Yale University
EliScholar – A Digital Platform for Scholarly Publishing at Yale

Yale Medicine Thesis Digital Library | School of Medicine

9-16-2010

Epigenetic Programming of Blood Vessel Identity

Aaron Weldon Aday
Yale University

Follow this and additional works at: http://elischolar.library.yale.edu/ymtdl

Recommended Citation
Aday, Aaron Weldon, "Epigenetic Programming of Blood Vessel Identity" (2010). *Yale Medicine Thesis Digital Library*. 103.
http://elischolar.library.yale.edu/ymtdl/103

This Open Access Thesis is brought to you for free and open access by the School of Medicine at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Medicine Thesis Digital Library by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

Epigenetic Programming of Blood Vessel Identity

A Thesis Submitted to the
Yale University School of Medicine
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

by

Aaron Weldon Aday

2010

ABSTRACT

EPIGENETIC PROGRAMMING OF BLOOD VESSEL IDENTITY. Aaron W. Aday, Lihua J. Zhu, and Nathan D. Lawson. Program in Gene Function and Expression, UMass Medical School, Worcester, MA. (Sponsored by Michael Simons, Section of Cardiovascular Medicine, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT).

Recent studies have revealed details of the signaling pathways controlling blood vessel development and function. However, little is known about what controls endothelial cell identity in different blood vessel types. It is important to identify transcriptional control elements that function in endothelial cells in order to examine their roles in differentiation and vascular development. Certain histone modifications can serve as molecular markers for these regulatory elements. Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) allows one to identify DNA sequences bound by these histones, and mapping to a reference genome permits localization of putative enhancer and promoter regions enriched for modified histones. By using this technology to identify global epigenetic modifications associated with transcriptional activation in endothelial-expressed genes, one can locate *cis*-regulatory elements that may play essential roles in controlling cell type-specific gene expression and defining blood vessel identity.

In an effort to identify *cis*-regulatory elements that control endothelial gene expression, we have performed ChIP-Seq on zebrafish embryos. Similar to previous studies, promoters are enriched for modifications such as trimethylation of histone 3 at the fourth lysine residue (H3K4me3). Monomethylation of histone 3 at the same position

(H3K4me1) is less strongly enriched at promoter elements and often localizes up and downstream of predicted gene sequences or in intronic regions. In several cases, these corresponding sequences are evolutionarily conserved and map to known transcription factor binding sites. We have also analyzed ChIP-Seq data from endothelial cells isolated from zebrafish embryos by fluorescence-activated cell sorting (FACS), and this vascular dataset has a unique epigenetic signature compared to whole embryos. Finally, we performed *in vivo* reporter assays and confirmed that some of the candidate enhancer elements identified through ChIP-Seq are able to drive gene expression. Together, these resources will allow us to better understand the transcriptional regulatory networks that are responsible for endothelial cell heterogeneity.

ACKNOWLEDGEMENTS

We would like to thank Michael Simons and Frank Giordano for critical reading of the manuscript. We also thank John Polli for excellent fish care and maintenance and Ellie Kittler and the UMass Deep Sequencing core for their sequencing and troubleshooting efforts. We also appreciate the efforts of the UMass Medical School Flow Cytometry Core Lab. We are grateful to all members of the Lawson lab for critical discussion and teaching. In addition, we appreciate the input of Dan Hart, Claude Gazin, and Kyu Choe. Aaron Aday is a Research Fellow supported by the Sarnoff Cardiovascular Research Foundation.

TABLE OF CONTENTS

Introduction……..1

Hypothesis/Aims……16

Methods……..17

Results……28

Conclusions……38

References……..45

INTRODUCTION

Endothelial cell diversity

Although the entire human body depends on the vasculature for circulation of oxygen, nutrients, and waste products, the endothelial cells lining the vasculature possess distinct physiologic and morphologic features depending on their location. For instance, endothelial cells in the central nervous system form tight junctions, thereby establishing the blood brain barrier (BBB) that protects neurons from toxic molecules in the blood. The endothelia of the renal glomeruli, small intestine, and endocrine tissues, however, have fenestrations, which allow transport of fluid and solutes. Finally, endothelial cells in the liver and spleen construct gap junctions, which permit transport of large macromolecules and cells (1).

Although we consider the differences in these endothelial cells in terms of their anatomic location and function in the mature vasculature, there is increasing evidence that this endothelial differentiation begins during embryogenesis. For example, the Wnt signaling pathway is responsible for creation of the BBB during embryonic development in the mouse (2). In this instance, the neural tube expresses the Wnt ligands *Wnt7a* and *Wnt7b* concomitantly with vascular invasion into the neural tube. Furthermore, *Wnt7a/b* double mutants exhibit cerebral hemorrhage and decreased expression of BBB markers. Similarly, another study has shown that the endocrine-gland-derived vascular endothelial growth factor (EG-VEGF) can induce proliferation and development of fenestrae in endothelial cell cultures derived from endocrine tissues (3). Of note, this molecule has little effect on endothelial cells harvested from other tissues, and EG-VEGF mRNA is restricted to the human ovaries, testes, adrenal glands, and placenta. These results

suggest a tissue-specific response to EG-VEGF similar to that of *Wnt7a/b* in the mouse neural tube.

In addition to molecular and morphological differences in organ-specific vascular systems, further evidence supports a role for endothelial differentiation in the early development of a somatic vascular network in vertebrates. Endothelial cells in the dorsal aorta of the zebrafish must differentiate into cells with distinct morphologies to form angiogenic sprouts (4; 5). Furthermore, tip cells at the leading edge of these new vessels preferentially express *flt4*, which is notably absent from the stalk cells making up the body of the sprouts as well as from the dorsal aorta, from which these tip cells originated (5). Disruption of the pathways responsible for this differentiation leads to severe vascular defects (5-7). Similarly, populations of endothelial precursors in the anterior lateral mesoderm that ultimately give rise to the anterior lateral dorsal aortae express both *kdrl* and *cxcr4*, while those in the posterior lateral mesoderm giving rise to the posterior lateral dorsal aortae express only *kdrl* (8). These different cell populations also exhibit distinct migratory patterns as they coalesce to form the patent vasculature. Given this body of evidence, endothelial cell differentiation plays a critical role in vertebrate vascular development.

Transcriptional control of endothelial cell identity

Although we now have more insight into the degree of endothelial diversity, we lack a complete understanding of the processes controlling development of this heterogeneous cell population. One critical research focus is the cell fate decision to transition from a common endothelial precursor into a committed arterial or venous fate. In vertebrates, this commitment occurs in embryogenesis prior to the onset of circulation

(9). Numerous studies have identified roles for components of the vascular endothelial growth factor-A (Vegf-A) and Notch signaling pathways in arterial endothelial differentiation (10-15). In addition, several transcription factors related to these pathways play key roles in arterial specification. Hey2 is a helix-loop-helix transcriptional repressor necessary for proper dorsal aorta development in the zebrafish (16; 17). Initially expressed in the lateral posterior mesoderm, it is later restricted to the dorsal aorta (16). Knockdown of *hey2* results in degradation of the axial arterial vasculature and both expansion of the venous vasculature and an increase in expression of *ephb4*, a venous marker (17). Notch signaling can activate *hey2*, although it appears this may be mediated by other factors and can also occur independently of Notch signaling (18). Similarly, the forkhead transcription factors Foxc1 and Foxc2 are important for specifying arterial fate (19; 20). In mouse, *foxc1/2* mutants exhibit a shunt connecting the dorsal aorta and cardinal vein, and these mutants also lack expression of the arterial markers *dll4*, *hey2, and ephB2* (19). Further biochemical studies show that Foxc1 and Foxc2 directly interact with Notch downstream activating elements to induce *hey2* expression (20). These analyses also indicate that VEGF-activated phosphoinositide 3-kinase (PI3K) modulates Foxc1/2 activation of *dll4* and *hey2*.

Several groups have examined the sex determining region Y-related high mobility group (SOX) transcription factors and their roles in zebrafish vascular development (21; 22). *Sox7* is primarily expressed in the dorsal aorta, while *sox18* is expressed throughout the vasculature (21). Whereas *sox7* expression is dependent on *Scl*, Vegf signaling and sonic hedgehog signaling, *sox18* is unaffected by loss of *Scl* or inhibition of either Vegf or sonic hedgehog signaling (21; 22). In addition, double knockdown of *sox7* and *sox18*

results in loss of trunk circulation and the formation of arteriovenous shunts (21; 22). More recent evidence suggests a role for *sox18* in regulating lymphatic endothelial cell differentiation in mouse (23).

Another important transcription factor for vascular development is the Ets transcription factor *etsrp*. Morpholino knockdown of *etsrp* in zebrafish results in the absence of angioblast differentiation and migration and a complete lack of circulation (24). Interestingly, Foxc2 and Etv2 (the mouse orthologue of Etsrp) can bind an endothelial-specific motif present in many endothelial enhancers (25). This study also shows that combinatorial expression of these factors induces ectopic vascular gene expression in *Xenopus*, and knockdown of both factors in the zebrafish leads to aberrant vascular development. Despite our steadily increasing knowledge of roles for these and other transcription factors in vascular development, there still remain numerous gaps in our understanding of this regulatory network and how it may contribute to endothelial heterogeneity.

Cis-*regulatory elements and their roles in development and human disease*

As the Foxc2/Etv2 study illustrates (25), the *cis*-regulatory elements to which transcription factors bind also play important yet poorly understood roles in cell lineage commitment. *Cis*-elements are a group of non-coding DNA regions that include promoters, enhancers, and insulators. Proximal promoter elements are genomic regions immediately upstream of genes that serve as sites for RNA polymerase II binding and recruitment of other transcription factors necessary for the initiation of transcription. Enhancers, however, can lie many kilobases up or downstream of their target genes and may even be present within the genes themselves. When bound by transcription factors,

enhancers may reorient chromatin architecture through looping to facilitate interaction with the transcriptional machinery occupying the promoter (26). This interaction is often tissue-specific depending on the binding motifs present in the enhancer. In arthropods, a conserved enhancer in the tinman gene complex drives expression of different transcriptional repressors in a species-dependent manner, and a chromosomal inversion, which redirects action of the enhancer, determines the specific expression pattern (27). Evidence from *Drosophila* also illustrates the role of promoter tethering elements and other proximal-promoter sequences to ensure specificity of such long-range interactions (28). It is possible that enhancers in close proximity to promoters may be able to function via direct interaction, thereby forgoing chromatin looping (26). Furthermore, proteins binding to enhancers may also interact with modified histones to facilitate transcription factor recruitment (26).

In contrast, insulators play more restrictive roles within the genome to ensure correct interactions between genomic regions. Enhancer-blocking insulators inhibit enhancer/promoter interaction through protein recruitment, and barrier insulators help prevent heterochromatin formation and subsequent gene silencing from occurring in critical regions (29). Interestingly, in *Drosophila*, the Homie insulator region blocks short-range promoter/enhancer interactions but can facilitate long-range interactions (30).

Several studies have illustrated the important roles of *cis*-elements in vascular development. In the zebrafish, a 5kb region upstream of the transcription start site (TSS) of *kdrl* (VEGFR-2) is both necessary and sufficient to drive expression of *kdrl* (31). This sequence contains binding sites for Ets factors as well as FoxH1, which functions as a repressor of vascular development. *Kdrl* upstream sequences in both mouse and human

also contain some of these binding sites, suggesting evolutionary conservation of these *cis*-elements. Similarly, the FOX:ETS motif, which synergistically binds both Foxc2 and Etv2, is present in many endothelial-specific enhancers, and mutations in these motifs eliminate their corresponding endothelial expression (25).

Previously, there has been little focus on the roles of non-coding genomic elements in human disease. However, it is increasingly clear that disease states cannot be completely explained by aberrations in coding sequences alone. In a subset of patients with α- or β-thalassemia, the disease results from deletions or rearrangements causing repositioning of enhancers required for globin gene expression (32). Similarly, single nucleotide variations in the ZRS enhancer are associated with human preaxial polydactyly (33). This enhancer normally drives expression of *Shh* even though it lies in the intron of a different gene approximately one megabase away from *Shh*. Deletion of this enhancer results in loss of *Shh* expression as well as severe limb truncation in the mouse (34). Finally, some individuals possess a common variant in an enhancer located in the first intron of the *RET* proto-oncogene; this variant confers a greater risk of developing Hirschprung's disease than the risk conferred by mutations in coding sequences (35).

Given the role of *cis*-elements in both normal development and human disease, there is a definite need for more research in this area. However, one problem facing researchers is the difficulty in identifying functionally active *cis*-elements with significant roles in any given process. Although some studies have used known transcription factor binding motifs to identify enhancers, not all predicted motifs actually bind these factors (26). Furthermore, occupancy studies have shown that some transcription factors bind

sequences lacking their predicted binding motifs (25). In many cases, this approach is also limiting because the relevant DNA-binding protein is unknown.

Computational approaches to identifying* cis-*elements

To address these problems, some studies have utilized alternative approaches. One study combined transcription factor occupancy data with known expression patterns of *cis*-regulatory elements at different timepoints in *Drosophila* (36). They trained an algorithm to predict spatiotemporal expression based on this dataset, and this algorithm was able to correctly predict expression patterns with >70% accuracy. Another study in zebrafish relied on the notion that highly conserved non-coding elements (NCEs) may represent candidate enhancers because evolutionary pressure to ensure their persistence over millions of years in multiple species suggests functional significance even though they do not encode gene products (37). In this report, the researchers chose 101 candidate enhancers based on their proximity to genes with known tissue-specific expression patterns in the brain. These enhancers were located up to 500 kilobases up or downstream of these genes, and they chose enhancers with a minimum of 60% shared identity and 100 base pairs of conservation between humans and zebrafish. Based on this approach, they were highly successful in predicting enhancers with tissue specificity, and they confirmed *in vivo* function of 76/101 of the predicted enhancers. A similar study in zebrafish examined a 200kb region up and downstream of *fgf8a* in the zebrafish genome (38). They identified 18 NCEs, including some within the nearby genes *slc2a15a* and *fbxw4*, and tested these using an *in vivo* reporter assay. Of their 18 candidates, 14 were able to drive gene expression, often in a pattern consistent with that of *fgf8a*. Finally, other research has used known *cis*-regulatory elements with common gene expression

patterns to train algorithms to predict other *cis*-elements that will yield similar expression patterns (39). Using this method, researchers were able to predict and validate *in vivo* enhancers in both *Drosophila* and mouse.

Despite advances in predicting *cis*-regulatory elements, these approaches have several limitations. Without knowledge of a relevant DNA-binding protein, it still remains difficult to predict spatiotemporal expression of enhancers. Although some *cis*-elements are evolutionarily conserved in different genomes, they may have moved relative to their target genes, further complicating efforts to determine their significance. Also, DNA conservation itself presents several problems. Highly conserved regions may function as insulators, not enhancers, or may encode non-coding RNAs (32). Furthermore, cross-species conservation can be a poor predictor of functional relevance. One study deleted four highly conserved regions in the mouse genome that were known to function as tissue-specific enhancers (40). Surprisingly, researchers observed no phenotypes in these mouse mutants, although deletion of their corresponding genes yielded a phenotype in all cases. Thus, it is important to utilize more reliable methods for predicting *cis*-regulatory elements.

Epigenetic modifications and* cis-*elements

Recently, much attention has focused on using epigenetic modifications as markers for *cis*-regulatory elements and other regions of the genome. One such marker is modified histones. Histones are proteins around which DNA wraps to form chromatin. In eukaryotes, the octameric histone core consists primarily of two copies each of histones 2A, 2B, 3, and 4; an additional histone, H1, resides outside the core. Typically, 146 base pairs of DNA bound to the octameric core comprise a nucleosome. By forming

chromatin, cells are able to compact DNA within the nucleus. Additionally, cells can restrict access of transcription factors to their respective binding sites through chromatin folding and unfolding, and this interplay between euchromatin and heterochromatin is a major mechanism of gene regulation (41; 42). Following translation, cells can enzymatically modify the N-termini of histones through phosphorylation, acetylation, methylation, and ubiquitylation, all of which affect chromatin conformation and, therefore, protein access to DNA (43). Since researchers are now able to reliably detect these modifications and map them to their corresponding genomic regions, epigenetic studies have emerged as powerful tools for genome-wide prediction of genomic elements.

Next-generation deep sequencing

In order to characterize the distribution of histone modifications, chromatin immunoprecipitation (ChIP) followed by either microarray analysis (ChIP-chip) or next generation sequencing (ChIP-Seq) have become the primary technologies for such studies. These experiments generally entail crosslinking of DNA to histones and other DNA-binding proteins by formaldehyde or other reagents. The chromatin is broken into small fragments of several hundred base pairs either through sonication or micrococcal nuclease digestion. Proteins and their corresponding DNA can be isolated following incubation with antibodies specific to the histone modification of interest. The DNA/protein bond is subsequently dissolved, and only DNA sequence previously associated with the modified histone remains.

Several options exist for detection of the immunoprecipitated DNA. Perhaps the most powerful of these options is next generation sequencing. Of the new sequencing technologies, one of the most popular is Illumina's Genome Analyzer (formerly Solexa).

To prepare samples for Illumina sequencing, blunt ends are created on the DNA fragments through enzymatic polymerase or exonuclease activity. Short, proprietary Illumina oligonucleotide adapters are ligated to the fragments, and the samples undergo PCR amplification. Quality control checks such as shotgun cloning and sequencing of fragments or PCR analysis of the final product ensure that the sample preparation was successful. At this point, the samples are ready for deep sequencing. The single-stranded samples are adhered to a flow cell surface and amplified using universal primers to create unique clusters of each fragment (44). The sequencing occurs through single-base pair reversible terminator technology. Each base pair contains a removable fluorophore. Although only one base pair is incorporated into each strand at a time, all four reversible terminators can be used simultaneously. After each incorporation step, lasers excite the fluorophores, and these images are recorded. Next, the fluorophores are chemically removed, and 3' hydroxyl groups are added to allow incorporation of the next base pair. Rather than sequence the entire fragment, the reaction continues for ~20-70 base pairs, depending on the machine (45; 46). In addition, the newer Genome Analyzers can sequence from both ends of each fragment cluster, thus providing more information for subsequent genome mapping. Each machine can sequence up to eight samples, or lanes, per run, and the newer machines can yield 4-5Gb of sequence (100-400 million sequence tags) with <1% base-calling error (46; 47).

Following the sequencing reactions, the fluorescent signals are quantified, and a base-calling algorithm determines the sequence. Subsequently, the sequences must be aligned to a reference genome, and several programs, such as ELAND, MAQ, and Bowtie, exist for such a purpose. These programs differ primarily in their cut-offs for

acceptable base calling quality and their methods of handling fragments that map to multiple regions (48). After mapping, there are several ways of handling this massive amount of sequence data. The sequence tags can be uploaded to the UCSC genome browser, which can visually display the distribution of sequences and, therefore, enrichment for the histone modification of interest around TSSs, enhancers, introns, and other elements. In such cases, it is often useful to first normalize the sequence data to an "input" sample, which represents sheared chromatin that was treated with preimmune serum or that did not undergo immunoprecipitation (49). Alternatively, one can create a tab-delimited file in which regions of significant enrichment for each histone modification are sorted by genomic coordinates or number of uniquely mapping sequence tags in that region.

The newer sequencing technologies such as Illumina's offer several key benefits over microarray-based sequencing. Because it is based on oligonucleotide hybridization, ChIP-chip is subject to factors such as probe GC content and length. As a result, ChIP-chip is more prone to hybridization artifacts that can affect sequence mapping (47; 50). Additionally, microarrays require a large number of probes to ensure high-resolution mapping, and this greatly increases the cost of experiments needing such detail. This is avoided in ChIP-Seq, which provides single base pair resolution with every run. Finally, in addition to being less costly, ChIP-Seq requires less starting DNA, which is crucial for experiments using small cell populations (45).

ChIP-Seq as a tool for identifying epigenetic modifications

Given the ability of ChIP-Seq to identify protein-bound regions of DNA with single-base resolution, many researchers have begun mapping histone modifications

throughout the genome. One study in particular used the Solexa 1G Genome Analyzer to localize numerous modified histones in $CD4^+$ T cells following micrococcal nuclease digestion of chromatin (51). This study found a strong enrichment of tri-methylated lysine 4 on the H3 subunit (H3K4me3) at the TSS, and this enrichment corresponded strongly to gene expression. Interestingly, there was a distinct paucity of this enrichment from -200 to +50 of active genes, which they attributed to a loss of nucleosome architecture in order to accommodate the transcriptional machinery. Similar reports found strong enrichment of H3K4me3 at the TSS in *Xenopus* as well as at the TSS in multiple different human cell lines (52; 53). Further evidence suggests that H3K4me3 modifications can persist beyond gene activation, thereby serving as a marker of past transcription and somehow modulating future gene regulation (54). Other research suggests that H3K4me3 combined with an additional modification conveys a signal different from a single modification alone, and this bivalent modification can interact with other modified histones and transcription factors to regulate gene expression (55).

In Barski *et al*. (51), mono-methylation of the lysine residue (H3K4me1) was also associated with the TSSs of active genes, although the enrichment was less than that of H3K4me3. Also, H3K4me1 enrichment extended well beyond 2kb up and downstream of the TSS, while H3K4me3 enrichment was largely confined within 1kb of the TSS. In contrast, the modification H3K27me3 was strongly enriched in silent regions of the genome, and its distribution did not vary surrounding the TSS. This finding is consistent with other studies using $CD4^+$ T cells (56; 57)

The group also examined several known enhancers and found strong enrichment of both H3K4me1 and H3K4me3 (51). Furthermore, they examined ~3500 sites of

DNase hypersensitivity, which is used as a marker for open chromatin available for protein interactions. At these sites, they discovered H3K4me1 enrichment over a broad region and H3K4me3 enrichment in a more narrowly delineated distribution, further suggesting these two marks can indicate enhancer elements. Interestingly, they also examined H3K4me1 and H3K4me3 enrichment at binding sites for CTCF, a known insulator-binding protein. Both modifications were enriched in these regions, although H3K4me3 enrichment was slightly stronger.

One limitation of this study is the small number of enhancers examined (51). Additionally, other studies have found conflicting evidence regarding H3K4me3's association with *cis*-regulatory elements. Using a different approach in HeLa cells, one group found H3K4me1 association with 96% of binding sites for the transcription factor STAT1 (58). Of note, these sites were downstream of the TSS. Another study used ChIP-chip to examine binding of P300, an enhancer-binding protein, in HeLa cells and found that the binding sites were primarily associated with H3K4me1, not H3K4me3 (59). Finally, a ChIP-chip study in human cell lines found enrichment of H3K4me1 at predicted enhancers, although the degree of enrichment varied among cell lines (53). Therefore, H3K4me1 may be a more reliable marker of *cis*-regulatory elements.

Given the association between histone modifications and *cis*-regulatory elements, it may be possible to use ChIP-Seq as a means to globally identify candidate elements based solely on epigenetic modifications. One group attempted a similar study using the p300 enhancer-binding protein instead of histone modifications (60). In this study, ChIP-Seq was used to map p300 binding in mouse forebrain, midbrain, and limb tissue. Approximately 10-21% of the identified peaks of enriched p300 binding were highly

conserved, and most of them were >10kb from the TSS. Although most of the peaks from each tissue were unique, there was some overlap. From this dataset, they identified 86 candidate enhancer elements and cloned them into LacZ reporter constructs. Of these candidates, 87-88% functioned as enhancers in their respective tissues. This is compared to a rate of ~50% when relying upon ultraconservation as a means of predicting candidate enhancers (61). In addition, candidates were 5-16 times more likely to drive reporter expression in their respective tissue of origin than if the enhancers were predicted based on computational means alone (60).

As this study demonstrates (60), ChIP-Seq allows one to globally identify enhancer elements with high accuracy and tissue specificity. However, relying upon enhancer-binding proteins has several disadvantages. Whereas histone modifications do not bias toward one class of enhancer, ChIP-Seq for a specific protein necessarily limits the scope of the experiment. Indeed, there may be critical enhancer elements for which the corresponding binding proteins are unknown. Also, in some cases, the protein may not be conserved to such a degree that available antibodies function in different species. Finally, focusing on enhancer-binding proteins prevents one from discovering other NCEs such as insulators or miRNAs. Therefore, ChIP-Seq studies using histone modifications are ideally suited for identifying enhancers and other *cis*-regulatory elements.

Zebrafish as a model system for vascular development

Since ChIP-Seq has already shown to be useful in identifying tissue-specific *cis*-regulatory elements, it is an ideal technique for identifying such elements in the vasculature (60). The zebrafish, *Danio rerio*, offers several key benefits in studying

vascular development. Zebrafish embryos develop external to the mother, which permits easy collection and visualization. At 24 hours post fertilization (hpf), the heart begins to beat, and the cardiovascular system is largely developed by 72hpf (62). Also, embryos possess transparent skin, which permits whole-mount *in situ* hybridization as well as direct visualization of the heart, vasculature, and other organs. Transgenic fish lines have been developed which allow visualization of the vasculature using GFP under the control of an endothelial-specific enhancer/promoter (63). Technology already exists for rapidly cloning candidate enhancers and confirming their function *in vivo* (37; 64). Given these advantages, the zebrafish constitutes an excellent model system to investigate the *cis*-regulatory network controlling vascular development.

HYPOTHESIS/AIMS

By using histone modifications as markers, we identified *cis*-regulatory elements that function to induce gene expression within the zebrafish vasculature.

1. Create an epigenetic map of histone modifications in the zebrafish embryo. Using ChIP-Seq, we mapped the distribution of various modified histones throughout the genome. We determined the correlation of these modifications to TSSs, NCEs, and known transcription factor binding sites. Also, we developed this epigenetic map for whole embryos as well as a purified population of zebrafish endothelial cells.

2. Determine epigenetic signatures for common transcription factor binding sites. We examined datasets composed of genes in common signaling pathways. Within these sets, we looked at regions of enrichment for specific modified histones and determined overrepresented transcription binding sites. By doing so, we demonstrated that binding sites for some critical transcription factors are significantly associated with specific modified histones, further demonstrating the utility of such epigenetic maps in predicting the functional relevance of non-coding genomic regions.

3. Identify *cis*-regulatory elements that function in the zebrafish vasculature. Using our epigenetic map, we identified candidate enhancer elements based on the degree of modified histone enrichment, evolutionary conservation, and proximity to genes with vascular expression. We then cloned these candidates into a reporter construct and validated their function *in vivo*, thereby demonstrating the utility of ChIP-Seq in globally identifying *cis*-regulatory elements.

METHODS

Except where noted, all experiments were performed by Aaron Aday.

Zebrafish husbandry

Fish lines were raised and maintained by Animal Medicine at UMass Medical School as previously described (65). Wild type (WT) lines were TL and CF. The *Tg(fli1:egfp)*^{y1} fish line has previously been developed (63).

Chromatin immunoprecipitation

For ChIP on whole embryos **(Fig. 1A)**, I used 200 WT embryos staged at 24hpf according to accepted standards (62). Embryos were dechorionated with a 3-minute pronase incubation and washed with calcium-free Ringers (0.5M NaCl, 0.3M KCl, 0.5M HEPES, pH 7.2). I then deyolked the embryos by passing them repeating through a P200 pipet tip, washed with calcium-free Ringers, and dissociated the embryos for one hour at 28°C in a 1X trypsin solution. This reaction was halted with collection medium (Leibovitz L-15, 0.3mM glutamine, 0.8mM $CaCl_2$, Pen 50U/mL/Strp 0.05mg/mL, and 11% fetal calf serum). I washed the cells twice with collection medium, centrifuging between each wash. Next, I created DNA/protein crosslinks with a 1% formaldehyde solution in 1X PBS for 10 minutes at room temperature. After adding 2M glycine, I centrifuged the samples and resuspended the pellets in SDS lysis buffer (50mM Tris-HCl, pH 8.1, 10mM EDTA, 1% SDS). In order to shear the chromatin, I sonicated each sample using a Microsonicator (Cole and Palmer Instruments) with a 6mm micro tip. Each sample underwent sonication for a minimum of five minutes at 30% intensity, and samples were kept in ice water during the sonication. After five minutes, I analyzed 1% of the sample using gel electrophoresis on a 1% TAE agarose gel to ensure the

distribution of chromatin fragments was between 200-1,000bp. If the range significantly exceeded 1,000bp, I performed additional sonication in intervals of 1 minute, again analyzing 1% of the sample after each additional round. I then centrifuged each sample and diluted the supernatant 1:10 in IP buffer (16.7mM Tris-HCl PH 8.1, 1.2mM EDTA, 1.1% Triton-X-100, 0.01% SDS, 167mM NaCl). 1% of this solution was set aside and stored at -20°C to serve as an input control for each corresponding immunoprecipitation. Next, each sample was incubated in an agarose bead solution. This "pre-clearing" helps remove any elements from the sample that may bind non-specifically to the agarose beads themselves (58). To accomplish this, I centrifuged ChIP-grade protein A agarose beads (Sigma), removed the supernatant, and washed the beads in IP buffer. After again centrifuging, I added block solution to the beads. This consisted of sheared salmon sperm (Invitrogen) at 20μg per 50μL of final volume, BSA at 1mg/mL, and IP buffer until the final volume was equal to that of the initial unwashed bead slurry. I added 80μL of this bead mixture to each sample and incubated 30 minutes at 4°C. After centrifuging, I collected the supernatant and split it evenly into microcentrifuge tubes. For every sample, half of this supernatant was treated with antibody and half served as a negative control. I used 8μg of antibody per sample, and our antibodies included anti-H3K4me1 (ChIP-grade, Abcam) and anti-H3K4me3 (ChIP-grade, Abcam). All samples were incubated overnight at 4°C. Both antibody-treated and untreated samples were then incubated with protein A agarose beads. These were washed and combined with blocking solution as before, but only 50μL was added to each sample. After a one hour incubation at 4°C, I centrifuged samples and washed with ChIP wash buffers A (20mM Tris-HCl pH 8.1, 2mM EDTA, 1% Triton-X-100, 0.1% SDS, 150mM NaCl), B (20mM

Tris-HCl pH 8.1, 2mM EDTA, 1% Triton-X-100, 0.1% SDS, 500mM NaCl), C (10mM Tris-HCl pH 8.1, 1mM EDTA, 1% NP-40, 1% sodium deoxycholate, 0.25M LiCl), and twice in TE solution (10mM Tris-HCl, 1mM EDTA). Next, I added 200µL elution buffer (0.1M $NaHCO_3$ and 1% SDS) to all samples. I also added 180µL of elution buffer to input samples for a final volume of 200µL. These were vortexed briefly and then incubated 15 minutes at room temperature. At this point, I set aside the input samples and centrifuged the remaining samples that contained agarose beads twice, collecting the supernatant each time. This ensured no agarose beads remained in our samples. Finally, I added 20µL of 5M NaCl to all samples (including input) and incubated overnight at 65°C to reverse the formaldehyde crosslinks. After this step, I purified all samples using a PCR purification kit (Qiagen) and eluted with 50µL EB. All ChIPs were performed in duplicate.

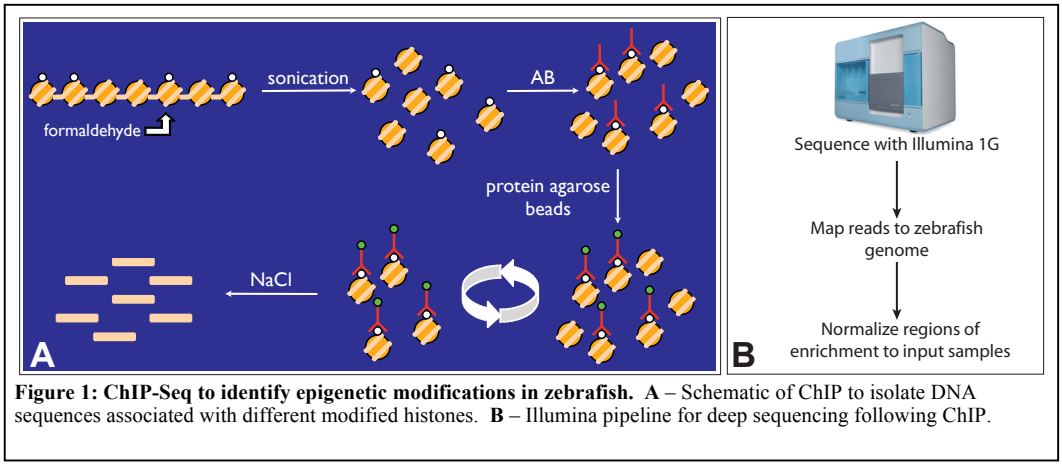

Figure 1: ChIP-Seq to identify epigenetic modifications in zebrafish. **A** – Schematic of ChIP to isolate DNA sequences associated with different modified histones. **B** – Illumina pipeline for deep sequencing following ChIP.

In order to determine the minimum number of cells needed for successful ChIP, I prepared WT embryos as described above. I counted cells using a hemacytometer and proceeded to ChIP experiments with $1x10^5$, $5x10^5$, and $1x10^6$ cells using the same protocol as for whole embryos.

For ChIP on endothelial cell populations, I began with 1,000-1,500 *Tg(fli1:egfp)*^{y1} embryos. These were dechorionated, deyolked, and trypsinized as described above. Fluorescence-activated cell sorting (FACS) was performed by the UMass Medical School Flow Cytometry Core Lab. I used ~500,000 GFP+ cells, which consisted almost entirely of endothelial cells, as well as ~500,000 GFP- cells as a negative control for each immunoprecipitation. Samples only required 4 minutes of sonication. Also, for these experiments, I only used the anti-H3K4me1 antibody, and I added 4μg instead of 8μg. The protocol is otherwise the same as for whole embryos. All experiments were performed in duplicate.

Confirmation of successful ChIP

In order to confirm the success of each ChIP experiment, I performed several quality checks. For whole embryos, I first measured each sample's concentration using a Nanodrop 2000 (Thermo Scientific). Next, I performed PCR amplification on all samples using primers specific for ~100bp regions upstream of the TSS for several housekeeping and vascular genes. These genes included *gapdh*, *ef1α*, *aqp8*, and *fli1b* **(Table 1)**. Primers were designed using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and synthesized by Invitrogen. For the PCR reactions, I used Platinum Taq polymerase (Invitrogen). I analyzed samples by gel electrophoresis on 2.5% TAE low-res ultra agarose gels.

The yield from ChIPs performed on small populations of embryonic cells as well as our purified endothelial cell populations was too low for detection by Nanodrop. Instead, for these samples, I performed quantitative PCR using Applied Biosystems instruments and primers specific to promoter regions of *fli1b*, *aqp8*, and *ef1α* **(Table 1)**.

Each reaction was performed in triplicate in a 96-well plate using Power SYBR green PCR Master Mix (Applied Biosystems) according to described methods (66). I quantified these promoter sequences for antibody, input, and negative control samples.

Gene	Forward Primer (5'→3')	Reverse Primer (5'→3')
fli1b	CTTGTTATGCGCTTACCTGGTC	GTTACATGCTGTGTATGCCATTG
aqp8	CTCACTGTGACGGCAAATATC	TTTAAGCAGTCGGCTACAGG
gapdh	AAATTACTTCTGCCTGGTTTCC	GGCCATGATTCTGTAAGCAC
ef1α	AGGCGGGGAGATTTTCAG	TTTATATGCGGGAGGAGGAC

Table 1: Primers for ChIP Validation and qPCR

After detection, I averaged the threshold cycles (Ct) for each sample. I next determined fold enrichment of antibody sample relative to negative control with normalization to the average Ct for input samples. This was accomplished using the following calculation:

$$2^{[(\text{Average input Ct}-\text{Average antibody Ct})-(\text{Average input Ct}-\text{Average negative control Ct})]}.$$

Genome Analyzer sample preparation

For whole embryos, I began with 80ng of ChIP-isolated DNA for both antibody and input samples. First, I created blunt ends on all fragment using an END-IT DNA Repair Kit (Epicentre). This reaction proceeded for 45 minutes at room temperature, after which samples were purified using a Gel Extraction Kit (Qiagen) and eluted with EB. Next, I added a single adenosine to the 3' ends using 50 units of Klenow Exo-minus (Epicentre). After one hour at room temperature, I again purified the samples using a Gel Extraction Kit. At this point, I concentrated the samples using a speed vacuum. Next, I ligated Genome Analyzer adapters (Illumina, P-GATCGGAAGAGCTCGTATGCCGTC TTCTGCTTG and ACACTCTTTCCCTACACGACGCTCTTCCGATCT) to the ends of our sample fragments. For this, I used a Fast Link Kit (Epicentre) along with the adapter mix at 0.3μM. After a two-hour incubation at room temperature, I added additional

ligase, buffer, and ATP and incubated the samples overnight at 16°C. I purified our samples with a Gel Extraction Kit and began PCR amplification. Initially, I performed test PCRs with various numbers of amplification steps to determine the correct number of steps that would provide us enough working material without increasing our risk of PCR-induced artifacts (67). Because of the high cost of Illumina's proprietary primers, I designed my own primers specific to the Illumina adapters (F – AATGATACGGCGACC ACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT and R – CAAGC AGAAGACGGCATACGAGCTCTTCCGATCT). For the PCR, I used PfuUltraTM II Fusion HS DNA polymerase (Stratagene) and 2μL of my adapter-ligated samples. I analyzed 10% of our sample following 15, 18, and 20 rounds of amplification on 2% TAE low-res ultra agarose. After deciding on an appropriate number of amplification steps, I repeated this PCR using Illumina's Genome Analyzer primers (primers 1.1 and 2.1, final concentration of 2.5μM).

Since I could not reliably quantify samples from our cell sorting experiments, I used 100% of our ChIP-isolated DNA (~30μL) instead of specifying 80ng for Genome Analyzer sample preparation. Also, for the adapter ligation step, the final concentration of adapters was 1.5μM. Otherwise, the preparation steps were identical.

Size selection and quality check

After PCR amplification, I used gel electrophoresis to analyze our samples on 2% TAE low-res ultra agarose. Using a clean razor blade, I excised the gel containing bands between 200-400bp, carefully avoiding smaller bands that contained self-ligated adapters (46). I separated this agarose equally into two microcentrifuge tubes and performed gel

extraction using a Qiagen Gel Extraction Kit. I eluted with 30μL EB, pooled identical samples, and quantified using the Nanodrop.

In order to confirm success of our sample preparation prior to sequencing, I performed PCR using Platinum Taq (Invitrogen) and primers for promoter regions of *aqp8*, *fli1b*, and *gapdh* **(Table 1)**. Afterward, I performed gel electrophoresis on 2.5% TAE low-res ultra agarose gels. Since I had already performed PCR amplification and gel extraction to ensure all remaining samples possessed properly ligated adapters, samples that also yielded bands from this PCR test represented successfully prepared samples.

Sequencing

All deep sequencing was performed on an Illumina 1G Genome Analyzer **(Fig. 1B)** operated by the UMass CFAR Deep Sequencing Core. Sequencing was performed using 36bp single-end reads. For whole embryos at 24hpf, we sequenced 6 total lanes of the H3K4me1, H3K4me3, and input samples as well as 4 lanes of each replicate. For the endothelial cell ChIP-Seq experiments, we sequenced 4 total lanes for the H3K4me1 and input GFP+ samples as well as 1 lane for each replicate.

Base calling and mapping

We used an existing pipeline consisting of the Firecrest and Bustard applications for base calling and generating sequence tag files. This analysis was performed by the UMass CFAR Deep Sequencing Core.

Next, we mapped the sequence tags to the zebrafish genome (Zv7) using BLAT. This software generated genomic coordinates for all tags (68). Only tags with unique genomic coordinates and less than 2bp of mismatch were included in subsequent

analyses. This analysis was performed by Dr. Julie Zhu, a bioinformatics programmer at UMass Medical School.

Computational analysis of putative* cis*-regulatory elements

To determine regions of significant enrichment for specific histone modifications, termed "peaks," we used the MACS algorithm, which allowed us to calculate significant differences in modified histone enrichment between our ChIP and input samples (69). The p-value cutoff for a significant peak was $1e^{-5}$. After these analyses, the genomic coordinates of all significant peaks were formatted into tab-delimited files, and I also uploaded this data to the UCSC genome browser (http://genome.ucsc.edu/) for visual analysis. The MACS analysis was performed by Dr. Julie Zhu.

We also analyzed our dataset for overrepresentation of transcription factor binding sites. To do this, we first used Clover, which is an algorithm that scans DNA sequences for over- and under-represented motifs of known transcription factor binding sites (71). We searched for NCEs as well as regions of H3K4me1 and H3K4me3 enrichment within sequences 40kb up and downstream of TSSs for several genes in the Notch pathway. For the seven genes from this set with NCEs or areas of enrichment (*nrarpa, nrarpb, her4.2, her6, notch2, notch3,* and *hes5*), we input sequences 750bp up and downstream of these loci into Clover. The algorithm output consisted of binding sites for Suppressor of Hairless (Rbpsuh), a component of the Delta-Notch pathway (72), that were overrepresented with relation to combinations of H3K4me1, H3K4me3, and NCEs. As a control, we compared this output to a set of FGF-responsive genes (*erm, fgf3, fgfr1, fgf8, dusp6, spry1, spry4,* and *pea3*) that also had H3K4me1 or H3K4me3 enrichment within 40kb of their TSSs. In addition, we analyzed the entire chromosome

15 as an additional control. Finally, we performed the same analysis on our ChIP-Seq data for Notch and FGF genes by examining for overrepresentation of a large set of known transcription factor binding sites (including Rbpsuh).

Similarly, we used the Multiple EM for Motif Elicitation (MEME) algorithm to look for overrepresented motifs (73). This algorithm does not use known binding site sequences and, instead, simply examines motifs within unaligned DNA sequences. Output from this algorithm consists of a weighted sequence matrix. For MEME analysis, we used the same sequences as with Clover. Clover and MEME analyses were performed by Nathan Lawson.

Enhancer cloning

To validate candidate enhancer elements, I used ChIP-Seq data from our whole embryo H3K4me1 dataset. I considered the best candidates those with the most significant enrichment either within or shortly up or downstream of our selected target genes: *her6*, *notch3*, and *dll4*. For *her6* and *notch3*, I designed PCR primers to amplify a region ~500-750bp up and downstream of each peak **(Table 2)**. I used three peaks for

Gene	Peak	Forward Primer (5'→3')	Reverse Primer (5'→3')
her6	A	GATCGATATCGAGAAAGAAACGCCAAGTCG	CTAGCTATAGGCAGGGATATTGCAAATGGT
	B	GATCGATATCTGCAACTCATCCTGTGGTGT	CTAGCTATAGATCCTTGCGGGCTTTAACTT
	C	GATCGATATCGCCAAAAGTTCAGGAGCAGA	CTAGCTATAGGTTCGCCTCTTTCTCACGAC
notch3	A	GATCTACGTAGGCAGTAGTGCAATCCCAAA	GATCTACGTACAGTGTGTGAAGTTGCAAAGG
	B	GATCGATATCCCTGTTCATCCCAATTACACG	CTAGCTATAGCACTGCTAAATTATTTTCTTTCCTTTG

Table 2: Primers for *her6* and *notch3* Candidate Enhancer Amplification

her6 and two peaks for *notch3*. The primers were designed with blunt-end restriction digest sites (SnaB1 or EcoRV, depending on the enhancer sequence) on each end.

After amplifying genomic DNA using HotMaster Taq Polymerase (5Prime), I performed gel extraction and restriction digested the fragment ends. I subsequently ligated each candidate enhancer upstream of an E1A basally active promoter in a

Gateway entry clone and confirmed the sequence and 5'→3' orientation by PCR colony screen and direct sequencing. I then performed a one hour LR reaction at room temperature (LR Clonase II, Invitrogen) to place this enhancer/promoter element upstream of the mCherry red fluorescent protein and an SV40 polyadenylation signal, thus creating a reporter construct to test enhancer function *in vivo* (64). I confirmed successful reporter construct synthesis by restriction digest.

For *dll4*, I designed attB4 and attB1 Gateway primers flanking a region ~500-750bp up and downstream of each peak **(Table 3)**. After PCR amplification, I performed a BP reaction (BP Clonase II, Invitrogen) for one hour at room temperature to recombine

Peak	Forward Primer (5'→3')	Reverse Primer (5'→3')
A	GGGGACAACTTTGTATAGAAAAGTTGAGTGCCAGGCCTCAATAATC	GGGGACTGCTTTTTTGTACAAACTTGCCGCGTAGTGGTGGGTATTA
B	GGGGACAACTTTGTATAGAAAAGTTGGATTTGCGTTCGAGCTTACC	GGGGACTGCTTTTTTGTACAAACTTGGTGCACGCTTTCAGTGATGT

Table 3: Gateway-Compatible Primers for *dll4* Candidate Enhancer Amplification

our candidate enhancers into a Gateway-compatible donor vector. I confirmed the sequence and 5'→3' orientation by PCR colony screen and direct sequencing. Next, I performed a one-hour LR reaction at room temperature to recombine this candidate enhancer into a Gateway expression clone upstream of the E1A promoter, EGFP, and SV40 polyadenylation signal. I confirmed the expression clone sequence by restriction digestion.

Injections

I injected reporter constructs along with Tol2 transposase mRNA into WT embryos at the one-cell stage, which permitted random incorporation into the zebrafish genome (74). This was accomplished using a microinjection needle mounted to a three-planar micromanipulator (World Precision Instruments) as previously described (75). Embryos were oriented in a Petri dish on 3% agarose medium for injections. I injected

25 and 50pg of all constructs into ~100 embryos in duplicate using a PV 820 Pneumatic Pico Pump (World Precision Instruments). At 24hpf, I added 0.003% 1-phenyl-2-thiourea to prevent pigmentation. I scored embryos for fluorescence from 24-48hpf.

Immunostaining

Because fluorescence driven by these candidate enhancers can sometimes be difficult to detect visually, I also immunostained embryos following injection with antibodies specific to the respective fluorescent proteins. Immunostaining was performed as previously described (65). I used a rabbit anti-GFP antibody (Invitrogen) for detection of EGFP and a polyclonal anti-DsRed antibody (Living Colors, Clontech) for detection of mCherry.

Microscopic imaging

I scored embryos for immunofluorescence using a Leica MZFLIII Fluo Combi fluorescence stereomicroscope. Nomarski imaging was performed using an Axioskop2 Plus compound microscope (Zeiss), and images were captured using an AxioCam MRc digital camera (Zeiss).

RESULTS

Because no previous studies have used ChIP-Seq to localize histone modifications in the zebrafish genome, we first needed to confirm such an approach was technically feasible. Initially, we defined proper sonication conditions, as it is important to ensure consistent fragment length generation without excessive sonication, which increases the risk of DNA damage (76). By sonicating chromatin derived from 24hpf WT whole embryos a total of five cycles, we were able to reproducibly obtain a distribution of chromatin ranging from 200-1,000bp **(Fig. 2)**. Because the DNA was protein-bound, the true range of fragment size was less than that displayed on the agarose gel. We found similar results using four rounds of sonication in our GFP+ and GFP- endothelial cell populations derived from *Tg(fli1:egfp)*^{y1} embryos (data not shown).

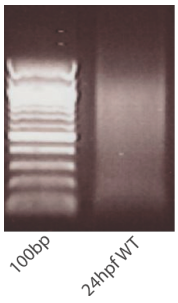

Figure 2: Sonication adequately shears chromatin for subsequent immunoprecipitation. Chromatin derived from 200 WT embryos at 24hpf and sonicated for five minutes compared to 100bp ladder. The majority of protein-bound fragments lie between 200-1,000bp.

Following each ChIP experiment, we also performed several quality checks before ligating adapters for Genome Analyzer sequencing. With ChIP experiments on WT embryos, we first determined the concentrations of each sample. For these experiments, we found DNA enrichment for input and antibody samples ranging between 2-4ng/μL, but the concentrations for negative controls lacking antibody precipitation were below the 2ng/μL threshold of detection for the Nanodrop **(Table 4)**.

In addition, we performed PCR analysis on each sample as another means of quality control. Because both H3K4me1 and H3K4me3 localize to promoter regions of active genes (51), we used primers specific to the promoter regions of several genes

expressed at 24hpf (data not shown). For ChIP using anti-H3K4me3, we detected a strong gel band for input and antibody samples for both vascular and housekeeping genes **(Fig. 3)**. Bands for ChIP samples were slightly less intense. However, these corresponding sequences were nearly undetectable for our negative controls. We found similar results with experiments using anti-H3K4me1 (data not shown).

Sample	Concentration (ng/μL)
Input	2.4
No Antibody	1.8
H3K4me1	3.8
Input	2.6
No Antibody	1.8
H3K4me3	2.9

Table 4: Spectrophotometry of Whole Embryo ChIP Samples

We also wanted to determine if we could perform ChIP on endothelial cells isolated from transgenic zebrafish. This presented a technical challenge because it would be difficult to isolate large numbers of cells from each clutch of embryos. Thus, we first performed ChIP on various numbers of WT cells to determine the feasibility of such an approach. Following ChIP, we subjected the remaining DNA to qPCR analysis. Through this approach, we found significant enrichment of immunoprecipitated DNA for promoter regions of both *fli1b* and *ef1α* relative to samples with no added antibodies **(Table 5)**. This was true for experiments using $5x10^5$ and $1x10^6$ cells with both anti-H3K4me1 and anti-H3K4me3 antibodies. Because it would be easier to collect a smaller number of cells, we used $5x10^5$ cells for subsequent experiments.

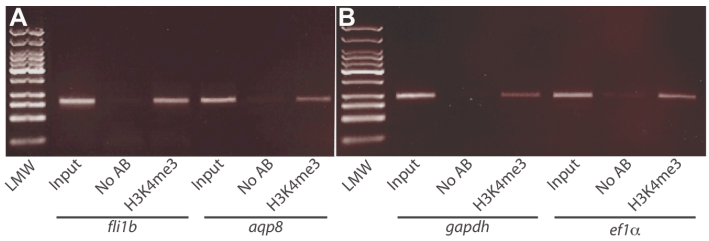

Figure 3: ChIP can be used to purify DNA associated with specific modified histones in zebrafish embryos. A – PCR performed on ChIP fragments isolated from 24hpf WT embryos using anti-H3K4me3. Primers were specific to promoter regions of *fli1b* and *aqp8*, which are both expressed in the vasculature. PCR amplified these sequences in both input and ChIP samples, but these sequences were nearly undetectable in negative controls. **B** – PCR performed on the same samples using primers specific to promoter regions of two housekeeping genes, *gapdh* and *ef1α*. Again, these sequences are present in the input and ChIP samples but not the negative controls.

We performed ChIP and qPCR analysis on GFP+ and GFP- cells isolated from *Tg(fli1:egfp)*^{y1} embryos. Because some evidence suggests H3K4me1 may be a better marker of enhancer elements (53; 58; 59) and because the anti-H3K4me3 antibody does not appear to be successful in ChIP with these small populations of cells **(Table 5)**, we only used anti-H3K4me1 for these experiments. As shown in **Table 6**, we again see strong enrichment of this histone modification in both GFP+ and GFP- cells using promoter-specific primers for *fli1b*, *aqp8*, and *ef1α*.

Number of Cells	Antibody	Fold Enrichment for *fli1b* (AB vs. no AB)	Fold Enrichment for *ef1α* (AB vs. no AB)
$1x10^6$	H3K4me1	45.20	18.83
	H3K4me3	5.59	7.80
$5x10^5$	H3K4me1	16.56	46.60
	H3K4me3	N.D.	N.D.

Table 5: qPCR Analysis of Dissociated 24hpf WT Embryonic Cells

Having confirmed that we could successfully perform ChIP on both whole zebrafish embryos as well as small populations of cells collected by FACS from transgenic embryos, we proceeded to prepare our samples for Illumina Genome Analyzer sequencing. One crucial step in this process is PCR amplification of adapter-ligated fragments. Although it is important that we have sufficient sample prior to deep sequencing, over-amplification can result in artifactual overrepresentation of certain fragments (67). Thus, we performed test PCR reactions with various cycles of amplification. By doing this for each sample, we could determine the appropriate number of cycles necessary. **Figure 4** shows three different

Cells	Antibody	Fold Enrichment for *fli1b* (AB vs. no AB)	Fold Enrichment for *aqp8* (AB vs. no AB)	Fold Enrichment for *ef1α* (AB vs. no AB)
$5x10^5$ GFP+	H3K4me1	3.39	16.19	5.14
$5x10^5$ GFP-	H3K4me1	5.02	8.30	5.56

Table 6: qPCR Analysis of 24hpf *Tg(fli1:egfp)*^{y1} Endothelial Cells

24hpf WT samples derived from whole embryos. In this case, eighteen rounds of amplification was the minimum number required for detection of a strong DNA smear for all samples. We found similar results for experiments using endothelial cell isolates (data not shown). This method also allowed us to clearly visualize self-ligated adapters, which were represented by a band slightly smaller than 100bp **(Fig. 4, arrow)**.

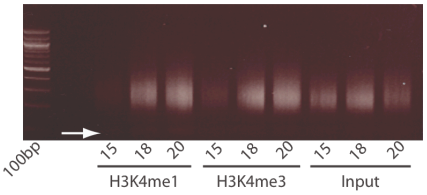

Figure 4: Few rounds of PCR amplification are needed for ChIP-Seq. Adapter-ligated ChIP-seq fragments from 24hpf WT whole embryos. Each sample was amplified for 15, 18, and 20 rounds, and 10% was analyzed by gel electrophoresis after each of these series. In this case, 18 rounds of PCR amplification provided adequate DNA for deep sequencing. Also, there was sufficient resolution between sample and self-ligated adapters **(arrow)**.

Once we confirmed that we could also successfully prepare our ChIP samples for deep sequencing, we sequenced our WT H3K4me1, H3K4me3, and input samples, which resulted in 11.9, 7.3, and 8.0 million uniquely mapping reads, respectively. For our endothelial cell samples, we sequenced 6.6 million uniquely mapping H3K4me1 tags and 7.4 million input tags. Subsequently, we uploaded this data to the UCSC Genome Browser for visualization. **Figure 5** shows the distribution of both H3K4me1 and H3K4me3 enrichment in 24hpf WT whole embryos near *notch3*. In this case, we see five

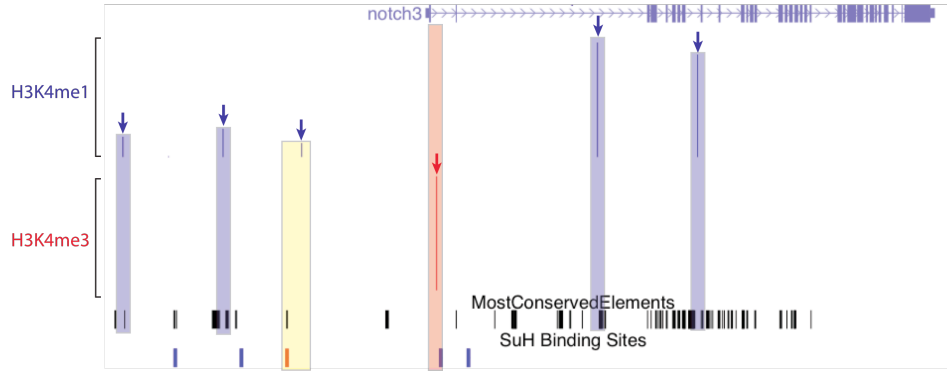

Figure 5: ChIP-Seq identifies regions of H3K4me1 and H3K4me3 enrichment that also correspond to NCEs and protein binding sites. ChIP-seq on 24hpf WT whole embryos with anti-H3K4me1 and anti-H3K4me3. Data for *notch3*, which is expressed by arterial endothelial cells as well as the CNS, is displayed in the UCSC Genome Browser. There are five regions **(blue arrows)** of H3K4me1 enrichment. Of note, four of the regions **(blue shading)** correspond to NCEs highly conserved in vertebrates, while one region **(yellow shading)** corresponds to both an NCE and an Rbpsuh binding site. We also observe strong enrichment of H3K4me3 **(red arrow)** at a region corresponding to both the TSS of *notch3* as well as an Rbpsuh binding site **(red shading)**.

regions of H3K4me1 enrichment **(Fig. 5, blue arrows)** both upstream and within the gene sequence. In all cases, these regions also correspond to NCEs **(Fig. 5, blue shading)**, and one region also maps to an Rbpsuh binding site **(Fig. 5, yellow shading)**. We also see enrichment of H3K4me3 at both the promoter region as well as an Rbpsuh binding site **(Fig. 5, red arrow and red shading)**.

We compiled our ChIP-Seq data to determine the distribution of H3K4me1 and H3K4me3 around TSSs **(Fig. 6)**. For H3K4me3, we see high amplitude enrichment at the TSS with the majority of this enrichment slightly skewed upstream **(Fig. 6A)**. This peak of enrichment falls sharply up and downstream of the TSS. With H3K4me1, we again see enrichment at the TSS, although this peak is broader and much lower in amplitude **(Fig. 6B)**. Furthermore, we see low amplitude enrichment within 10kb up and downstream of the TSS, although this enrichment is greater in amplitude upstream.

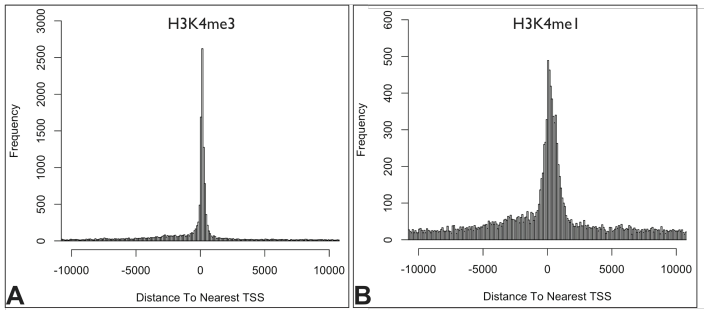

Figure 6: ChIP-Seq characterizes unique distributions of H3K4me3 and H3K4me1 around TSSs in the zebrafish genome. ChIP-Seq on 24hpf WT whole embryos. **A** – Graph A shows the correlation between H3K4me3 peaks and TSSs. These peaks cluster at the TSS with very few peaks up or downstream. **B** – Graph B shows the correlation between H3K4me1 and TSSs. Note the much smaller amplitude of enrichment at the TSS compared to H3K4me3. Also, the enrichment up and downstream of the TSS does not taper as drastically as that of H3K4me3.

Next, we examined all genes in the zebrafish genome and characterized the presence of H3K4me1 and H3K4me3 peaks and well as NCEs +/- 350bp of their TSSs **(Table 7)**. Although the vast majority of genes exhibited no epigenetic enrichment near their TSSs, we found ~2,500 H3K4me1 and ~7,100 H3K4me3 peaks in these regions.

Approximately 1,800 TSSs were marked by both H3K4me1 and H3K4me3, and we found several hundred regions of epigenetic enrichment that also corresponded to NCEs.

	No Peaks	H3K4me1 Peak	H3K4me3 Peak	NCE	me1 + me3 Peaks	me1 Peak + NCE	me3 Peak + NCE	me1 + me3 + NCE
Number of Genes with Corresponding Features	16,056	2,514	7,106	2,250	1,779	320	806	211

Table 7: Characterization of Regions +/- 350bp of TSSs in 24hpf WT Embryos

We subsequently broadened our analysis to include all regions of the genome, not just TSSs **(Table 8)**. By doing so, we identified nearly ten times more regions of H3K4me1 enrichment compared to H3K4me3. There are ~11,500 regions with both, of which we already determined ~1,800 are located near the TSS **(Table 7)**. Similarly, whereas only a few hundred regions with epigenetic marks in addition to NCEs were located near TSSs **(Table 7)**, we identified several thousand more within the whole genome **(Table 8).**

	H3K4me1 Peak	H3K4me3 Peak	me1 + me3	me1 + NCE	me3 + NCE	me1 + me3 + NCE
Number of Regions with Corresponding Elements	227,742	22,450	11,483	11,465	7,893	4,275

Table 8: Epigenetic Characterization of the Entire WT Genome at 24hpf

Using the Clover algorithm, we scanned regions corresponding to NCEs as well as H3K4me1 and H3K4me3 enrichment near genes within the Notch signaling pathway for overrepresentation of Rbpsuh binding sites. Within this pathway, we identified seven genes characterized by flanking sequences containing both enrichment for an NCE or a specific histone modification as well as a statistically significant overrepresentation of Rbpsuh binding sites **(Table 9)**. Of note, the majority of these Rbpsuh binding sites did not map to NCEs. When the same analysis was performed using genes within the FGF

pathway, there was no detectable overrepresentation of Rbpsuh binding sites (data not shown). Finally, when the Notch gene set was again analyzed by Clover for a large set of transcription factor binding motifs instead of Rbpsuh alone, the algorithm again identified sites of Rbpsuh enrichment (data not shown). Using MEME, we analyzed these same flanking regions for motif-overrepresentation **(Fig. 7)**. This analysis revealed an overrepresented motif in the flanking regions near *her4.2, her6, nrarpa,* and *notch3* **(Fig. 7B)** very similar to the Rbpsuh consensus binding sequence **(Fig. 7A)**. Again, the majority of these *cis*-elements associated with Rbpsuh binding sites did not correspond with NCEs.

Gene	H3K4me1 + Rbpsuh	H3K4me3 + Rbpsuh	NCE + Rbpsuh	me1 + me3 + Rbpsuh	me1 + NCE + Rbpsuh	me3 + NCE + Rbpsuh	me1 + me3 + NCE + Rbpsuh
nrarpa	—	1	1	2	—	—	—
nrarpb	—	4	—	—	1	—	—
her4.2	4	—	2	2	—	—	—
her6	—	—	2	1	—	—	—
hes5	—	—	1	—	—	—	—
notch2	2	—	—	—	—	—	—
notch3	—	1	2	—	1	—	—

Table 9: Frequency of overrepresented Rbpsuh binding sites near Notch signaling gene set and relation to epigenetic modifications and evolutionary conservation, p<0.001

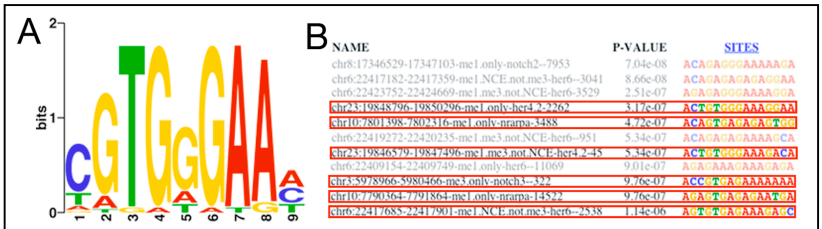

Figure 7: MEME analysis reveals overrepresentation of Rbpsuh consensus sequences. A – Logo of mouse Rbpsuh binding site. **B** – Putative *cis*-elements containing Rbpsuh binding sites identified using MEME algorithm.

We also compared our ChIP-Seq data to a set of known enhancers. A previous study identified several functional *cis*-elements near the *fgf8a* locus in zebrafish (38). Using the UCSC Genome Browser, we found that many of these confirmed enhancers are closely associated with areas of significant H3K4me1 enrichment **(Fig. 8, blue boxes)**. Interestingly, all six of these regions lie downstream of *fgf8a*.

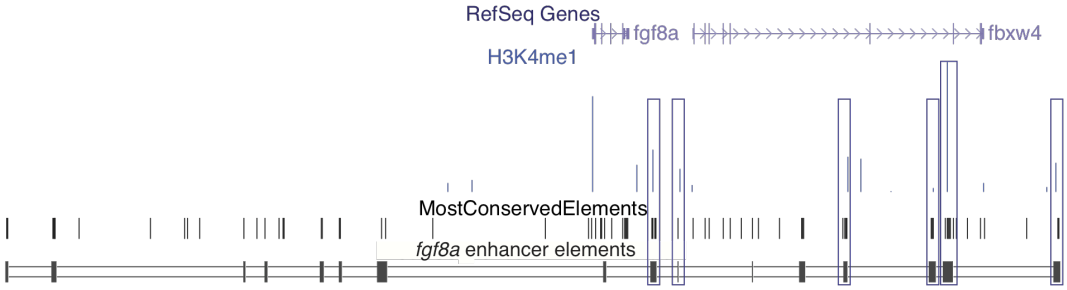

Figure 8: H3K4me1 is enriched at sites of known zebrafish enhancers. By comparing ChIP-Seq data from H3K4me1 in whole embryos to a dataset of confirmed zebrafish enhancers near *fgf8a*, we found six regions **(blue boxes)** of overlap.

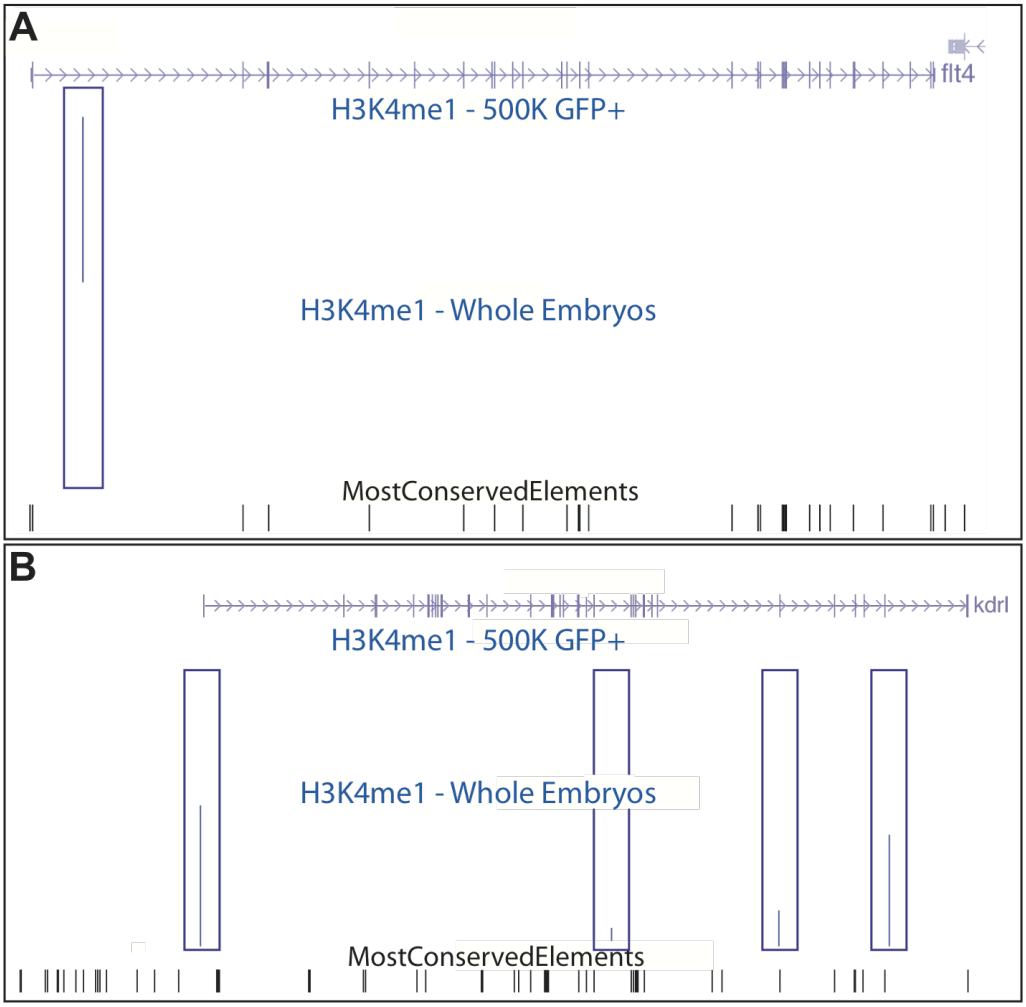

Figure 9: ChIP-Seq on an endothelial cell population reveals unique regions of epigenetic modifications. We compared H3K4me1 modifications from an endothelial cell population to the whole embryo at 24hpf. **A** – For the region surrounding *flt4*, the endothelial population was markedly enriched for H3K4me1 within the first intron **(blue box).** This enrichment was absent from the whole embryo analysis. **B** – For the region surrounding *kdrl*, the endothelial population lacked four regions of significant H3K4me1 enrichment seen in whole embryo analysis **(blue boxes).**

Next, we compared our H3K4me1 ChIP-Seq data from both whole embryos and *Tg(fli1:egfp)*^{y1} GFP+ cells **(Fig. 9)**. For *flt4* (VEGFR3), whole embryo analysis initially revealed no areas of significant H3K4me1 enrichment. By contrast, endothelial cells exhibited marked enrichment for this modification within the first intron of *flt4* **(Fig. 9A, blue box)**. On the other hand, analysis of *kdrl* in whole embryos initially showed four regions of significant H3K4me1 enrichment. The endothelial cell population, however, exhibited no such enrichment **(Fig. 9B, blue boxes)**.

Finally, we used our H3K4me1 ChIP-Seq data to identify candidate enhancers and validate their function *in vivo* using a reporter construct. For this analysis, we used our WT rather than our endothelial cell data because we had performed deeper sequencing on the former. Also, this dataset permitted us to include H3K4me3 enrichment. By using Gateway recombination-based technology, we cloned our candidate enhancers into expression clones driving expression of either mCherry or GFP **(Fig. 10)**. We initially chose seven candidate enhancers, three for *her6* and two each for *notch3* and *dll4*. One such ~1200bp candidate for *notch3* lay upstream of the TSS. It was characterized by H3K4me1 enrichment and mapped to an NCE as well as an Rbpsuh binding site **(Fig. 11A, blue box)**. After injecting this construct into embryos and assaying at ~28hpf, we found expression in the notochord **(Fig. 11C, arrow)** as well as the floor plate of the

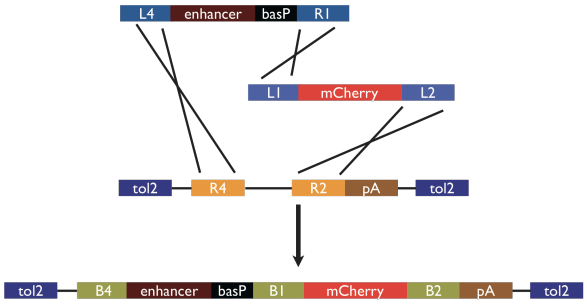

Figure 10: Gateway technology permits rapid synthesis of *in vivo* expression clones for enhancer validation. By using Gateway recombination-based technology, we were able to quickly clone putative enhancers upstream of a basal promoter driving the expression of a fluorescent protein, in this case mCherry. This construct also contains transposable elements for incorporation into the zebrafish genome. We injected these constructs into 1-cell WT zebrafish embryos and examined for expression of fluorescence at 24-48hpf.

neural tube **(Fig. 11C, arrowheads)**. This corresponds to early Notch activation in these same regions in a Notch-indicator transgenic zebrafish line (data not shown). Another candidate lay downstream of *her6* and mapped to a region of both H3K4me1 and H3K4me3 enrichment **(Fig. 11B)**. This 1200bp region also corresponded to two Rbpsuh binding sites **(Fig. 11B, blue box)**. At ~28hpf, this enhancer drove expression of mCherry in the fin fold **(Fig. 11D, arrowhead)** as well as in tissue posterior to the eye **(Fig. 11E, arrows)**. The other 5 candidates yielded no appreciable expression. Therefore, in total, we were able to validate 2/7 candidate enhancers.

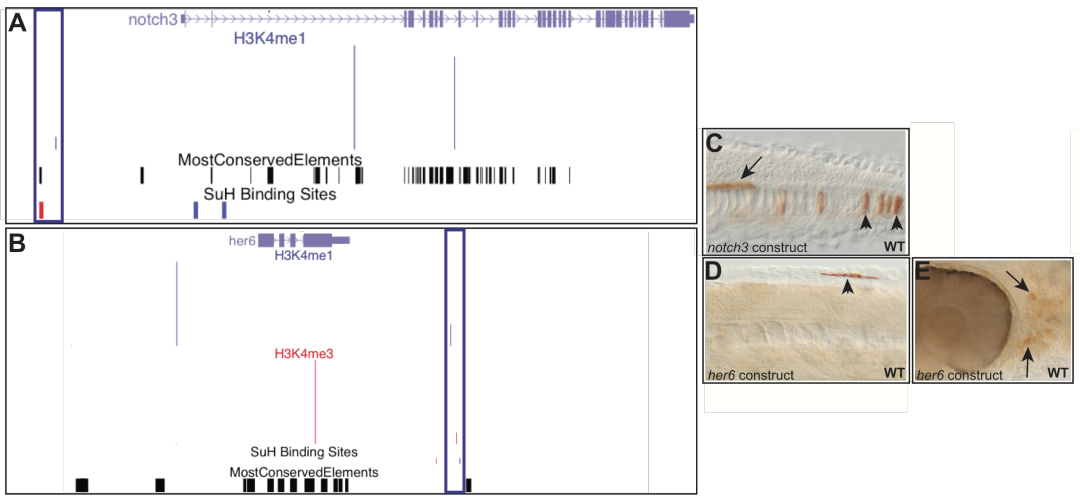

Figure 11: Enrichment of modified histones identifies functional enhancers. **A** – One candidate enhancer corresponds to a region of H3K4me1 enrichment upstream of the *notch3* TSS. This ~1200bp region **(boxed)** also maps to an NCE and an Rbpsuh binding site. **B** – The other candidate enhancer localizes to a region of both H3K4me1 and H3K4me3 enrichment downstream of *her6*, which is a Notch signaling target gene widely expressed in the CNS. This ~1200bp region **(boxed)** also contains two Rbpsuh binding sites. **C, D, & E** – After injecting our enhancer constructs, embryos were fixed at 28hpf and immunostained with a dsRed antibody, which amplified the fluorescent signal. **C** – Our *notch3* candidate enhancer yielded expression in the notochord **(arrowheads)** as well as the floor plate of the neural tube **(arrow)** in the tail. This corresponds to Notch activation in zebrafish embryos in both the notochord and neural tube. **D & E** – With our *her6* candidate enhancer, we see expression in the fin fold **(arrowhead)** and in tissue posterior to the eye **(arrows)**.

CONCLUSIONS

Few studies have used ChIP-Seq to study modified histones in vertebrates. Thus, we faced a technical hurdle in developing a viable ChIP-Seq protocol for whole embryos as well as extremely small cell populations. Some studies have used micrococcal nuclease digestion instead of sonication for chromatin shearing (51). There is even some evidence that micrococcal digestion yields a better signal to noise ratio than sonication (77). However, some studies using micrococcal nuclease (51) digest the chromatin prior to crosslinking, which may permit rearrangement of the nucleosomes during the endonuclease treatment (78). As shown in **Figure 2**, we found sonication to be an ideal means of quickly, cheaply, and reproducibly shearing chromatin into an appropriate size range. Other large centers also utilize mechanoacoustic methods of chromatin shearing in their standard protocols (46).

Given the expense and time required for next generation sequencing, it was critical that we confirm success of our ChIPs before proceeding. The concentrations displayed in **Table 4** indicate that we were successful in precipitating DNA fragments from whole embryos. Both our samples treated with antibody and our input samples were detectable by spectrophotometry, but our negative controls were not. Although the values were not drastically different, we consistently observed similar results for samples that successfully underwent subsequent deep sequencing, and our confidence in initially using spectrophotometry analysis was confirmed by the results in **Figure 3**. Because previous evidence indicated both H3K4me1 and H3K4me3 were enriched at promoter regions (51), we used promoter-specific primers for this PCR analysis. Amplification yielded strong gel bands for our ChIP and input samples. We expect input samples to

still contain these sequences, which the electrophoresis confirmed. Bands were nearly undetectable for our negative controls, thus confirming success of the ChIPs. Although this method is less precise than qPCR, the differences in band intensity were striking, and standard PCR requires less chromatin and is also less expensive.

For ChIP on small populations of cells, we determined effectiveness using qPCR, as the chromatin yield was much lower than with whole embryos. As shown in **Table 5**, we were able to successfully perform ChIP with $5x10^5$ cells. This is a significant feat, as many other ChIP experiments require as many as $5x10^7$ cells (48). Also, because we could use such a small number of cells, we were able to collect fewer clutches of embryos for our subsequent ChIP experiments. Of note, the enrichment for anti-H3K4me1 was much stronger than anti-H3K4me3, indicating this antibody was more effective with a smaller pool of epitopes. We next turned to ChIP on endothelial cells isolated by FACS from *Tg(fli1:egfp)*^{y1} embryos, and we compared enrichment of H3K4me1 (but not H3K4me3) at promoter regions from both GFP+ endothelial cells and GFP- cells **(Table 6)**. Enrichment at the promoter of *ef1α*, a housekeeping gene, was nearly identical in both populations. This was expected, since nearly all cells should be utilizing this gene and controlling its associated epigenetic marks in similar ways. However, we saw differences for *fli1b* and *aqp8*, indicating that these different cell populations control epigenetic marks at these loci differently.

Another major concern was avoiding overamplification of our adapter-ligated library, as this would increase the likelihood of PCR artifacts (67). As shown in **Figure 4**, we performed several trial runs of PCR amplification. In examining the gel, we looked for the minimum number of PCR cycles that yielded an easily detectable smear of

fragments, in this case 18 cycles. Using gel electrophoresis also allowed us to visualize self-ligated adapters and eliminate them during gel extraction. One must take care to avoid these sequences, because they will otherwise be included in cluster amplification and deep sequencing (46).

After compiling our ChIP-Seq data, we used the UCSC genome browser to explore the distribution of modified histone enrichment near genes of interests. In **Figure 5**, one can see how closely H3K4me1 enrichment corresponds to NCEs and Rbpsuh binding sites near *notch3*, and this occurs both upstream of the TSS and within introns. Also, H3K4me3 is enriched at the TSS. Although this visual snapshot only highlights one gene, it clearly shows the wealth of information such analysis can provide.

In order to understand our ChIP-Seq findings throughout the genome, we compiled distribution maps for H3K4me1 and H3K4me3 enrichment relative to all TSSs **(Fig. 6)**. Consistent with other studies (51-53), we found H3K4me3 to be sharply enriched near TSSs. As in other studies (51), H3K4me1 is also enriched at TSSs, although not to as great a degree. We also see lower level H3K4me1 enrichment up and downstream of TSSs, which may represent an association between this mark and *cis*-elements, which other studies have also found (51; 53; 58; 59). Next, we further characterized TSS flanking regions **(Table 7)** and found that the majority lacked H3K4me1 and H3K4me3 enrichment. This finding suggests these marks are not widely distributed around promoter regions, and cells are tightly controlling their placement. There are many more instances of H3K4me3 enrichment associated with these sequences than H3K4me1. Interestingly, we found 211 regions characterized by both H3K4me1 and H3K4me3 enrichment as well as evolutionary conservation. Such a high degree of

epigenetic regulation in these highly conserved areas may indicate these sequences are particularly important in the organism.

We expanded this same analysis to include the entire genome instead of just regions surrounding TSSs **(Table 8)**. By doing so, we located 10-fold more regions of H3K4me1 enrichment versus H3K4me3, which supports previous findings of H3K4me1 enrichment primarily at *cis*-elements (51; 53; 58; 59). Again, we found more than 4,000 highly conserved regions of H3K4me1 and H3K4me3 enrichment. Such regions may represent *cis*-elements with more significant or widespread roles than those marked by only one type of modified histone.

In order to determine if our ChIP-Seq data was able to locate known transcription binding sites, we used the Clover algorithm to examine genes in the Notch pathway **(Table 9)**. Upon Notch activation, Rbpsuh functions with other proteins to drive target gene expression (79). Therefore, we would expect this Notch gene dataset to contain Rbpsuh binding sites, and these sites should correspond to epigenetic modification. Clover found Rbpsuh binding sites enriched for various combinations of modified histones near five of the Notch genes examined, which indicates the power of our experimental approach in identifying *cis*-elements. The majority of these sites were not associated with NCEs, suggesting histone modification is a better predictor of *cis*-elements. When we performed the same analysis on genes in the FGF pathway, Clover did not identify any Rbpsuh sites, which is what we expected since that protein does not have a known role in FGF signaling. Additionally, we searched for multiple transcription factor binding sites using this same analysis on both Notch and FGF genes. We included this analysis as a way to avoid biasing toward Rbpsuh. In this case, we again found

regions of modified histone enrichment mapping to Rbpsuh binding sites in the Notch gene set but not with FGF, further supporting our findings. Similarly, we subjected these same sequences to MEME analysis **(Fig. 7**), which also identified several regions that closely resembled Rbpsuh binding sites.

Despite the results of our Clover and MEME analyses, an association between histone modifications and transcription factor binding sites does not necessarily indicate proteins actually occupy these sites (26). Therefore, we compared our data to a set of known enhancers **(Fig. 8)**. We found six regions of H3K4me1 enrichment corresponding to known enhancers for *fgf8a*, indicating our ChIP-Seq approach is successful at identifying functional enhancer elements. Although other known enhancers from this dataset are not marked by enrichment, it is possible that deeper sequencing of our sample may yield enrichment at these areas.

Since we are primarily interested in vascular development, we wanted to determine if epigenetic marks from endothelial cells alone would provide more pertinent information than ChIP-Seq on whole embryos. To do this, we performed ChIP-Seq on endothelial cells harvested from *Tg(fli1:egfp)*^{y1} embryos and compared the results to our whole embryo analysis **(Fig. 9)**. In some cases, the endothelial cells exhibited H3K4me1 enrichment that was absent in whole embryos **(Fig. 9A)**. This may occur because the signal is only present in a small population of cells and is diluted in whole embryo samples. In other cases, enrichment present in whole embryos was absent from endothelial cells **(Fig. 9B)**. Sequences at these loci may not be functioning as promoters or *cis*-elements in the endothelial cell population. Using endothelial-specific epigenetic profiles may be a more reliable approach for identifying candidate *cis*-elements, and in

the future, we plan to perform deeper sequencing of these samples in order to obtain more sequence tags and increase our ability to accurately predict *cis*-elements functioning in the zebrafish vasculature.

Several previous studies examining the relationship between modified histones and enhancers relied exclusively on predicted rather than functional enhancers (51; 53; 58; 59). Those that do functionally validate predicted enhancers rely upon ChIP-Seq data using enhancer binding proteins or CNEs, not epigenetic modification (37; 60). We felt functionally validating candidate enhancers identified from our ChIP-Seq data was essential, and our approach using modified histones represented a more global survey than one relying on known enhancer binding proteins. Using H3K4me1 enrichment, we assayed seven candidate enhancers near three genes in the Notch pathway: *her6*, *notch3*, and *dll4* **(Fig. 11)**. Although these genes are expressed in the vasculature (data not shown), the candidate enhancers drove reporter gene expression elsewhere. One *notch3* enhancer drove expression in the notochord and neural tube **(Fig. 11C)**, while a *her6* enhancer drove expression in the fin fold **(Fig. 11D)** and in tissue posterior to the eye **(Fig. 11E)**. It is possible these enhancers normally function to facilitate expression of distant genes instead of those in close proximity, as enhancers are known to function at distances of up to one megabase (33). Of the seven candidates tested, we were only able to validate two. Previous studies using CNEs and P300 binding to predict enhancers yielded much higher success rates (37; 60). Our success rate may improve if we tested more candidates and examined for expression at different timepoints. Also, using regions of H3K4me1 enrichment specifically from our endothelial cell dataset may prove a better means of identifying enhancers for vascular gene expression. In addition, it is possible

some of these candidates are in fact insulators or repressors. H3K4me1 is associated with insulator elements (51), and there are also instances of long-range repression via *cis*-elements in zebrafish and other organisms (32; 80). We plan to design constructs to test for insulator and repressor function, and results from these experiments may reveal an additional use of our dataset for predicting *cis*-elements.

By using ChIP-Seq technology, we have created the first map of H3K4me1 and H3K4me3 enrichment in WT zebrafish embryos as well as endothelial cells collected from transgenic zebrafish. In addition, we have characterized the relation of these epigenetic marks to each other as well as CNEs and TSSs. Predicted enhancers from our data map to transcription factor binding sites as well as to confirmed zebrafish enhancers. Finally, we ourselves functionally assayed several predicted enhancers and confirmed their ability to drive target gene expression. We believe our approach represents a more thorough method of predicting and screening candidate *cis*-regulatory elements in the zebrafish genome. Our method avoids the limited scope of experiments using known enhancer-binding proteins, and it is less speculative than those relying simply on predicted protein binding sites or evolutionary conservation. Through further analysis, our data may also reveal a role in predicting insulator and repressor elements. By making our data publicly available, we will have created an invaluable resource for numerous researchers. Also, it will permit us to further explore the roles of specific *cis*-regulatory elements in vascular development as well as the role of modified histones in embryonic development. This is part of a new wave of research focusing on epigenetics and technology that permits genome-wide analyses, and it represents a powerful tool for ultimately studying human disease.

REFERENCES

1. Boulpaep, E. (2002). The Cardiovascular System. Medical Physiology. Walter Boron and Emile Boulpaep, eds. Philadelphia, W.B. Saunders: 463-465.

2. Stenman, J.M., Rajagopal, J., Carroll, T.J., Ishibashi, M., McMahon, J.*, et al.* 2008. Canonical Wnt signaling regulates organ-specific assembly and differentiation of CNS vasculature. *Science.* 322(5905): 1247-1250.

3. LeCouter, J., Kowalski, J., Foster, J., Hass, P., Zhang, Z.*, et al.* 2001. Identification of an angiogenic mitogen selective for endocrine gland endothelium. *Nature.* 412(6850): 877-884.

4. Childs, S., Chen, J.-N., Garrity, D.M. and Fishman, M.C. 2002. Patterning of angiogenesis in the zebrafish embryo. *Development.* 129(4): 973-982.

5. Siekmann, A.F. and Lawson, N.D. 2007. Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature.* 445(7129): 781-784.

6. Gerhardt, H., Golding, M., Fruttiger, M., Ruhrberg, C., Lundkvist, A.*, et al.* 2003. VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J Cell Biol.* 161(6): 1163-1177.

7. Hellström, M., Phng, L.-K., Hofmann, J.J., Wallgard, E., Coultas, L.*, et al.* 2007. Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature.* 445(7129): 776-780.

8. Siekmann, A.F., Standley, C., Fogarty, K.E., Wolfe, S.A. and Lawson, N.D. 2009. Chemokine signaling guides regional patterning of the first embryonic artery. *Genes Dev.* 23(19): 2272-2277.

9. Lawson, N.D. and Weinstein, B.M. 2002. Arteries and veins: making a difference with zebrafish. *Nat Rev Genet.* 3(9): 674-682.

10. Krebs, L.T., Xue, Y., Norton, C.R., Shutter, J.R., Maguire, M.*, et al.* 2000. Notch signaling is essential for vascular morphogenesis in mice. *Genes & Development.* 14(11): 1343-1352.

11. Xue, Y., Gao, X., Lindsell, C.E., Norton, C.R., Chang, B.*, et al.* 1999. Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Human Molecular Genetics.* 8(5): 723-730.

12. Uyttendaele, H., Ho, J., Rossant, J. and Kitajewski, J. 2001. Vascular patterning defects associated with expression of activated Notch4 in embryonic endothelium. *Proc Natl Acad Sci USA.* 98(10): 5643-5648.

13. Lawson, N.D., Mugford, J.W., Diamond, B.A. and Weinstein, B.M. 2003. phospholipase C gamma-1 is required downstream of vascular endothelial growth factor during arterial development. *Genes & Development.* 17(11): 1346-1351.

14. Lawson, N.D., Scheer, N., Pham, V.N., Kim, C.H., Chitnis, A.B.*, et al.* 2001. Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development.* 128(19): 3675-3683.

15. Lawson, N.D., Vogel, A.M. and Weinstein, B.M. 2002. sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation. *Dev Cell.* 3(1): 127-136.

16. Zhong, T.P., Rosenberg, M., Mohideen, M.A., Weinstein, B. and Fishman, M.C. 2000. gridlock, an HLH gene required for assembly of the aorta in zebrafish. *Science.* 287(5459): 1820-1824.

17. Zhong, T.P., Childs, S., Leu, J.P. and Fishman, M.C. 2001. Gridlock signalling pathway fashions the first embryonic artery. *Nature.* 414(6860): 216-220.

18. Nakagawa, O., McFadden, D.G., Nakagawa, M., Yanagisawa, H., Hu, T.*, et al.* 2000. Members of the HRT family of basic helix-loop-helix proteins act as transcriptional repressors downstream of Notch signaling. *Proc Natl Acad Sci USA.* 97(25): 13655-13660.

19. Seo, S., Fujita, H., Nakano, A., Kang, M., Duarte, A.*, et al.* 2006. The forkhead transcription factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. *Developmental Biology.* 294(2): 458-470.

20. Hayashi, H. and Kume, T. 2008. Foxc transcription factors directly regulate Dll4 and Hey2 expression by interacting with the VEGF-Notch signaling pathways in endothelial cells. *PLoS ONE.* 3(6): e2401.

21. Pendeville, H., Winandy, M., Manfroid, I., Nivelles, O., Motte, P.*, et al.* 2008. Zebrafish Sox7 and Sox18 function together to control arterial-venous identity. *Dev Biol.* 317(2): 405-416.

22. Cermenati, S., Moleri, S., Cimbro, S., Corti, P., Del Giacco, L.*, et al.* 2008. Sox18 and Sox7 play redundant roles in vascular development. *Blood.* 111(5): 2657-2666.

23. François, M., Caprini, A., Hosking, B., Orsenigo, F., Wilhelm, D.*, et al.* 2008. Sox18 induces development of the lymphatic vasculature in mice. *Nature.* 456(7222): 643-647.

24. Sumanas, S. and Lin, S. 2006. Ets1-related protein is a key regulator of vasculogenesis in zebrafish. *PLoS Biol.* 4(1): e10.

25. Deval, S., Chi, N., Meadows, S., Minovitsky, S., Anderson, J.*, et al.* 2008. Combinatorial Regulation of Endothelial Gene Expression by Ets and Forkhead Transcription Factors. *Cell.* 135(6): 1053-1064.

26. Farnham, P.J. 2009. Insights from genomic profiling of transcription factors. *Nat Rev Genet.* 10(9): 605-616.

27. Cande, J.D., Chopra, V.S. and Levine, M. 2009. Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, Tribolium castaneum. *Development.* 136(18): 3153-3160.

28. Kwon, D., Mucci, D., Langlais, K.K., Americo, J.L., Devido, S.K.*, et al.* 2009. Enhancer-promoter communication at the Drosophila engrailed locus. *Development.* 136(18): 3067-3075.

29. Gaszner, M. and Felsenfeld, G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet.* 7(9): 703-713.

30. Fujioka, M., Wu, X. and Jaynes, J.B. 2009. A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication. *Development.* 136(18): 3077-3087.

31. Choi, J., Dong, L., Ahn, J., Dao, D., Hammerschmidt, M.*, et al.* 2007. FoxH1 negatively modulates flk1 gene expression and vascular formation in zebrafish. *Dev Biol.* 304(2): 735-744.

32. Visel, A., Rubin, E.M. and Pennacchio, L.A. 2009. Genomic views of distant-acting enhancers. *Nature.* 461(7261): 199-205.

33. Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P.*, et al.* 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics.* 12(14): 1725-1735.

34. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. and Shiroishi, T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development.* 132(4): 797-803.

35. Emison, E.S., McCallion, A.S., Kashuk, C.S., Bush, R.T., Grice, E.*, et al.* 2005. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature.* 434(7035): 857-863.

36. Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M. and Furlong, E.E.M. 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature.* 461(7269): 65-70.

37. Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H.*, et al.* 2010. A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Dev Biol.* 337(2): 484-495.

38. Komisarczuk, A.Z., Kawakami, K. and Becker, T.S. 2009. Cis-regulation and chromosomal rearrangement of the fgf8 locus after the teleost/tetrapod split. *Developmental Biology.* 336(2): 301-312.

39. Kantorovitz, M.R., Kazemian, M., Kinston, S., Miranda-Saavedra, D., Zhu, Q.*, et al.* 2009. Motif-blind, genome-wide discovery of cis-regulatory modules in Drosophila and mouse. *Dev Cell.* 17(4): 568-579.

40. Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V.*, et al.* 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* 5(9): e234.

41. Li, B., Carey, M. and Workman, J.L. 2007. The role of chromatin during transcription. *Cell.* 128(4): 707-719.

42. Kim, J.K., Samaranayake, M. and Pradhan, S. 2008. Epigenetic mechanisms in mammals. *Cell. Mol. Life Sci.*: 1-17.

43. Cedar, H. and Bergman, Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet.* 10(5): 295-304.

44. Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J.*, et al.* 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 456(7218): 53-59.

45. Mardis, E.R. 2007. ChIP-seq: welcome to the new frontier. *Nat Meth.* 4(8): 613-614.

46. Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J.*, et al.* 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Meth.* 5(12): 1005-1010.

47. Park, P.J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet.* 10(10): 669-680.

48. Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J.*, et al.* 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods.* 48(3): 240-248.

49. Kharchenko, P.V., Tolstorukov, M.Y. and Park, P.J. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol.* 26(12): 1351-1359.

50. Park, P. 2008. Epigenetics meets next-generation sequencing. *Epigenetics : official journal of the DNA Methylation Society.* 3(6): 318-321.

51. Barski, A., Cuddapah, S., Cui, K., Roh, T., Schones, D.*, et al.* 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell.* 129(4): 823-837.

52. Akkers, R.C., Heeringen, S.J.v., Jacobi, U.G., Janssen-Megens, E.M., Francoijs, K.-J.*, et al.* 2009. A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos. *Developmental Cell.* 17(3): 425-434.

53. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A.*, et al.* 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature.* 459(7243): 108-112.

54. Martin, C. and Zhang, Y. 2005. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol.* 6(11): 838-849.

55. Yu, H., Zhu, S., Zhou, B., Xue, H. and Han, J.-D.J. 2008. Inferring causal relationships among different histone modifications and gene expression. *Genome Research.* 18(8): 1314-1324.

56. Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., Desalle, R.*, et al.* 2009. Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics.* 10(143).

57. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A.*, et al.* 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet.* 40(7): 897-903.

58. Robertson, A.G., Bilenky, M., Tam, A., Zhao, Y., Zeng, T.*, et al.* 2008. Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Research.* 18(12): 1906-1917.

59. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W.*, et al.* 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 39(3): 311-318.

60. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A.*, et al.* 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 457(7231): 854-858.

61. Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D.*, et al.* 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet.* 40(2): 158-160.

62. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. and Schilling, T.F. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn.* 203(3): 253-310.

63. Lawson, N.D. and Weinstein, B.M. 2002. In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev Biol.* 248(2): 307-318.

64. Villefranc, J.A., Amigo, J. and Lawson, N.D. 2007. Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev Dyn.* 236(11): 3077-3087.

65. Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). Eugene, Univ. of Oregon Press.

66. Bookout, A.L., Cummins, C.L., Mangelsdorf, D.J., Pesola, J.M., and Kramer, M.F. (2006). High-Throughput Real-Time Quantitiative Reverse Transcription PCR. Current Protocols in Molecular Biology. **73:** 15.18.11-15.18.28.

67. Cha, R.S. and Thilly, W.G. 1993. Specificity, efficiency, and fidelity of PCR. *PCR Methods Appl.* 3(3): S18-29.

68. Kent, W.J. 2002. BLAT--the BLAST-like alignment tool. *Genome Research.* 12(4): 656-664.

69. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S.*, et al.* 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9(9): R137.

70. Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological).* 57(1): 289-300.

71. Frith, M.C., Fu, Y., Yu, L., Chen, J.-F., Hansen, U.*, et al.* 2004. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.* 32(4): 1372-1381.

72. Fortini, M.E. 2009. Notch signaling: the core pathway and its posttranslational regulation. *Dev Cell.* 16(5): 633-647.

73. Bailey, T.L. and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 2: 28-36.

74. Kawakami, K. 2007. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol.* 8(Suppl 1): S7.

75. Xu, Q. (1999). Microinjection into Zebrafish Embryos. Molecular Methods in Developmental Biology. M. Guille, Humana Press. **127:** 125-132.

76. Gilchrist, D.A., Fargo, D.C. and Adelman, K. 2009. Using ChIP-chip and ChIP-seq to study the regulation of gene expression: Genome-wide localization studies reveal widespread regulation of transcription elongation. *Methods.* 48(4): 398-408.

77. Huang, H.-S., Matevossian, A., Jiang, Y. and Akbarian, S. 2006. Chromatin immunoprecipitation in postmortem brain. *J Neurosci Methods.* 156(1-2): 284-292.

78. O'Neill, L.P. and Turner, B.M. 2003. Immunoprecipitation of native chromatin: NChIP. *Methods.* 31(1): 76-82.

79. Adams, R.H. and Alitalo, K. 2007. Molecular regulation of angiogenesis and lymphangiogenesis. *Nat Rev Mol Cell Biol.* 8(6): 464-478.

80. Jessen, J.R., Willett, C.E. and Lin, S. 1999. Artificial chromosome transgenesis reveals long-distance negative regulation of rag1 in zebrafish. *Nat Genet.* 23(1): 15-16.