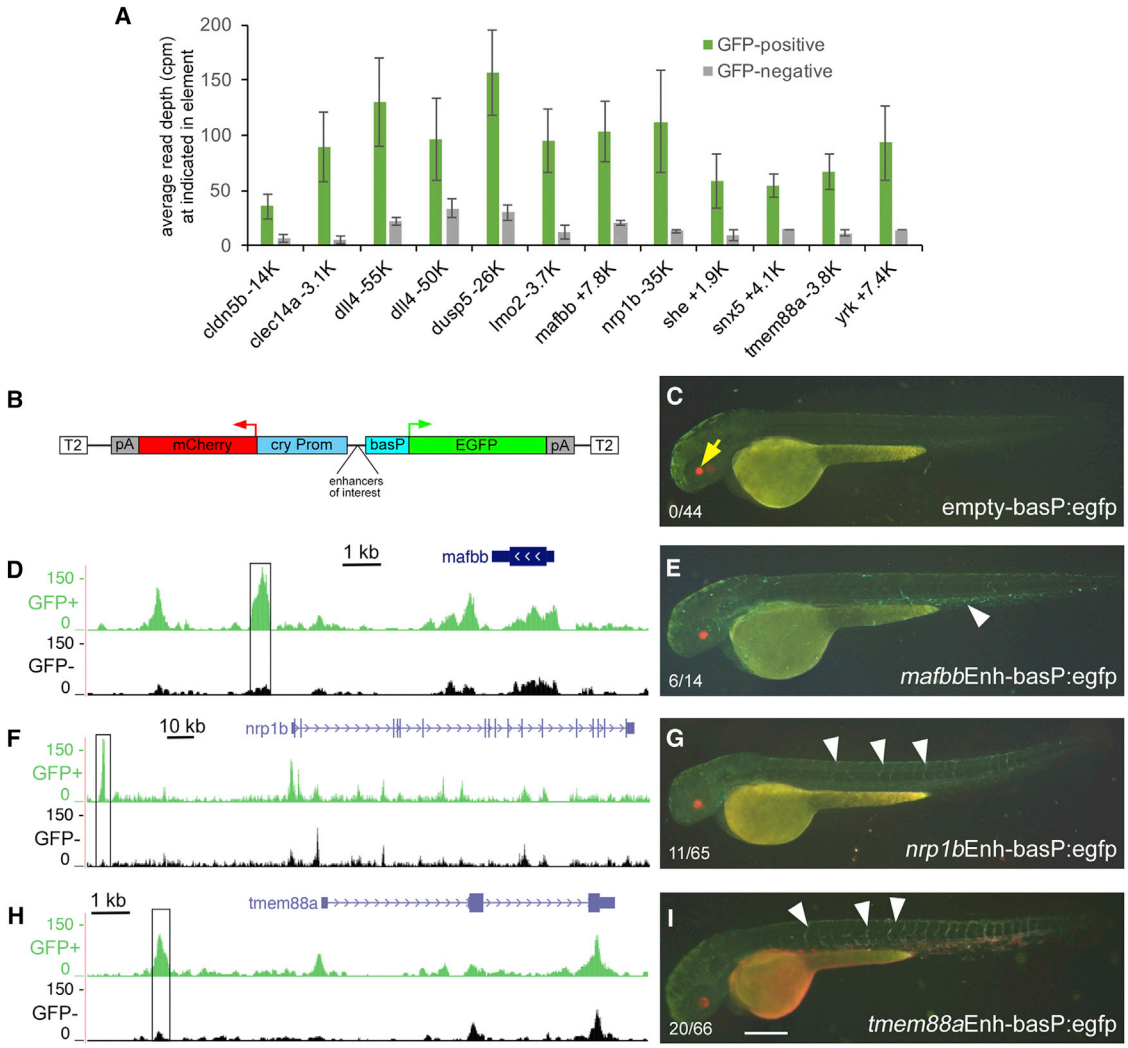


Figure 4. GFP-Positive Elements Can Drive Endothelial Gene Expression

(A) Average read depth, in counts per million (cpm), at each element used for reporter assays. Data are represented as mean ± SD from triplicate ATAC-seq libraries. All elements display GFP^+/GFP^- $\log_2$ fold change >1 ($p < 0.0001$; FDR < 0.05; biological triplicates). Name of adjacent endothelial gene shown on x axis with distance (in kilobases) and direction ("+" = downstream; "–" = upstream) of enhancer relative to TSS.

(B) Tol2 plasmid backbone used for reporter assays.

(C, E, G, and I) Overlays of green and red fluorescent images from embryos injected with reporter constructs. Lateral views: dorsal is up, and anterior is to the left. Ratios in left bottom denote number of embryos with GFP expression over number of *cryaa:mcherry*-expressing embryos from replicate injections. Embryos injected with (C) reporter with only basal promoter driving EGFP, or reporter with elements (E) downstream of *mafbb*, (G) upstream of *nrp1b* or (I) *tmem88a*. (D, F, and H) Mapped reads flanking indicated genes from GFP-positive and -negative nuclei isolated from *Tg(fli1a:egfp)*^{y1} embryos.

(E, G, and I) White arrowheads denote low-level expression in trunk endothelial cells. Black boxes are elements used in reporter assays. Scale bar, 250 μm.

these endothelial-specific enhancers would not otherwise have been detected using available methodology.

Detecting developmentally active lineage-specific enhancers is essential to gain a better understanding of transcriptional regulatory networks required for organogenesis. At the same time, such enhancers provide important tools for driving transgene expression in desired cell populations. Previous studies have demonstrated the utility of using conservation to identify endothelial enhancers in several zebrafish studies (Bussmann et al., 2010; Veldman and Lin, 2012). However, the lack of closely related genomes for accurate homology alignment limits a more comprehensive identification of putative enhancers using this approach (Hiller et al., 2013). Indeed, analysis of our endothelial ATAC-seq elements indicates that more than 80% do not bear known conserved sequences, suggesting that available CNE annotations only detect a very small proportion of tissue-specific enhancer elements. Similarly, whole-embryo ChIP-seq datasets likely underrepresent lineage-specific enhancers from cells that comprise a small proportion of the embryo. In any case, ATAC-seq on cell-type-specific nuclei reveals a large number of putative enhancer elements that would otherwise not be detected using previous approaches. A recent study in

Table 2. ATAC-Seq Elements Used for Reporter Assays in This Study

					Reporter Activity					
			Chromatin Mark[b]		enh-basP		prom		enh-prom	
Gene	Coordinates[a]	Relative to TSS	K27ac	CNE[c]	end	ect	end	ect	end	ect
cldn5b	chr10:45481551-45481866	−14,717	no	no	−	−	+	+	+	+
clec14a	chr17:10362331-10362838	−3,187	no	no	+	−	+	−	+++	−
***dll4* (E1)**	chr20:28219030–28219622	−55,088	no	yes	+	−	−	+	+++	+
***dll4* (E2)**	chr20:28223130–28223785	−50,993	no	no	+	+	−	+	+	++
dusp5	chr22:32639298–32640073	−26,698	yes	no	+	−	−	+	+++	++
lmo2	chr18:36722032–36722499	−3,681	no	no	+	−	+	−	+++	−
mafbb	chr11:26309960–26310474	7,807	no	yes	+	−	−	+	+	++
nrp1b	chr2:43535170–43535764	−34,648	yes	no	+	−	+	−	+++	+
she	chr16:24769545–24769995	1,942	no	no	−	++	−	−	+	++
snx5	chr13:33875217–33875816	4,149	no	no	−	−	+	−	+	−
tmem88a	chr10:22967947–22968466	−3,870	no	yes	+	−	−	+	+++	+
yrk	chr19:45217027–45217893	7,443	no	yes	−	−	+	−	+	+

end, endothelial expression; ect, ectopic expression; enh-basP, enhancer upstream of basal promoter; prom, cognate gene promoter only; enh-prom, enhancer and cognate promoter.

[a]Coordinates are based on location of PCR primers, which sometimes extend beyond annotations shown in Table S3.

[b]Chromatin modifications in whole embryos by ChIP-seq at 24 hpf as reported by Bogdanovic et al. (2012).

[c]Annotated as conserved element by Hiller et al. (2013).

mouse has made similar efforts to identify endothelial-specific enhancer elements (Zhou et al., 2017). In this case, more than 2,000 endothelial-specific regulatory elements were identified using tissue-specific biotinylation of the histone acetyltransferease, p300, followed by bioChIP-seq. However, this approach requires generation of numerous transgenic lines and relies on ChIP-seq, which is more technically demanding and requires more cells than ATAC-seq. While such "Biotagging" approaches have begun to be applied in zebrafish (Housley et al., 2014; Trinh et al., 2017), FANS-assisted ATAC-seq is less technically demanding and can be applied to available zebrafish transgenic lines using a small number of cells.

In addition to demonstrating proof of concept, our work also provides a collection of endothelial-specific enhancers that serves as an important resource to the vascular development community. Further functional dissection of elements within this collection will undoubtedly lead to unique insights into transcriptional control of vascular development and endothelial differentiation. At the same time, continued application of FANS-assisted ATAC-seq at different developmental stages, and in mutants that affect particular aspects of vascular development, will provide dynamic information on the genome-wide regulation of transcriptional networks that control gene expression during vascular development.

EXPERIMENTAL PROCEDURES

Fish Care

Fish were maintained in accordance with the University of Massachusetts Medical School Institutional Animal Care and Use Committee. The *Tg(fli1a:egfp)*^{y1} line has been described (Lawson and Weinstein, 2002). For injections, embryos were derived from group in-crosses of the EK wild-type line.

Isolation of Nuclei

Tg(fli1a:egfp)^{y1} embryos were dechorionated with pronase at 24 hpf and deyolked in calcium-free Ringers. Embryos were pelleted and resuspended in 2 mL of homogenization buffer (HB) (15 mM Tris-HCl [pH 7.4], 0.34 M sucrose, 15 mM NaCl, 60 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, and Roche Complete protease inhibitors added before use), which was pre-chilled at 4°C. Embryos were transferred to a 2-mL Dounce Tissue Grinder (Sigma-Aldrich, D8938) on ice and carefully dissociated with ten strokes with the loose pestle and 15 times with the tight pestle. Lysate was filtered through 100-μm cell strainers (Corning Life Sciences, Product# 352360) and spun at 3,500 × *g* for 5 min at 4°C to pellet nuclei. The supernatant was carefully decanted, and nuclei were resuspended in 10 mL of HB buffer and transferred to a 15-mL conical tube. Nuclei were pelleted at 3,500 × *g* for 5 min at 4°C and resuspended in 3 mL of pre-chilled (4°C) PBTB buffer (0.1% Triton X-100 in 1 × PBS; 5% BSA added prior to use and filtered through a 0.22-μm pore filter). Nuclei were transferred to a 15-mL conical tube and further dissociated by gently passing them ten times through a 21-gauge needle. The nuclei were passed through 20 μm cell strainers (EMD Millipore, SCNY00020) and sorted on a FACS Aria (BD Bioscience) in the UMass Med Flow Cytometry Lab. GFP-positive and -negative nuclei were collected into separate tubes and maintained on ice.

ATAC-Seq

20,000 GFP-positive or -negative nuclei were used for ATAC-seq as described elsewhere (Buenrostro et al., 2013). Paired-end ATAC-seq libraries were sequenced by the Genome Technology Access Center at Washington University in St. Louis. Three biological replicates were generated for GFP-positive and -negative nuclei.

RNA-Seq

For RNA-seq, we applied a modified version of the MARIS protocol (Hrvatin et al., 2014). 300 *Tg(fli1a:egfp)*^{y1} embryos at 24 hpf were deyolked in calcium-free Ringers and dissociated in 4 mL of TrypLE Express (Invitrogen; pre-warmed to 28.5°C) at 28.5°C with pipetting every 5 min for 20 min. All subsequent steps were performed on ice. Cells were passed through a 70-μm filter (BD Falcon 352340) into a 50-mL conical tube and transferred to a 15-mL tube. Cells were centrifuged at 3,000 rpm for 3 min, resuspended in 1 mL 1 × PBS, and transferred to a 2-mL Eppendorf tube. Cells were spun down, resuspended in 1 mL of 4% paraformaldehyde/PBS/0.1% saponin (Sigma-Aldrich) for 30 min at 4°C, and transferred to a 15-mL Falcon to which 3 mL of wash buffer (PBS/0.25% BSA/0.1% saponin, 1:100 RNasin Plus RNase Inhibitor [Promega]) was added. Cells were spun down and resuspended in wash buffer

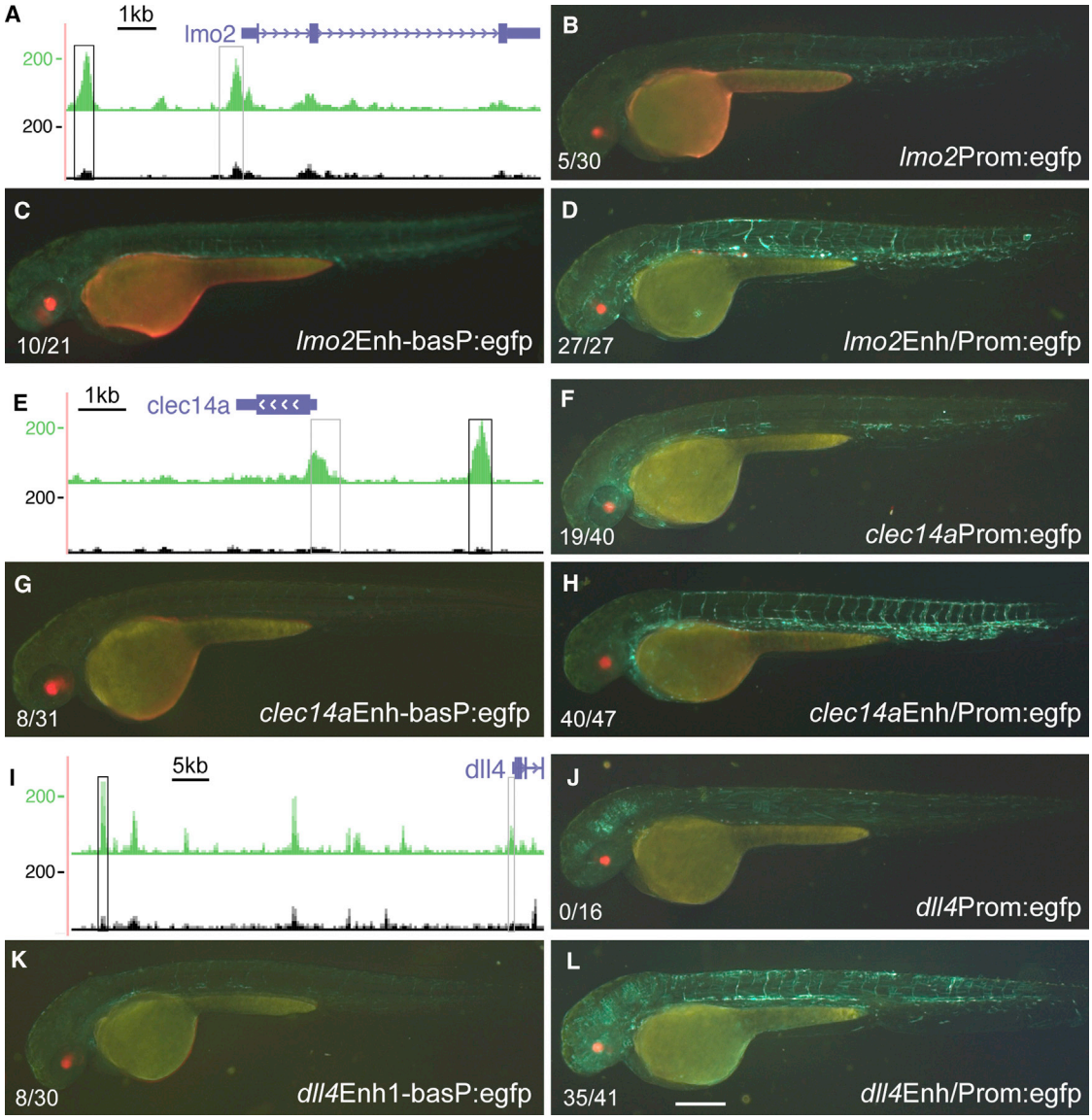


Figure 5. Pairing Cognate Promoters and Enhancers Improves Endothelial Expression

(A, E, and I) GFP-positive (green) and GFP-negative (black) ATAC-seq read density in nuclei at (A) *lmo2*, (E) *clec14a*, and (I) *dll4* loci. Black boxes are putative enhancer elements; gray boxes denote region used as a promoter.

(B–D, F–H, and J–L) Overlays of green and red fluorescence from embryos injected with enhancer reporter constructs. Ratios in left bottom are number of embryos with endothelial GFP over the total number of *cryaa:mcherry*-expressing embryos from replicate injections. Lateral views, dorsal is up, anterior to the left. Embryos injected with reporter construct containing (B, F, J) gene-specific promoter, (C, G, K) enhancer and basal promoter, or (D, H, L) enhancer and cognate promoter for indicated genes upstream of EGFP. Scale bar, 250 μm.

followed by sorting on a FACS Aria (BD Biosciences). Gates were set with reference to non-GFP controls. Sorting speed was adjusted to ensure >90% efficiency. 1 × 10^5 GFP-positive and -negative cells were collected in tubes coated with a small amount of wash buffer. Cells were pelleted, supernatant was discarded, and total RNA was isolated using the RecoverAll Total Nucleic Acid Isolation kit (Ambion) according to the manufacturer's protocol with the following modifications: the protocol was started at the protease digestion stage and cells were incubated in 100 μL of digestion buffer with 4 μL proteinase K for 1 hr at 50°C. RNA integrity number (RIN), and quantity was determined by Bioanalyzer. 5 ng of total RNA (RIN >7) was used for library construction using the TotalScript kit (Epicenter). FACS and RNA isolation from GFP-positive and -negative cells was performed twice to give biological replicates. Library sequencing was performed at the UMass Med Deep Sequencing Core Lab.

Data Analysis

For ATAC-seq, adaptor sequences were removed using cutadapt (version 1.3; Martin, 2011) and reads mapped onto Zv9 using Bowtie2 (Langmead and Salzberg, 2012). Default settings were modified to allow paired-end fragments up to 2 kb. For quality assessment of ATAC-seq libraries, we developed and applied ATACseqQC (available through Bioconductor; https://www.bioconductor.org/packages/release/bioc/html/ATACseqQC.html). To visualize mapped reads, we generated bigwig files from BAM output using deepTools2 (Ramírez et al., 2016). MACS (version 2.1; Zhang et al., 2008)

was used to call enrichment, with the following settings: callpeak -g 1.35e9 -qvalue 0.05 -bw 250 -mfold 10 30. Fragments of desired length were extracted from SAM files. Heatmaps and density plots were generated using ChIPpeakAnno (Zhu et al., 2010). To identify differentially mapped regions, we used Diffbind (Bioconductor; Ross-Innes et al., 2012). As input, we used BAM files outputted from Bowtie2 and peaks from MACS2 for each replicate. An element was GFP-positive if GFP^+/GFP^- log_2 fold change in read density ≥ 1 and FDR ≤ 0.05. Conversely, GFP-negative elements were log_2 fold change ≤ -1 and FDR ≤ 0.05. An element was "Common" if fold-change thresholds fell below that for GFP-positive and -negative and the mean read concentration was greater than four in either sample. To determine concordance, we used ChIPpeakAnno (Zhu et al., 2010) or bedTools (Quinlan and Hall, 2010). For embryo ChIP-seq, we utilized available datasets (GEO: GSE32483; Bogdanovic et al., 2012). Enriched peaks were called from these data using MACS with default settings. For CNEs, we utilized previously described elements (Hiller et al., 2013). For RNA-seq analysis, reads were mapped onto Zv9 using Tophat2 (v.2.0.9; Kim et al., 2013). Differentially expressed genes were identified using Cuffdiff (v.2.1.1) with a custom transcript annotation. To visualize genomic data, bed or bigwig files were uploaded to a local mirror site of the UCSC Genome Browser. Statistical comparison of fragment per kilobase per million reads (FPKM), and TSS accessibility was performed using an unpaired Student's t test. Correlation of flanking (2.5×10^5 bp up- and downstream of the TSS) GFP-positive, -negative, or common elements and log_2 fold change from RNA-seq was analyzed by Spearman's rank correlation. GREAT analysis was performed on the web interface (http://bejerano.stanford.edu/great/public/html/). For test regions, GFP-positive or -negative ATAC-seq elements were used as input. For background, all elements were used. To identify over-represented transcription factor sites, we used Hypergeometric Optimization of Motif EnRichment (HOMER; Heinz et al., 2010).

Cloning

Putative enhancer elements were PCR amplified using primers containing attB4 and attB1 sequences and cloned by Gateway recombination into pDONR-P4-P1R (Thermo Fisher Scientific). See Table S7 for all primer information. Enhancer plasmids are named p5E with the name of the adjacent gene and "E" (e.g., p5E-dusp5E1; see Table S8 for all entry plasmids). To generate reporter constructs in which an enhancer is upstream of a basal promoter and EGFP, p5E-E constructs were used in a Gateway reaction with pENTRbasEGFP (Addgene #22453; Villefranc et al., 2007), p3E-mcs (Addgene #49004), and pDestTol2pACrymCherry (Addgene #64023; Berger and Currie, 2013). To generate enhancer/reporter constructs with their cognate promoter, we constructed middle entry clones for each gene promoter. Each promoter was amplified by PCR with primers bearing overlap sequences flanking the Age I site in pENTRegfp2 (Addgene #22450; Table S7). Following PCR, each fragment was individually used in a HiFi Assembly reaction (New England Biolabs) with pENTRegfp2 digested with Age I. The resulting pME-promoter-Egfp plasmids (Table S8) were sequence-verified and used in Gateway LR reactions with cognate p5E enhancer plasmids, p3E-mcs, and pDestTol2pACrymCherry as above to generate reporter constructs for injection (Table S9). To generate promoter-alone reporters, LR reactions were performed as above, but using empty p5E-mcs plasmid (Addgene, #26029). All plasmids constructed in this study are available at Addgene (http://www.addgene.org/Nathan_Lawson/).

Reporter Assays

Tol2 reporter constructs were injected into 1-cell stage zebrafish embryos as previously (Villefranc et al., 2007). Injected embryos were observed at 48–55 hr post-fertilization to detect expression. Only morphologically normal embryos with lens expression of *mcherry* were used to score for Egfp. Representative images of green and red fluorescence were captured as described elsewhere (Villefranc et al., 2007). For comparison between enhancer, promoter-only, and chimeric enhancer/promoter elements, exposure time settings to detect fluorescence were kept constant. A qualitative rank (see Table 2) was assigned based on the penetrance of detectable endothelial expression in *cryaa:mCherry*-positive embryos (Figure S3A). Images of injected embryos in Figures 4, 5, and S3 are representative of qualitative ranks in Table 2 and penetrance in Figure S3A.

ACCESSION NUMBERS

The accession number for the ATAC-seq and RNA-seq data reported in this paper is GEO: GSE97257.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, nine tables, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.06.070.

AUTHOR CONTRIBUTIONS

A.Q. developed and performed FANS-assisted ATAC-seq and RNA-seq. M.A. constructed plasmids and performed and analyzed reporter injections. J.Y. ran sequencing analysis pipelines. J.O. and L.J.Z. developed the ATAC-seq analysis pipelines. N.D.L. contributed to plasmid construction, bioinformatics, designed the study, and wrote the manuscript. All authors edited the manuscript.

ACKNOWLEDGMENTS

pDESTtol2pACrymCherry was a gift from Joachim Berger & Peter Currie (Addgene plasmid # 64023). We thank Manuel Garber and Alper Kucukural for maintaining the local UCSC Genome Browser mirror. We are indebted to John Polli and Pat White for zebrafish care and maintenance. We thank Tom Fazzio for helpful comments on the manuscript. This work was supported by National Heart, Lung, and Blood Institute grants R01HL093467 and R01HL122599 (to N.D.L.).

Received: April 8, 2017
Revised: June 8, 2017
Accepted: June 22, 2017
Published: July 18, 2017

REFERENCES

Aday, A.W., Zhu, L.J., Lakshmanan, A., Wang, J., and Lawson, N.D. (2011). Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. Dev. Biol. *357*, 450–462.

Adli, M., and Bernstein, B.E. (2011). Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. Nat. Protoc. *6*, 1656–1668.

Berger, J., and Currie, P.D. (2013). 503unc, a small and muscle-specific zebrafish promoter. Genesis *51*, 443–447.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169–181.

Bogdanovic, O., Fernandez-Miñán, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J., and Gómez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. *22*, 2043–2053.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bussmann, J., Bos, F.L., Urasaki, A., Kawakami, K., Duckers, H.J., and Schulte-Merker, S. (2010). Arteries provide essential guidance cues for lymphatic endothelial cells in the zebrafish trunk. Development *137*, 2653–2657.

Covassin, L., Amigo, J.D., Suzuki, K., Teplyuk, V., Straubhaar, J., and Lawson, N.D. (2006). Global analysis of hematopoietic and vascular endothelial gene expression by tissue specific microarray profiling in zebrafish. Dev. Biol. *299*, 551–562.

De Val, S., and Black, B.L. (2009). Transcriptional control of endothelial cell development. Dev. Cell *16*, 180–195.

Dickmeis, T., Mourrain, P., Saint-Etienne, L., Fischer, N., Aanstad, P., Clark, M., Strähle, U., and Rosa, F. (2001). A crucial component of the endoderm formation pathway, CASANOVA, is encoded by a novel sox-related gene. Genes Dev. *15*, 1487–1492.

Ferg, M., Armant, O., Yang, L., Dickmeis, T., Rastegar, S., and Strähle, U. (2014). Gene transcription in the zebrafish embryo: Regulators and networks. Brief. Funct. Genomics *13*, 131–143.

Gehrig, J., Reischl, M., Kalmár, E., Ferg, M., Hadzhiev, Y., Zaucker, A., Song, C., Schindler, S., Liebel, U., and Müller, F. (2009). Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. Nat. Methods *6*, 911–916.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al.; modENCODE Consortium (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science *330*, 1775–1787.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Göttgens, B., Barton, L.M., Chapman, M.A., Sinclair, A.M., Knudsen, B., Grafham, D., Gilbert, J.G., Rogers, J., Bentley, D.R., and Green, A.R. (2002). Transcriptional regulation of the stem cell leukemia gene (SCL)–comparative analysis of five vertebrate SCL loci. Genome Res. *12*, 749–759.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hiller, M., Agarwal, S., Notwell, J.H., Parikh, R., Guturu, H., Wenger, A.M., and Bejerano, G. (2013). Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: Application to zebrafish. Nucleic Acids Res. *41*, e151.

Housley, M.P., Reischauer, S., Dieu, M., Raes, M., Stainier, D.Y., and Vanhollebeke, B. (2014). Translational profiling through biotinylation of tagged ribosomes in zebrafish. Development *141*, 3988–3993.

Hrvatin, S., Deng, F., O'Donnell, C.W., Gifford, D.K., and Melton, D.A. (2014). MARIS: Method for analyzing RNA following intracellular sorting. PLoS ONE *9*, e89459.

Jessen, J.R., Meng, A., McFarlane, R.J., Paw, B.H., Zon, L.I., Smith, G.R., and Lin, S. (1998). Modification of bacterial artificial chromosomes through chi-stimulated homologous recombination and its application in zebrafish transgenesis. Proc. Natl. Acad. Sci. USA *95*, 5121–5126.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev. Cell *7*, 133–144.

Kikuchi, Y., Agathon, A., Alexander, J., Thisse, C., Waldron, S., Yelon, D., Thisse, B., and Stainier, D.Y. (2001). casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish. Genes Dev. *15*, 1493–1505.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Komisarczuk, A.Z., Kawakami, K., and Becker, T.S. (2009). Cis-regulation and chromosomal rearrangement of the fgf8 locus after the teleost/tetrapod split. Dev. Biol. *336*, 301–312.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lawson, N.D., and Weinstein, B.M. (2002). In vivo imaging of embryonic vascular development using transgenic zebrafish. Dev. Biol. *248*, 307–318.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Meng, A., Tang, H., Ong, B.A., Farrell, M.J., and Lin, S. (1997). Promoter analysis in living zebrafish embryos identifies a cis-acting motif required for neuronal expression of GATA-2. Proc. Natl. Acad. Sci. USA *94*, 6267–6272.

modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787–1797.

Müller, F., Chang, B., Albert, S., Fischer, N., Tora, L., and Strähle, U. (1999). Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. Development *126*, 2103–2116.

Pham, V.N., Lawson, N.D., Mugford, J.W., Dye, L., Castranova, D., Lo, B., and Weinstein, B.M. (2007). Combinatorial function of ETS transcription factors in the developing vasculature. Dev. Biol. *303*, 772–783.

Phng, L.K., Potente, M., Leslie, J.D., Babbage, J., Nyqvist, D., Lobov, I., Ondr, J.K., Rao, S., Lang, R.A., Thurston, G., and Gerhardt, H. (2009). Nrarp coordinates endothelial Notch and Wnt signaling to control vessel density in angiogenesis. Dev. Cell *16*, 70–82.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: How is it established, and why does it matter? Dev. Biol. *339*, 258–266.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Reischauer, S., Stone, O.A., Villasenor, A., Chi, N., Jin, S.W., Martin, M., Lee, M.T., Fukuda, N., Marass, M., Witty, A., et al. (2016). Cloche is a bHLH-PAS transcription factor that drives haemato-vascular specification. Nature *535*, 294–298.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

Trinh, L.A., Chong-Morrison, V., Gavriouchkina, D., Hochgreb-Hägele, T., Senanayake, U., Fraser, S.E., and Sauka-Spengler, T. (2017). Biotagging of specific cell populations in zebrafish reveals gene regulatory logic encoded in the nuclear transcriptome. Cell Rep. *19*, 425–440.

Veldman, M.B., and Lin, S. (2012). Etsrp/Etv2 is directly regulated by Foxc1a/b in the zebrafish angioblast. Circ. Res. *110*, 220–229.

Villefranc, J.A., Amigo, J., and Lawson, N.D. (2007). Gateway compatible vectors for analysis of gene function in the zebrafish. Dev. Dyn. *236*, 3077–3087.

Wasserman, W.W., Palumbo, M., Thompson, W., Fickett, J.W., and Lawrence, C.E. (2000). Human-mouse genome comparisons to locate regulatory sites. Nat. Genet. *26*, 225–228.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652–657.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

Zhou, P., Gu, F., Zhang, L., Akerberg, B.N., Ma, Q., Li, K., He, A., Lin, Z., Stevens, S.M., Zhou, B., and Pu, W.T. (2017). Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. eLife, Published online January 25, 2017. http://dx.doi.org/10.7554/eLife.22039.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.