

Identification of Vascular and Hematopoietic Genes Downstream of *etsrp* by Deep Sequencing in Zebrafish

Gustavo Gomez¹, Jae-Hyung Lee², Matthew B. Veldman¹, Jing Lu¹, Xinshu Xiao², Shuo Lin^{1*}

¹ Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, California, United States of America, ² Department of Integrative Biology and Physiology, University of California Los Angeles, Los Angeles, California, United States of America

Abstract

The transcription factor *etsrp/Er71/Etv2* is a master control gene for vasculogenesis in all species studied to date. It is also required for hematopoiesis in zebrafish and mice. Several novel genes expressed in vasculature have been identified through transcriptional profiling of zebrafish embryos overexpressing *etsrp* by microarrays. Here we re-examined this transcriptional profile by Illumina RNA-sequencing technology, revealing a substantially increased number of candidate genes regulated by *etsrp*. Expression studies of 50 selected candidate genes from this dataset resulted in the identification of 39 new genes that are expressed in vascular cells. Regulation of these genes by *etsrp* was confirmed by their ectopic induction in *etsrp* overexpressing and decreased expression in *etsrp* deficient embryos. Our studies demonstrate the effectiveness of the RNA-sequencing technology to identify biologically relevant genes in zebrafish and produced a comprehensive profile of genes previously unexplored in vascular endothelial cell biology.

Citation: Gomez G, Lee J-H, Veldman MB, Lu J, Xiao X, et al. (2012) Identification of Vascular and Hematopoietic Genes Downstream of *etsrp* by Deep Sequencing in Zebrafish. PLoS ONE 7(3): e31658. doi:10.1371/journal.pone.0031658

Editor: Henry H. Roehl, University of Sheffield, United Kingdom

Received: October 28, 2011; **Accepted:** January 17, 2012; **Published:** March 16, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This work was supported by grants from the National Institutes of Health (NIH) (F31HL091713 to GAG and DK54508 to SL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: shuolin@ucla.edu

Introduction

The cardiovascular system, which includes the heart, vessels and blood, function together to deliver oxygen and nutrients to cells throughout the body and remove metabolic waste. Understanding the development of this system is instrumental to the advancement of both basic and clinical sciences. The zebrafish, *Danio rerio*, is an excellent model organism for such studies due to embryo transparency, high fecundity, and fast development of organogenesis. In particular, the cardiovascular system is formed within one day of birth [1,2]. Through genetic, cellular and molecular studies in zebrafish, a great deal of knowledge regarding the molecular components and cellular events that establish this system has been obtained. It is notable that many of the key molecular players and events that drive organogenesis in zebrafish are evolutionarily and functionally conserved with other organisms, including mammals.

A critical transcription factor required for the development of vascular endothelial cells is *ets related protein*, *etsrp*, which was originally discovered through transcriptional profiling of the cardiovascular mutant line, *cloche*, that lacks blood and vasculature [3] and independently by a mutagenesis screen [4]. It was subsequently found that *etsrp* is both necessary and sufficient to induce both the vascular endothelial and primitive myelopoietic program in zebrafish [5,6]. Its significance was further underscored by the identification and characterization of the functional homolog *ets variant 2*, *etv2*, in mammals [7,8] and *Xenopus* [9,10]. The *etsrp* transgenic fish line has been useful to reveal the cellular events that establish the cranial vasculature and myelopoiesis [11], and it is possible that *etsrp* might be associated with the initiation of the definitive hematopoietic program [12].

The zebrafish genome is now in its ninth version and remains incompletely annotated. Nonetheless bioinformatic approaches have been used to identify important genes encoding transcription factors containing the ETS box DNA binding domain in hematopoietic and endothelial development [13]. The use of microarrays by the zebrafish research community has increased the identification of genes expressed in the developing cardiovascular system [3,14,15,16,17,18,19]. The arrival of higher throughput next generation sequencing has expanded the possibilities to deepen our understanding and identification of novel genes, and has already proven its utility for studies in zebrafish. Although the majority of the published studies combining RNA-sequencing and zebrafish have focused on immunity [20,21,22], the use of high throughput sequencing to study other biological subjects in zebrafish has begun to increase [23,24,25]. The ability of this technological approach to examine the zebrafish transcriptome at greater depth than microarrays without an *a priori* bias prompted us to re-examine the transcriptional profile of embryos overexpressing *etsrp*. In this study we present a panel of 39 more genes that are expressed in the developing zebrafish vasculature that were identified by this approach.

Results

As previously demonstrated, the overexpression of *etsrp* results in the induction of vascular related genes during gastrulation stages of development, before the onset of angioblast specification [5]. In this study *etsrp* RNA was injected into one cell *flkl1-gfp* transgenic embryos and their transcriptome profiles were examined at later stages of gastrulation, 70–90% epiboly, when ectopic induction of *flkl1-gfp* is detected (Figure 1). Total RNA was extracted from pools

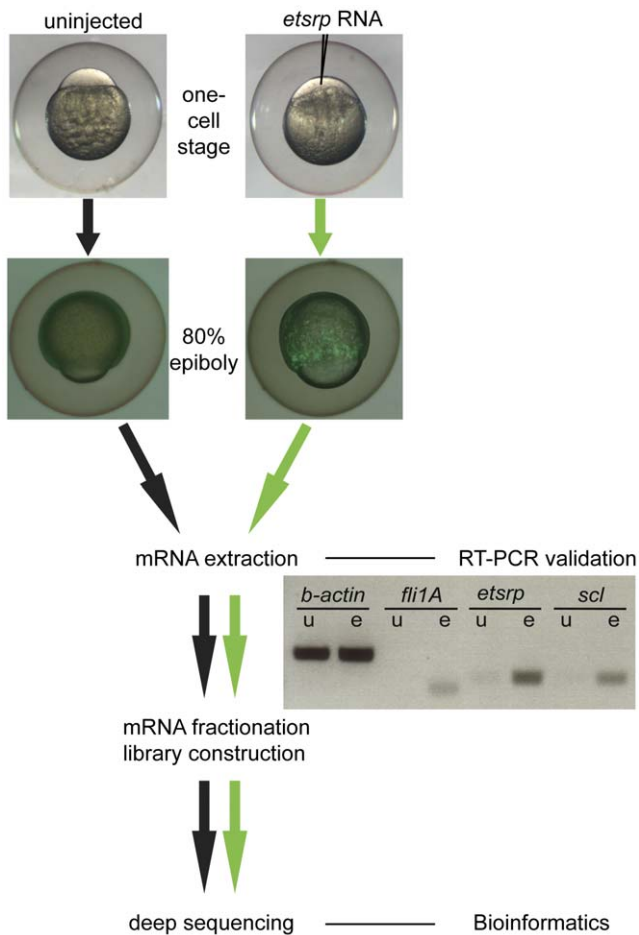


Figure 1. Workflow diagram. Embryos were either uninjected or injected with *etsrp* RNA at the one cell stage, then raised until the late gastrulation stages when the *flk1-gfp* transgenic reporter is induced ectopically. Equal paired groups of embryos were pooled for mRNA extraction, library construction, and solexa mRNA-sequencing. Samples were pre-validated prior to sequencing by RT-PCR with *fli1A*, the 3'UTR of *etsrp*, as well as *scl*, which were all preferentially induced as expected by *etsrp* overexpression (e) relative to uninjected controls (u). doi:10.1371/journal.pone.0031658.g001

of 125 embryos, and the induction of known hemangioblast markers, *fli1a*, *scl* and *etsrp*, was verified in the *etsrp* overexpressed group by RT-PCR before library construction and sequencing. Two separate sets of RNA-seq libraries were constructed from independent samples that were pre-validated using the same validation scheme as demonstrated in the workflow diagram on Figure 1.

A total of 123 million single-end or paired-end reads were obtained and 35% of the reads were mapped onto the unigenes transcript sequence database (Build #117) [26] using the Burrows Wheeler Aligner (BWA) alignment program [27]. Table S1 lists the number of reads in each control and *etsrp* overexpressed (*etsrp* oe) samples and the mapping results. About 43 million reads were uniquely mapped in total, which covered 77% of the unigenes database (39,784 out of 51,481). The expression level difference between the control and *etsrp* oe samples were compared, resulting in the upregulation of 849 unigenes entries above a 1.9-fold cutoff value in the *etsrp* oe group with a corrected p-value < 0.05, while 726 entries were downregulated using the same parameters. A sample of the top 35 hits of annotated upregulated genes obtained in this dataset is listed in Table 1.

In order to determine the biological relevance of the gene pool obtained here, a gene ontology analysis was performed with the Database for Annotation, Visualization and Integrated Discovery Gene Ontology (DAVID GO) program, as demonstrated in the identification of novel genes expressed in myeloid cells in zebrafish [28]. Of the 849 upregulated unigenes hits examined by functional annotation clustering, 797 were queried successfully, but because some unigenes entries are associated with more than one gene, and some genes are represented by more than one unigenes identifier, a total of 787 entries were evaluated resulting in 148 clusters. The remaining 52 entries were not identified either because they represent expressed sequence tags not associated with a characterized gene, they are unique to zebrafish, or their orthologs were unidentifiable by DAVID GO gene conversion tools. Conversely, the 726 downregulated hits resulted in the successful query of 306 entries divided among 50 clusters, but relative to the induced clusters, these had rather low fold-enrichment scores.

The top two clusters from the upregulated data contained genes associated with "vascular development", "angiogenesis," and other terms related to vascular endothelial cell biology. These data were combined resulting in 68 genes after duplicate deletion, and examined individually for involvement in vascular development. 18 genes were excluded because although several are expressed in mesoderm or heart, their specific expression or function within endothelial cells remains to be demonstrated. We did examine one of these, *xirp2a*, which we found to have endothelial expression. These were noted as a subset of potential vascular genes in Table S2, which also includes 12 that were removed because there is no indication that they are associated with endothelial cells. By searching the list generated containing 849 unigenes entries, 15 more vascular related genes were identified (Table 2), bringing the total number of genes associated with endothelial cells in this dataset to 53.

In addition to endothelial specific genes, there were 4 genes of the myeloid lineage that were preferentially induced by *etsrp* oe. They are *lplastin*, *csf3r*, *ceb1*, *hsd3b7*. This observation is consistent with what has been previously reported by our group [6]. Overall using the DAVID GO program and visual inspection of the RNA-seq dataset we observed that a substantial number of genes (6.2%) are associated with developing endothelial biology, confirming the biological relevance of the upregulated dataset. Contrarily, DAVID GO analysis of the downregulated genes did not result in a distinct biological category of genes, as genes associated with the development of all three germ layers including endoderm, mesoderm, and ectoderm were repressed.

In order to further biologically validate the data and identify new genes expressed in endothelial cells, several genes were selected from the upregulated dataset for analysis by RNA whole mount in situ hybridization (WISH) at the 80% epiboly stages, when they are predicted to be upregulated by overexpressing *etsrp* (Figure 2). For gene selection, the potential of the top 350 hits to be expressed in endothelial cells was evaluated thoroughly, while the remaining portion was assessed at random. Genes examined were selected based on a PUBMED literature search using the following criteria: 1. Not previously shown to be expressed in vascular endothelial cells. 2. Previously detected in endothelial cells but evidence is limited to *in vitro* studies. 3. Has not been examined in developing vasculature. One of the top ranked genes (highest fold-induction), *similar to Src homology 2 domain containing E*, *she*, was used as a positive control since its expression in the vasculature has been reported previously [19]. The gene set examined is listed on Table 3. This assay resulted in the clear ectopic induction by *etsrp* overexpression in 49 of the 50 genes examined (Figure 2). The high percentage of validated genes induced by *etsrp* overexpression

Table 1. Top 35 annotated genes induced by *etsrp*.

Unigene ID	Gene ID	Gene Description	Fold Change	p-value
Dr.24158	<i>srgn</i>	<i>serglycin</i>	1920.8	7.7E-53
Dr.153782	<i>LOC100151241</i>	<i>similar to loc559821 protein</i>	1899.9	4.5E-26
Dr.83594	<i>crabp1a</i>	<i>cellular retinoic acid binding protein 1a</i>	917.4	5.1E-17
Dr.141228	<i>LOC558783</i>	<i>similar to cell death activator cide-a</i>	452.8	1.1E-05
Dr.149118	<i>LOC798186</i>	<i>similar to cdc42 gtpase-activating protein</i>	441.5	1.9E-21
Dr.118570	<i>LOC561493</i>	<i>similar to src homology 2 domain containing e</i>	315.3	1.8E-196
Dr.40434	<i>ms4a17a.11</i>	<i>membrane-spanning 4-domains, subfamily a, member 17a.11</i>	300.4	3.5E-06
Dr.133475	<i>LOC100151130</i>	<i>similar to rasip1 protein</i>	254.8	3.6E-10
Dr.37870	<i>ckmb</i>	<i>creatine kinase, muscle b</i>	231.6	3.8E-07
Dr.92232	<i>efcab2</i>	<i>ef-hand calcium binding domain 2</i>	231.5	8.4E-03
Dr.84343	<i>zgc:153721</i>	<i>zgc:153721</i>	225.9	1.1E-09
Dr.85502	<i>mrc1</i>	<i>mannose receptor c1-like protein</i>	225.7	4.0E-20
Dr.47591	<i>etv2</i>	<i>ets variant gene 2</i>	219.0	4.9E-237
Dr.110840	<i>LOC100006361</i>	<i>similar to preprogalanin 1b</i>	217.2	4.9E-03
Dr.103328	<i>zgc:162298</i>	<i>zgc:162298</i>	198.7	2.0E-05
Dr.86665	<i>nhlh2</i>	<i>nescient helix loop helix 2</i>	180.6	2.1E-08
Dr.114377	<i>zgc:173594</i>	<i>zgc:173594</i>	175.7	2.8E-03
Dr.110713	<i>LOC100149611</i>	<i>similar to loc495463 protein</i>	165.4	3.8E-08
Dr.84866	<i>LOC100005159</i>	<i>similar to hemicentin 1</i>	159.4	9.6E-04
Dr.111220	<i>gpr183</i>	<i>g protein-coupled receptor 183</i>	151.4	1.9E-04
Dr.113631	<i>LOC568486</i>	<i>similar to regulating synaptic membrane exocytosis 1</i>	136.1	2.5E-02
Dr.82184	<i>LOC569386</i>	<i>platelet endothelial cell adhesion molecule-like</i>	123.4	6.2E-06
Dr.124732	<i>LOC100148619</i>	<i>similar to myelin associated glycoprotein</i>	113.9	1.4E-02
Dr.152900	<i>LOC572378</i>	<i>similar to transforming growth factor-b type ii receptor</i>	110.1	4.9E-03
Dr.142587	<i>csf3r</i>	<i>colony stimulating factor 3 receptor (granulocyte)</i>	104.9	6.5E-05
Dr.100658	<i>LOC100004843</i>	<i>hypothetical loc100004843</i>	104.1	2.8E-03
Dr.149641	<i>LOC558126</i>	<i>similar to hcg2040171</i>	103.1	4.9E-03
Dr.151919	<i>si:ch211-10e8.4</i>	<i>si:ch211-10e8.4</i>	102.9	8.4E-03
Dr.81739	<i>kcnh2</i>	<i>potassium voltage-gated channel, subfamily h, member 2</i>	101.5	2.0E-09
Dr.78041	<i>xirp2</i>	<i>xin actin-binding repeat containing 2</i>	95.0	7.6E-56
Dr.118849	<i>zgc:63958</i>	<i>zgc:63958</i>	91.4	1.7E-03
Dr.102595	<i>LOC568153</i>	<i>similar to chemokine receptor-like 1</i>	86.4	2.8E-03
Dr.115871	<i>LOC569038</i>	<i>similar to cholecystokinin a receptor</i>	85.0	2.5E-02
Dr.83127	<i>LOC100001838</i>	<i>similar to loc100001772 protein</i>	79.5	1.4E-02

Several unigene hits identified in the top list correspond to unannotated expressed sequence tags, est's, and these were removed from this table. The dataset containing the data in its entirety including the est entries that were removed from this table are on Table S7. doi:10.1371/journal.pone.0031658.t001

by WISH supports the validity of both the deep sequencing and bioinformatic methods used to examine the transcriptional profiles of upregulated genes. The genes predicted to be suppressed by this approach was also evaluated by quantitative RT-PCR, but only 2 of 8 genes examined were validated (Table S3). This suggests that this approach is more amenable to the examination of gene induction by *etsrp*. Nevertheless, the complete dataset containing downregulated unigene entries is listed in Table S8.

To determine whether the 50 selected ectopically induced genes are expressed in the developing embryonic zebrafish vasculature, embryos were processed for WISH at a developmental stage when most of the primitive vasculature has formed, 24-hour post fertilization (24hpf). Marked expression in the vasculature was noted for 39 of the 50 genes (Figure 3). To confirm that these genes are functionally downstream of *etsrp* in the vasculature, we

examined their expression in *etsrp* morphants, in which case the expression in the axial vasculature is prominently reduced as demonstrated in Figure 4. Sparse expression in myeloid cells was only detected for one of these genes, *myo1f* (Figure 3L). Although there is some ubiquitous expression of some of these genes, the more pronounced expression in the vasculature is clearly demonstrated in higher magnification images of the trunk regions (Figure S1), where the axial vascular expression in morphants is clearly reduced in all genes. Of the 39 genes with vascular expression only *flh3* (Figure 3 AD' and Figure S1AD') and *acsbg2* (Figure 3AL' and Figure S1AL') are preferentially expressed in the axial venous system at this stage.

In the course of gene selection and WISH probe generation the Ensembl databases were utilized together with those at NCBI to examine gene structure and evolutionary ontology. Several of the

Table 2. Known vascular genes induced by *etsrp* in RNA-seq dataset.

Unigene ID	Gene ID	Gene Description	Fold Change	p-value
Dr.118570	LOC561493	similar to src homology 2 domain containing e	315.3	1.8E-196
Dr.133475	LOC100151130	similar to rasip1 protein	254.8	3.62E-10
Dr.85502	<i>mrc1</i>	mannose receptor c1-like protein	225.7	4.0E-20
Dr.47591	<i>etv2</i>	ets variant gene 2	219.0	4.9E-237
Dr.84866	LOC100005159	similar to hemicentin 1	159.4	9.6E-04
Dr.152900	LOC572378	similar to transforming growth factor-b type ii receptor	110.1	4.90E-03
Dr.107226	<i>sc:d0254</i>	<i>sc:d0254</i>	37.9	5.0E-10
Dr.78408	<i>fli1a</i>	friend leukemia integration 1a	37.6	3.0E-61
Dr.86352	<i>zgc:113016</i>	<i>zgc:113016</i>	36.3	1.5E-09
Dr.36543	<i>aqp8a</i>	aquaporin 8a	32.6	4.9E-237
Dr.83871	<i>rasgrp3</i>	ras guanyl releasing protein 3 (calcium and dag-regulated)	32.6	2.6E-08
Dr.89996	<i>egfl7</i>	egf-like-domain, multiple 7	21.6	1.5E-10
Dr.79866	<i>yrc</i>	yes-related kinase	20.9	7.0E-151
Dr.103153	<i>plxnd1</i>	plexin d1	20.1	4.9E-237
Dr.118013	<i>cdh5</i>	cadherin 5	19.1	3.5E-18
Dr.80968	<i>fli1b</i>	friend leukemia integration 1b	16.3	1.8E-07
Dr.77989	<i>dusp5</i>	dual specificity phosphatase 5	14.8	1.2E-16
Dr.76027	<i>ker18</i>	keratin 18	12.5	2.1E-05
Dr.75812	<i>tal1</i>	t-cell acute lymphocytic leukemia 1	12.4	6.9E-08
Dr.150623	<i>f10</i>	coagulation factor x	9.8	7.2E-04
Dr.89035	<i>tmem88a</i>	transmembrane protein 88 a	9.2	7.2E-16
Dr.75094	<i>kdr1</i>	kinase insert domain receptor like	7.8	1.6E-36
Dr.80363	<i>clcc14a</i>	c-type lectin domain family 14, member a	6.3	4.0E-06
Dr.151971	<i>sox7</i>	sry-box containing gene 7	6.1	4.4E-02
Dr.74559	<i>scarf1</i>	scavenger receptor class f, member 1	5.6	1.4E-02
Dr.75958	<i>robo4</i>	roundabout homolog 4	5.4	3.7E-14
Dr.87001	<i>cldn5b</i>	claudin5b	5.1	1.6E-15
Dr.83306	<i>mcam</i>	melanoma cell adhesion molecule	4.7	4.0E-39
Dr.52827	<i>zfp2b</i>	zinc finger protein, multitype 2b	4.4	1.5E-05
Dr.5660	<i>crip2</i>	cysteine-rich protein 2	4.4	9.4E-11
Dr.107483	<i>sb:cb911</i>	<i>sb:cb911</i>	4.3	3.1E-03
Dr.75385	LOC563577	similar to novel apoptosis-stimulating protein of p53	4.2	3.1E-06
Dr.132454	<i>smox</i>	spermine oxidase	3.2	4.8E-55
Dr.599	<i>ldb2a</i>	lim-domain binding factor 2a	3.2	3.7E-03
Dr.81683	<i>rbpms2</i>	rna binding protein with multiple splicing 2	3.1	1.2E-09
Dr.81298	<i>flt4</i>	fms-related tyrosine kinase 4	3.0	2.1E-29
Dr.79626	<i>ildr2</i>	immunoglobulin-like domain containing receptor 2	2.9	1.4E-21
Dr.132331	<i>nrp1b</i>	neuropilin 1b	2.9	2.5E-15
Dr.78142	<i>acvr1</i>	activin a receptor type ii-like 1	2.8	6.7E-19
Dr.135121	<i>stab2</i>	stabilin	2.7	4.5E-10
Dr.76054	<i>tpm4</i>	tropomyosin 4	2.6	1.9E-37
Dr.79413	<i>jam2</i>	junctional adhesion molecule 2	2.6	1.5E-21
Dr.91385	<i>kdr</i>	kinase insert domain receptor (a type iii receptor tyrosine kinase)	2.5	2.5E-29
Dr.88777	<i>pdlim4</i>	pdz and lim domain 4	2.5	2.3E-02
Dr.76395	<i>C8orf4</i>	chromosome 8 open reading frame 4	2.4	1.0E-13
Dr.37960	<i>fgfr1b</i>	fibroblast growth factor receptor-like 1b	2.4	2.1E-07
Dr.75409	<i>gapdhs</i>	glyceraldehyde-3-phosphate dehydrogenase, spermatogenic	2.3	2.7E-03
Dr.22604	<i>amot</i>	angiomin	2.3	2.4E-08
Dr.80539	<i>elov1b</i>	elongation of very long chain fatty acids-like 1b	2.3	2.7E-09
Dr.82429	LOC563907	similar to tumor endothelial marker 8	2.1	6.5E-03

Table 2. Cont.

Unigene ID	Gene ID	Gene Description	Fold Change	p-value
Dr.104822	<i>sidkey-261h17.1</i>	<i>sidkey-261h17.1</i>	2.1	4.6E-107
Dr.78553	<i>micall2</i>	<i>mical-like 2</i>	2.0	5.6E-18

Genes in bold text were identified within the *etsrp* overexpressed dataset by visual inspection of the data. Those not in bold were identified by DAVID cluster analysis. doi:10.1371/journal.pone.0031658.t002

genes probed for were noted to have redundant gene names as of the current genome build, Zv9 in Ensembl. These paralogs have been noted in their description below with an asterisk and further depicted in Table S4 with chromosomal location and identity. Following is a brief description of 10 selected genes in the order that they are presented in Figures 3 and 4. Descriptions for the remaining 29 genes are continued in Text S1.

A. *lyg2*

The goose type lysozyme ortholog *lyg2* was originally identified in goose egg white, where it functions in the catalysis of bacterial cell wall break down, and is expressed ubiquitously in flounder [29]. In chicken, it is expressed in lung and non-adherent fraction of bone marrow cells, possibly reflecting expression within myelocytes where it exhibits an innate immune function [30].



Figure 2. Verification of RNA-seq data. Wild type and *etsrp* injected, *etsrp* oe, embryos were collected at late stages of gastrulation and processed by WISH. Ectopic induction was detected as random positively labeled cells in embryos injected with *etsrp* RNA at the one cell stage. Endogenous expression was observed in the presumptive forerunner cells (arrows) of wild-type uninjected embryos in (C) *she* and (E) *spa17*. Endogenous expression was also observed in (K) *tmem119b* (L) *myl9* (AD') *sh3bp4* (AN') *hmha1* (AT') *aff3* and (AW') *acsbg2*. Ectopic induction is clearly detected in 49 of the 50 genes examined, but not (AT') *aff3*, which has a relatively high level of endogenous expression at this stage. Wild type uninjected embryos are positioned on the left column with their *etsrp* oe counterparts on the right for each gene, anterior is facing left where possible. Scale bar: 250 μ m.

doi:10.1371/journal.pone.0031658.g002

Table 3. Genes examined from the RNA-seq *etsrp* overexpression dataset.

Unigene ID	Gene ID	Gene Description	Fold Change	p-value
Dr.153782	<i>lyg2</i>	<i>lysozyme G like 2</i>	1899.9	4.5E-26
Dr.149118	<i>cdc42GAP1</i>	<i>similar to Cdc42 GTPase-activating protein</i>	441.5	1.9E-21
Dr.118570	<i>she</i>	<i>similar to Src homology 2 domain containing E</i>	315.3	1.8E-196
Dr.40434	<i>ms4a17a.11</i>	<i>membrane-spanning 4-domains, subfamily A, member 17A.11</i>	300.4	3.5E-06
Dr.84343	<i>spa17</i>	<i>sperm auto antigen 17</i>	225.9	1.1E-09
Dr.103328	<i>klhl4</i>	<i>kelch like 4</i>	198.7	2.0E-05
Dr.110713	<i>LOC100149611</i>	<i>similar to LOC495463 protein ; "CD93"</i>	165.4	3.8E-08
Dr.111220	<i>gpr183</i>	<i>g protein-coupled receptor 183</i>	151.4	1.9E-04
Dr.78041	<i>xirp2</i>	<i>xin actin-binding repeat containing 2</i>	95.0	7.6E-56
Dr.134501	<i>rassf4</i>	<i>ras association (RalGDS/AF-6) domain family member 4</i>	51.4	1.4E-02
Dr.91332	<i>tmem119b</i>	<i>transmembrane protein 119</i>	50.7	1.8E-42
Dr.89765	<i>myl9</i>	<i>myosin, light chain 9, regulatory</i>	50.1	5.7E-113
Dr.115399	<i>cntn4</i>	<i>contactin 4</i>	38.4	1.4E-02
Dr.84022	<i>myo1E</i>	<i>similar to myosin IE</i>	37.9	8.4E-03
Dr.75719	<i>iclp2</i>	<i>invariant chain-like protein 2</i>	32.2	1.3E-16
Dr.88587	<i>rgl2</i>	<i>ral guanine nucleotide dissociation stimulator-like 2</i>	28.2	1.9E-36
Dr.76656	<i>capn8</i>	<i>calpain 8</i>	23.5	1.1E-05
Dr.11010	<i>mhc1uea</i>	<i>major histocompatibility complex class I UEA gene</i>	21.8	2.4E-21
Dr.77849	<i>dgki</i>	<i>similar to diacylglycerol kinase, iota</i>	17.4	4.4E-04
Dr.117215	<i>ifit5</i>	<i>similar to Interferon-induced protein with tetratricopeptide repeats 5</i>	15.1	2.1E-03
Dr.134371	<i>plscr2</i>	<i>similar to phospholipid scramblase 2</i>	10.6	1.5E-47
Dr.7340	<i>slc16a9a</i>	<i>solute carrier family 16 (monocarboxylic acid transporters), member 9a</i>	9.3	2.1E-23
Dr.89035	<i>tmem88a</i>	<i>transmembrane protein 88a</i>	9.2	7.2E-16
Dr.100033	<i>rasa4</i>	<i>ras p21 protein activator 4-like</i>	7.8	2.3E-04
Dr.96217	<i>samd10</i>	<i>similar to sterile alpha motif domain containing 10</i>	7.2	1.2E-02
Dr.114623	<i>cald1</i>	<i>caldesmon 1</i>	6.8	1.6E-22
Dr.91020	<i>sept5b</i>	<i>septin 5b</i>	6.7	1.3E-05
Dr.119058	<i>zgc:171494</i>	<i>zgc:171494</i>	6.4	2.8E-02
Dr.84654	<i>ankdd1a</i>	<i>ankyrin repeat and death domain containing 1A</i>	6.4	2.5E-09
Dr.91634	<i>sh3bp4</i>	<i>novel protein similar to vertebrate SH3-domain binding protein 4</i>	6.0	2.0E-06
Dr.82145	<i>myof</i>	<i>similar to fer-1-like 3, myoferlin</i>	5.7	4.1E-32
Dr.84960	<i>grtp1b</i>	<i>growth hormone regulated TBC protein 1b</i>	5.6	1.4E-02
Dr.85673	<i>tagap</i>	<i>t-cell activation GTPase activating protein</i>	5.5	2.8E-03
Dr.124255	<i>plc-12</i>	<i>similar to Inactive phospholipase c-like protein 2</i>	5.2	2.9E-05
Dr.80073	<i>fhl3</i>	<i>four and a half LIM domains 3</i>	5.1	5.4E-37
Dr.133138	<i>irf9</i>	<i>interferon regulatory factor 9</i>	4.9	9.4E-12
Dr.18530	<i>FAM166B</i>	<i>family with sequence similarity 166, member B</i>	4.4	1.8E-04
Dr.5660	<i>crip2</i>	<i>cysteine-rich protein 2.</i>	4.4	9.4E-11
Dr.135601	<i>zcchc4</i>	<i>zinc finger, CCHC domain containing 4</i>	4.0	4.4E-03
Dr.77065	<i>hmha1</i>	<i>histocompatibility (minor) HA-1</i>	4.0	9.1E-12
Dr.47691	<i>ankrd58</i>	<i>ankyrin repeat domain-containing protein 58-like</i>	3.8	3.1E-06
Dr.92393	<i>ccdc135</i>	<i>similar to coiled-coil domain-containing protein 135</i>	3.6	5.2E-05
Dr.83578	<i>tmem151b</i>	<i>similar to transmembrane protein tmem151B</i>	3.5	1.9E-03
Dr.81141	<i>si:ch211-132b12.7</i>	<i>hypothetical protein LOC564531</i>	3.5	1.4E-02
Dr.78875	<i>sept9b</i>	<i>septin 9b</i>	3.3	1.3E-05
Dr.83297	<i>aff3</i>	<i>af4/fmr2 family, member 3; LAF</i>	2.4	1.5E-10
Dr.90997	<i>tmem179</i>	<i>transmembrane protein 179</i>	2.0	2.8E-09
Dr.34190	<i>si:ch211-250g4.3</i>	<i>si:ch211-250g4.3</i>	2.0	4.5E-04
Dr.107097	<i>acsbg2</i>	<i>Acyl-CoA synthetase bubblegum family member 2</i>	1.9	2.8E-64
Dr.77920	<i>arhgap27</i>	<i>similar to Rho GTPase-activating protein 27</i>	1.9	9.1E-07

Bold are expressed in vascular endothelial cells at the time examined (24 hpf).
doi:10.1371/journal.pone.0031658.t003

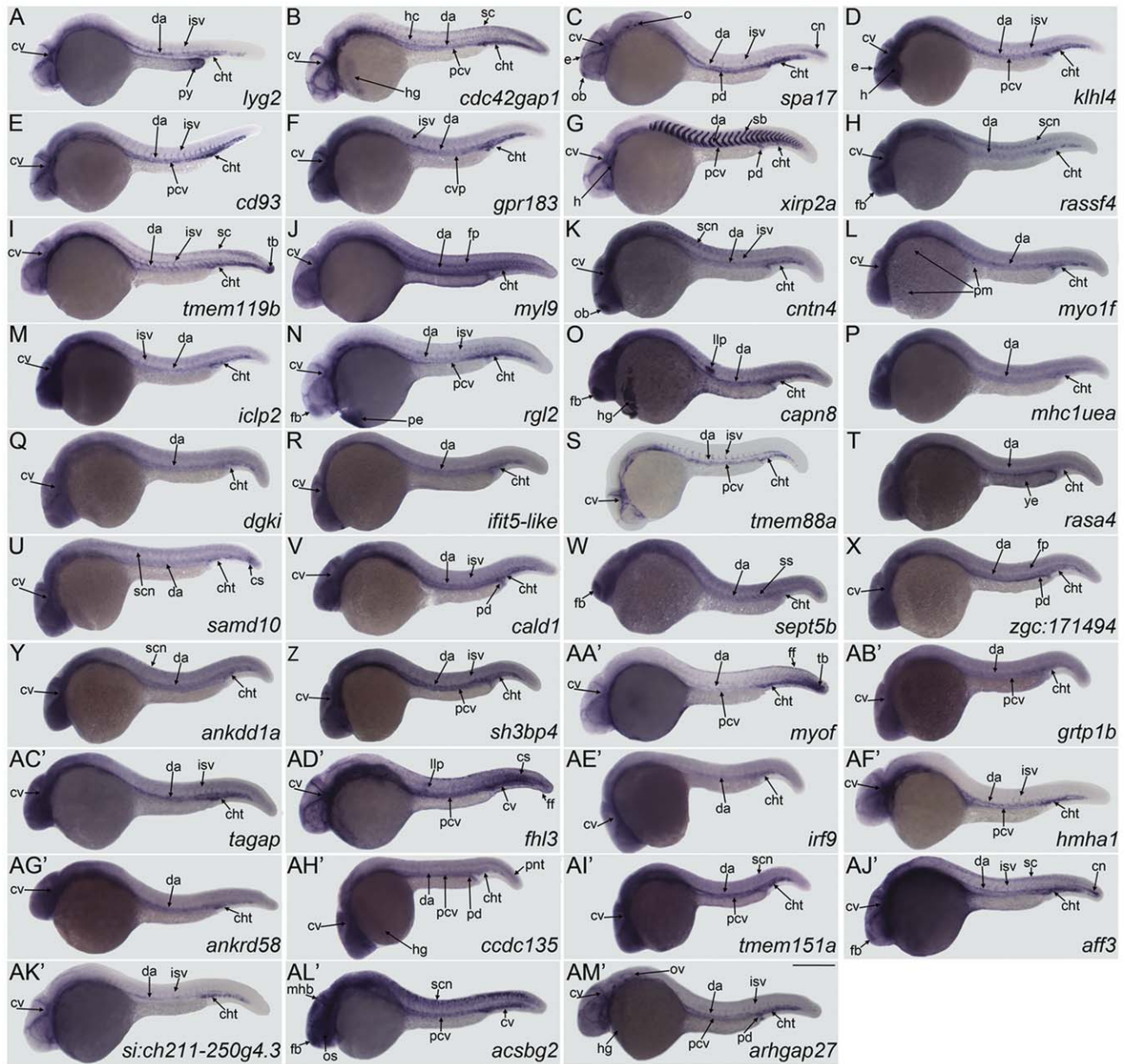


Figure 3. Genes with expression in vascular endothelial cells. The following genes are restricted to the vasculature and are expressed in the cranial vasculature, dorsal aorta, caudal hematopoietic tail region, posterior cardinal vein, and intersegmental vessels: (E) *cd93*; (F) *gpr183*; (S) *tmem88a*; (AF') *hmha1*, while the following are expressed in the same structures but exclude the posterior cardinal vein: (A) *lyg2* is also expressed in the posterior yolk extension; (AC') *tagap*; (AK') *si:ch211-250g.4.3*. These genes are expressed in the cranial vasculature, dorsal aorta, caudal hematopoietic tail region, posterior cardinal vein and other tissues as noted: (B) *cdc42gap1*, is also expressed in the hypochord, hatching gland, and sparsely in spinal cord; (D) *klhl4* is also expressed in the heart, intersegmental vessels, and epiphysis; (G) *xirp2a* is also expressed in the heart, somite boundaries and pronephric duct; (N) *rgl2*, is also expressed in the intersegmental vessels, forebrain and primitive erythrocytes; (AA') *myof* is also expressed in the fin fold, and tailbud; (AM') *arhgap27* is also expressed in the intersegmental vessels, otic vesicles, pronephric duct, and hatching gland. The following genes are expressed in the cranial vasculature, dorsal aorta, caudal hematopoietic tail region, and intersegmental vessels as well as (C) *spa17*, the olfactory bulb, epiphysis, otoliths, pronephric duct, and caudal notocord; (H) *rassf4* was not detected in the intersegmental vessels but is observed in the forebrain, floor plate of neural tube, and spinal cord neurons; (I) *tmem119b* is also expressed in spinal cord, and tailbud; (AJ') *aff3* is also expressed in the forebrain, spinal cord, and caudal notocord. (AD') *fh13* is expressed in the cranial vasculature, posterior cardinal vein, caudal vein, fin fold, lateral line primordium, and lightly in caudal somites. (AL') *acsbg2* is expressed in the posterior cardinal vein, caudal vein, spinal cord neurons, optic stalk, forebrain, and midbrain-hindbrain boundary. The following genes have basal levels of ubiquitous expression with prominent labeling of the cranial vasculature, dorsal aorta, caudal hematopoietic tail region and other structures and tissues where indicated: (J) *myl9* is also expressed in the floor plate; (K) *cntn4*, is also expressed in the intersegmental vessels, olfactory bulb, and spinal cord neurons; (L) *myo1f* is also expressed in primitive myeloid cells; (M) *iclp2*, also has slight expression in anterior intersegmental vessels; (Q) *dgki*; (R) *ifit5-like*; (U) *samd10* is also expressed in the caudal somites, and neurons of the anterior spinal cord; (V) *cald1* is also expressed in the intersegmental vessels and pronephric duct; (X) *zgc:171494* is also expressed in the floor plate, and pronephric duct; (Y) *ankdd1a* is also expressed in the spinal cord neurons; (Z) *sh3bp4* is also expressed in the posterior cardinal vein, and intersegmental vessels; (AB') *grtp1b* is also expressed in the posterior cardinal vein; (AE') *irf9*; (AG') *ankrd58*; (AH') *ccdc135* is also expressed in the posterior cardinal vein, pronephric duct, posterior neural tube, and slightly in the hatching gland; (AI') *tmem151a* is also expressed in the posterior cardinal vein, spinal cord neurons and intersegmental vessels (see Figure S1A1'). Basal ubiquitous expression with darker

staining of the dorsal aorta and caudal hematopoietic tail region is noted in: (O) *capn8* which includes expression in the forebrain, hatching gland, and lateral line primordium; (P) *mhc1uea*; (T) *rasa4* is also expressed in the yolk extension; (W) *sept5b* is also expressed in the forebrain and somites. Embryos are positioned laterally with the anterior facing left. Abbreviations: cht, caudal hematopoietic tail region; cn, caudal notochord; cs, caudal somites; cv, cranial vasculature; da, dorsal aorta; e, epiphysis; ff, fin fold; fp, floor plate; h, heart; hg, hatching gland; isv, intersegmental vessels; llp, lateral line primordium; mhb, midbrain hindbrain boundary; o, otoliths; ob, olfactory bulb; os, optic stalk; ov, otic vesicle; pd, pronephric duct; pe, primitive erythroid cells; pm, primitive myeloid cells; py, posterior yolk extension; sb, somite boundaries; sc, spinal cord; scn, spinal cord neurons; ss, somites; tb, tailbud; ye, yolk extension. Scale bar: 250 μ m.
doi:10.1371/journal.pone.0031658.g003

Contrary to mammalian *lysozyme G*, the *lysozyme G2* in flounder and zebrafish lack a signal peptide and a conserved cystein catalytic triad, suggesting an intracellular role with a potentially distinct function [31]. There are 2 other paralogs for this gene. *lyg1* (*zgc:92608*), is expressed in macrophages [32], while the expression of *zgc:162941* has not been reported.

B. *cdc42GAP1/ARHGAP1*

Originally cloned in mice, *cdc42GAP1* catalyzes the hydrolysis of GTP bound by the Rho GTPases *cdc42* and *rac1*, inactivating them [33]. In mice and humans it is expressed ubiquitously but exhibits marked expression in heart and lungs [34]. Mice knockouts result in generally reduced organ size relative to wild-type siblings due to increased rates of basal apoptosis, and display reduced viability [35]. The few survivors experience premature aging associated with heightened genomic instability [36]. In vitro evidence suggests *cdc42GAP1* negatively modulates angiogenesis [37]. Although it is expressed ubiquitously in mammals, it is predominantly expressed in the developing vasculature and heart in zebrafish at the stage examined here (Figure 3B).

C. *spa17/zgc:153721*

The protein is homologous with the human *sperm protein antigen 17*, *spa17*, and both contain a cAMP dependent protein kinase

regulatory subunit domain and an IQ motif that binds proteins with EF-hand motifs, but *zgc:153721/spa17* of zebrafish has a glutamine rich region between these motifs that is absent in *spa17* in humans and mice. Although in mice it is expressed ubiquitously and is expressed in embryos, in rabbits, it is most highly expressed in the tissue from which it was originally identified, testis [38,39]. Increased expression of *spa17* has been detected in esophageal, ovarian, and cervical cancers and has been used as a clinical marker for cancer in these tissues [40,41,42]. In zebrafish, *spa17* is also expressed in the pronephros (Figure 3C).

D. *klhl4*

kelch like homolog like 4, *klhl4* is a member of the kelch motif containing family of proteins [43]. The founding member, *kelch*, was cloned from *Drosophila*, and mediates cytoplasmic streaming through ring canals in the developing oocyte [44]. In human fetal tissue, *klhl4* shows ubiquitous expression, and the protein is structurally similar to the founding member, *kelch1* [45]. As in humans, the *klhl4* of zebrafish has a BTB/POZ like domain and a BACK domain at the N-terminus, but while human *klhl4* has 6 kelch domains at the carboxy terminus, zebrafish has 3. The function of this protein has not been characterized.

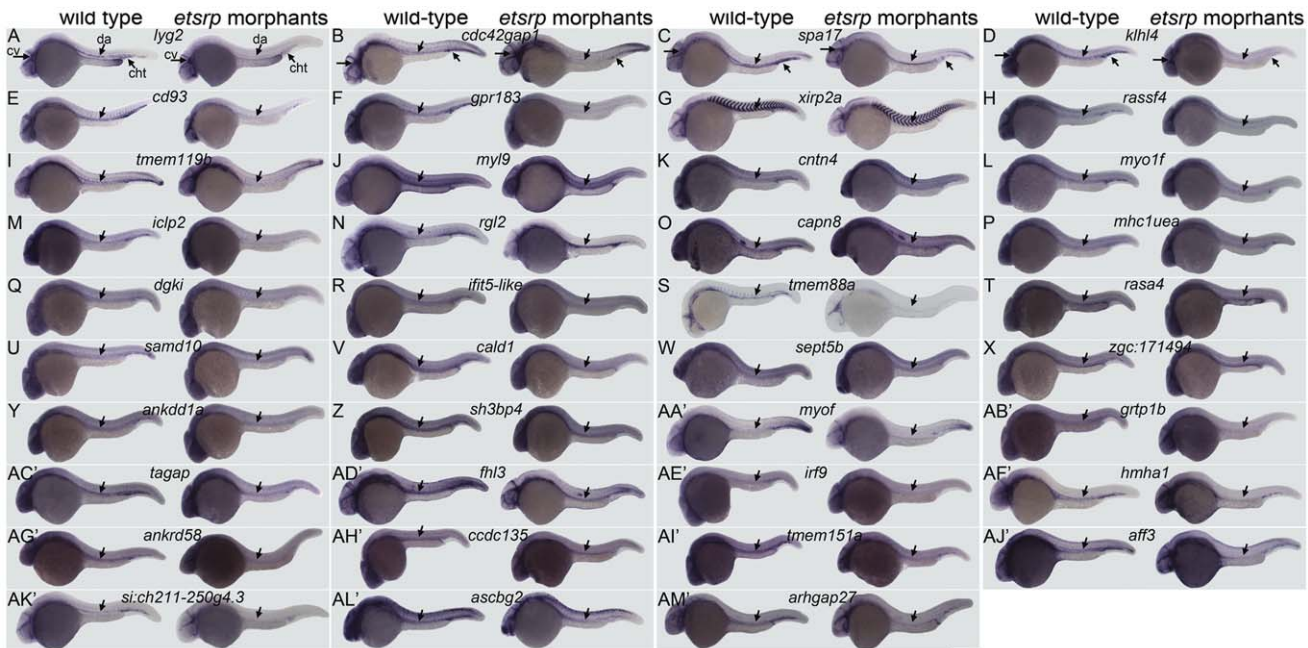


Figure 4. Analysis of gene expression in *etsrp* morphants. *Flk1-gfp* embryos were injected at the one cell stage with a mixture of translation blocking morpholinos for *etsrp* and analyzed by WISH at the 24hpf stage. Marked reduction is apparent in all genes examined in the axial vasculature, which includes the dorsal aorta and posterior cardinal vein, and is marked with a downward facing arrow in all images. The primitive myeloid cells stained in (L) *myo1f* wild-type controls are absent in their *etsrp* morphant counterparts. The staining in the axial trunk region of *rgl2* in *etsrp* morphants marks primitive erythrocytes that are trapped due to lack of circulation. The expression of non-vascular structures is not affected otherwise in *etsrp* morphants. Embryos are positioned laterally with the anterior facing left. Abbreviations: cht, caudal hematopoietic tail region; cv, cranial vasculature; da, dorsal aorta. Scale bar: 250 μ m.
doi:10.1371/journal.pone.0031658.g004

E. *cd93*

Other aliases of *cd93* include *complement component C1q receptor like* (*C1qr*), *lymphocyte antigen 68* (*ly68*), and *AA4.1*. *cd93* is a widely conserved member of the c type lectin family whose expression in endothelial and other cell types has been documented in mice and humans [46,47]. In previous reports the name *complement receptor C1qR-like* (*crl*) has been used for a different gene in zebrafish, which is the likely homolog of *c-type lectin domain family 14, member a* (*clec14a*) [3,18,19]. Although both of these genes are members of the c-type lectin family, they are paralogs of each other that deserve distinction, and we note that both are expressed in the vasculature, since *clec14a* [3] and *cd93* shown here, are vascular specific at the stage examined (Figure 3E).

F. *gpr183**

Epstein bar induced 2, *ebi2*, or *g protein coupled receptor 183*, *gpr183*, is an orphan receptor that was originally identified as a gene induced by Epstein Bar Virus infection in Buritt lymphoma cells [48]. Knockout mice are viable without any notable gross phenotype, however they experience defects in B-cell mobility within lymphoid tissues, and consequently have difficulty mounting rapid antibody responses [49,50]. Northern blot evaluation of its expression was originally limited to peripheral blood mononuclear cells (PBMCs), lymphocytic tissues and lungs [48]. However, in separate assays it has been detected in the aorta, atria, gastrointestinal tract, but not in the hematopoietic bone marrow or fetal liver [51]. At the stage examined, *gpr183* is clearly expressed in the developing vasculature of zebrafish embryos (Figure 3F), but not in the heart.

G. *xirp2a*

xin actin binding repeat containing 2a, *xirp2a*, is one of three orthologs for the mammalian counterpart, *xirp1 β* [52]. The paralogs in zebrafish are *xirp2b*, which is restricted to the myotome borders during the larval stages of development (ZFIN), and *xirp2c*, which has not been characterized. Functionally, xin repeat motifs cross-link F-actin with β -catenin at adherens junctions [53]. *Xirp2a* expression in the myoseptum, pronephric duct, heart, otic vesicle, head mesoderm, and liver has been demonstrated in developing zebrafish larvae [32]. In this study the expression in cranial and axial vasculature in the trunk becomes more evident. The most likely explanation for the differences observed might be a longer in situ color developing time used here.

H. *rassf4*

ras association (RalGDS/AF-6) domain family member 4, *rassf4*, is a member of Ras association domain family of proteins, which contain a Ras association domain and in the case of *rassf4*, a SARAH domain (reviewed in [54]). *rassf4* is expressed ubiquitously in humans, but is downregulated in tumorous cells due to the hypermethylation of its promoter [55]. The idea that *rassf4* is a tumor suppressor is supported by the apoptotic death of tumor cells following forced expression of *rassf4* [55]. Besides expression in the vasculature of 1 dpf zebrafish embryos, *rassf4* is also highly expressed in spinal cord neurons (Figure 3H, and Figure S1H).

I. *tmem119b*

Neither the protein encoded by *tmem119b* nor its paralog, *tmem119a*, have been characterized. *tmem119b* encodes a single pass type I transmembrane protein while *tmem119a* encodes a multipass transmembrane protein with two transmembrane domains. As a conserved gene in vertebrates, it is currently unclear whether *tmem119b* or *tmem119a* is the ortholog of their counterpart in mice, *osteoblast induction factor*, *obif*, which promotes osteoblast differenti-

ation [56]. Although *obif* and *tmem119b* share a structural topological profile, *tmem119a* has a negligible higher amount of amino acid conservation with *obif*. It is also currently unknown whether *tmem119/obif* plays the same roles in humans.

J. *myl9*

myosin light chain 9 regulatory, *myl9*, is a calcium-regulated protein that regulates the contraction of myosin heavy chains in non-skeletal muscle cells that was originally cloned from human umbilical artery [57]. Currently there is evidence that the expression of *myl9* may be directly regulated by *jumb* [58], and *runx1* [59]. In zebrafish there is a close paralog, *myosin light polypeptide 9-like* (*zgc: 77916*), with 95% gene identity to *myl9* that also remains to be characterized.

Discussion

In this report we provide further support for the utility of combining the relatively unbiased approach of high throughput next generation RNA-sequencing to analyze whole embryo transcriptomes in zebrafish. The high correlation between genes induced by *etsrp* as predicted by RNA-seq and their confirmation by WISH reinforces the effectiveness of the methods used to interrogate transcriptional profiles in zebrafish. Of the genes selected from the dataset obtained, we identified a very high percentage of genes expressed in the vasculature of embryos, validating this approach to identify biologically relevant genes. However there may be non-vascular genes in this data that are induced due to the induction of off-target genes by the artificially elevated level of *etsrp*. This issue can be addressed by WISH using wild type embryos at a proper embryonic stage. Additionally, this method does not discriminate direct and indirect targets induced by *etsrp*. Further analysis by ChIP-Seq should reveal those genes that are transcriptionally regulated by *etsrp*.

Both the ectopic induction of these genes by *etsrp* overexpression and down-regulation by *etsrp* deficiency in vascular endothelial and primitive myeloid cells highlights their genetic epistatic relationship to *etsrp* and show that *etsrp* is a potent transcriptional activator whose expression and activity must be tightly regulated. Although several genes exhibit basal levels of ubiquitous expression their level in the vasculature is still higher. These genes deserve as much attention as those only specific to the vascular lineage, as several studies have shown that the down-regulation of such genes resulted in the disruption of vascular development. For instance, knockdown of *nrarp-a* and *nrarp-b* [60,61], *lpa₁* [62], and *dep1a/dep1b* [63] respectively, resulted in defects in intersegmental vessels, lymphangiogenesis, and arterial specification. We obtained approximately 850 genes that were induced 1.9 fold or more by *etsrp* expression. Based on previous annotation and studies we can classify 6.2% of them as vascular genes. Experimental studies identified 39 more vascular genes and further increased the percentage to 10.8%. This leaves over 700 unigene entries in the dataset, some of which may be expressed and actively involved in the development of the zebrafish vasculature, and are potentially worth examining. The complete dataset is available in Table S5 and we hope this will benefit research on the development of the vascular system. The panel of genes found here remain to be characterized further and will require subsequent functional studies.

Materials and Methods

Embryo injection

Flk1-gfp embryos were injected with 75 pg *etsrp* RNA at the one cell stage as described [5], and collected at the late gastrulation

stages, 70–90% epiboly, when ectopic *flkl1-gfp* expression is notable in injected embryos. Total RNA was extracted with the PerfectPure RNA Tissue kit (5 Prime). For pre-sequencing validation, cDNA was made with Oligo(dT) primers and Superscript II reverse transcriptase (Invitrogen). The following primers pairs were used for verification by RT-PCR: *β-actin*: 5′ -TGTTTTCCCCTCCATTGTTG-3′, 5′-ACATACATGGCAGGGGTGTT-3′; *etsrp* 3′UTR: 5′-GAGGAATTCTCGAAGGATTGG-3′; TGGTTTTCTAAAGGCACCTAG-3′; *Fli1a*: 5′-CCGAGGTCCTGCTCTCACAT-3′; 5′-GGGACTGGTCAGCGTGAGAT-3′; *scl*: 5′-GGAGATGCGGAACAGTATGG-3′, 5′-GAAGGCACCGTTCACTTCT-3′. To test gene expression in *etsrp* morphants, *flkl1-gfp* embryos were injected with a mixture of 4 ng each of translation blocking *etsrp* morpholinos as described in [5] at the one cell stage, and embryos were harvested at 24hpf for WISH.

Cloning and Whole Mount In Situ Hybridization (WISH)

Different segments of the genes selected for evaluation were cloned into the pCRII Topo vector with the TOPO TA cloning kit (dual promoter) according to the manufacturer's instructions (Invitrogen). Amplicons were obtained from an *etsrp* overexpression cDNA library at late stages of gastrulation (70%–90% epiboly). Primers used for amplification are noted in Table S5. Positive plasmids were confirmed by sequencing. WISH was performed as described [64]. DIG labeled RNA probes were generated by linearizing TOPO-cloned genes with restriction endonucleases (New England Biolabs), and transcribing with SP6 or T7 RNA polymerase (Promega).

Image acquisition and processing

In situ stained embryos were further processed by a serial dehydration in ethanol, followed by rehydration into 1XPBS. Embryos were imaged in 2% methylcellulose in depression microscope slides. Images were captured with a color digital CCD camera (AxioCam, Zeiss) mounted on an upright microscope (Axioskop2 plus, Zeiss) with Openlab 4.0.2 software (Improvision, Lexington, MA). Serial images were combined, merged and processed with Adobe Photoshop CS5.

Deep sequencing library construction and bioinformatics

mRNA-seq libraries were constructed with the Tru-Seq RNA prep kit (Illumina) according to manufacturer's instructions. After this project was initiated, the Illumina sequencer has gone through a series of improvements. As a result, we obtained both single-end and paired-end reads with different lengths (76, 51, and 46 bp) using the Illumina GA II sequencer and its earlier versions. The obtained reads were mapped to 51,481 Unigene transcript sequences (Build #117, October 2009) [26] using BWA [27] allowing up to 4 mismatches. In case of the paired-end reads, we determined whether the mapped results support correct pairing of the reads according to the unigene annotation. If one read in a pair is mapped to a gene, the other one should be mapped to the same gene. Since each read in a pair may map to multiple locations in the genome, all possible combinations of their mappings were examined for correct pairing. The pair of reads is considered uniquely mapped only if one unique pair of mapped locations was identified.

To determine expression levels of genes and exons, we used the variable RPKM (reads per kilobase of exon per million mapped reads) defined by Mortazavi et al. 2008 [65]. Analysis of differential gene expression was carried out to obtain the differentially expressed genes between control sample and *etsrp* oe sample. The number of uniquely mapped read-pairs for each gene in each sample was stored. The total number of mapped

reads in each lane was normalized using the total mapped reads in each lane. Fisher's exact test was then performed using the above read counts for each gene. The resulting p-values were corrected via the Benjamini and Hochberg method as implemented in R. Finally, differentially expressed genes were defined as those with changes of at least 1.9-fold between a pair of samples at a false discovery rate (FDR) of 5%. Table S7 lists all 849 upregulated unigene hits in the *etsrp* oe sample, while the 726 downregulated hits are tabulated in Table S8.

Quantitative PCR (qPCR)

qPCR was performed and analyzed exactly as stated in [12], with three independent biological replicates of either uninjected or *etsrp* oe groups of embryos harvested at late gastrulation stages. Primers used are listed in Table S6.

Supporting Information

Figure S1 Higher magnification of Figure 4. The axial trunk vasculature of embryos displayed in figure 4 were imaged at higher magnification to highlight the changes observed. Wild-type embryos are on the left half of each column with their *etsrp* morphant counterparts on the right for each gene. Embryos were positioned with anterior facing left.

(TIF)

Table S1 Total RNA-Seq reads obtained in control and *etsrp* oe samples and mapping results.

(XLS)

Table S2 DAVID GO derived gene subset without clear vascular expression or not expressed in developing vasculature.

(XLS)

Table S3 Quantitative PCR of unigene hits predicted to be suppressed by *etsrp* oe via RNA-seq.

(XLS)

Table S4 Genes with double entries in Ensembl.

(XLS)

Table S5 Primers used for cloning to make WISH probes.

(XLS)

Table S6 Primers used to test downregulated gene set following *etsrp* overexpression by qPCR.

(XLS)

Table S7 Complete *etsrp* mRNA-seq dataset (induced greater than or equal to 1.9 fold by *etsrp* overexpression).

(XLS)

Table S8 Downregulated mRNA-seq dataset (reduced greater than or equal to 1.9 fold by overexpression of *etsrp*).

(XLS)

Text S1 Supplementary Text. A continuation of the results section that describes information regarding the genes examined in this study.

(DOC)

Acknowledgments

We thank Scott Doherty for excellent technical support and Anqi Liu for help with animal husbandry.

Author Contributions

Conceived and designed the experiments: GAG SL. Performed the experiments: GAG MBV. Analyzed the data: GAG J-HL MBV JL.

Contributed reagents/materials/analysis tools: XX. Wrote the paper: GAG.

References

1. Stainier DY, Fishman MC (1994) The zebrafish as a model system to study cardiovascular development. *Trends Cardiovasc Med* 4: 207–212.
2. Weinstein BM (2002) Plumbing the mysteries of vascular development using the zebrafish. *Semin Cell Dev Biol* 13: 515–522.
3. Sumanas S, Joraniak T, Lin S (2005) Identification of novel vascular endothelial-specific genes by the microarray analysis of the zebrafish cloche mutants. *Blood* 106: 534–541.
4. Pham VN, Lawson ND, Mugford JW, Dye L, Castranova D, et al. (2007) Combinatorial function of ETS transcription factors in the developing vasculature. *Dev Biol* 303: 772–783.
5. Sumanas S, Lin S (2006) Ets1-related protein is a key regulator of vasculogenesis in zebrafish. *PLoS Biol* 4: e10.
6. Sumanas S, Gomez G, Zhao Y, Park C, Choi K, et al. (2008) Interplay among Etsrp/ER71, Scl, and Alk8 signaling controls endothelial and myeloid cell formation. *Blood* 111: 4500–4510.
7. Lee D, Park C, Lee H, Lugas JJ, Kim SH, et al. (2008) ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. *Cell Stem Cell* 2: 497–507.
8. Kataoka H, Hayashi M, Nakagawa R, Tanaka Y, Izumi N, et al. (2011) Etv2/ER71 induces vascular mesoderm from Flk1+PDGFR α + primitive mesoderm. *Blood*.
9. Neuhaus H, Müller F, Hollemann T (2010) Xenopus er71 is involved in vascular development. *Dev Dyn* 239: 3436–3445.
10. Salanga MC, Meadows SM, Myers CT, Krieg PA (2010) ETS family protein ETV2 is required for initiation of the endothelial lineage but not the hematopoietic lineage in the Xenopus embryo. *Dev Dyn* 239: 1178–1187.
11. Proulx K, Lu A, Sumanas S (2010) Cranial vasculature in zebrafish forms by angioblast cluster-derived angiogenesis. *Dev Biol* 348: 34–46.
12. Ren X, Gomez GA, Zhang B, Lin S (2010) Scl isoforms act downstream of etsrp to specify angioblasts and definitive hematopoietic stem cells. *Blood* 115: 5338–5346.
13. Liu F, Patient R (2008) Genome-wide analysis of the zebrafish ETS family identifies three genes required for hemangioblast differentiation or angiogenesis. *Circulation Research* 103: 1147–1154.
14. Mathavan S, Lee SG, Mak A, Miller LD, Murthy KR, et al. (2005) Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genet* 1: 260–276.
15. Covassin L, Amigo JD, Suzuki K, Teplyuk V, Straubhaar J, et al. (2006) Global analysis of hematopoietic and vascular endothelial gene expression by tissue specific microarray profiling in zebrafish. *Dev Biol* 299: 551–562.
16. Qian F, Zhen F, Ong C, Jin S-W, Meng Soo H, et al. (2005) Microarray analysis of zebrafish cloche mutant using amplified cDNA and identification of potential downstream target genes. *Dev Dyn* 233: 1163–1172.
17. Weber GJ, Choe SE, Dooley KA, Paffett-Lugassy NN, Zhou Y, et al. (2005) Mutant-specific gene programs in the zebrafish. *Blood* 106: 521–530.
18. Gomez GA, Veldman MB, Zhao Y, Burgess S, Lin S (2009) Discovery and characterization of novel vascular and hematopoietic genes downstream of etsrp in zebrafish. *PLoS ONE* 4: e4994.
19. Wong KS, Proulx K, Rost MS, Sumanas S (2009) Identification of vasculature-specific genes by microarray analysis of Etsrp/Etv2 overexpressing zebrafish embryos. *Dev Dyn* 238: 1836–1850.
20. Hegedus Z, Zakrzewska A, Agoston VC, Ordas A, Rác P, et al. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol* 46: 2918–2930.
21. Ordas A, Hegedus Z, Henkel CV, Stockhammer OW, Butler D, et al. (2010) Deep sequencing of the innate immune transcriptomic response of zebrafish embryos to Salmonella infection. *Fish Shellfish Immunol*.
22. Stockhammer OW, Rauwerda H, Wittink FR, Breit TM, Meijer AH, et al. (2010) Transcriptome analysis of Traf6 function in the innate immune response of zebrafish embryos. *Mol Immunol* 48: 179–190.
23. Aanes H, Winata CL, Lin CH, Chen JP, Srinivasan KG, et al. (2011) Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome research*.
24. Soares AR, Pereira PM, Santos B, Egas C, Gomes AC, et al. (2009) Parallel DNA pyrosequencing unveils new zebrafish microRNAs. *BMC Genomics* 10: 195.
25. Aday AW, Zhu LJ, Lakshmanan A, Wang J, Lawson ND (2011) Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Dev Biol*.
26. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, et al. (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31: 28–33.
27. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
28. Zakrzewska A, Cui C, Stockhammer OW, Benard EL, Spaink HP, et al. (2010) Macrophage-specific gene functions in Spi1-directed innate immunity. *Blood* 116: e1–11.
29. Hikima J, Minagawa S, Hirono I, Aoki T (2001) Molecular cloning, expression and evolution of the Japanese flounder goose-type lysozyme gene, and the lytic activity of its recombinant protein. *Biochim Biophys Acta* 1520: 35–44.
30. Nakano T, Graf T (1991) Goose-type lysozyme gene of the chicken: sequence, genomic organization and expression reveals major differences to chicken-type lysozyme gene. *Biochim Biophys Acta* 1090: 273–276.
31. Irwin DM, Gong Z (2003) Molecular evolution of vertebrate goose-type lysozyme genes. *J Mol Evol* 56: 234–242.
32. Thisse B, Heyer V, Lux A, Alunni V, Degraeve A, et al. (2004) Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods Cell Biol* 77: 505–519.
33. Lamarche-Vane N, Hall A (1998) CdGAP, a novel proline-rich GTPase-activating protein for Cdc42 and Rac. *J Biol Chem* 273: 29172–29177.
34. Tcherkezian J, Triki I, Stenne R, Danek EI, Lamarche-Vane N (2006) The human orthologue of CdGAP is a phosphoprotein and a GTPase-activating protein for Cdc42 and Rac1 but not RhoA. *Biol Cell* 98: 445–456.
35. Wang L, Yang L, Burns K, Kuan C-Y, Zheng Y (2005) Cdc42GAP regulates c-Jun N-terminal kinase (JNK)-mediated apoptosis and cell number during mammalian perinatal growth. *Proc Natl Acad Sci USA* 102: 13484–13489.
36. Wang L, Yang L, Debidda M, Witte D, Zheng Y (2007) Cdc42 GTPase-activating protein deficiency promotes genomic instability and premature aging-like phenotypes. *Proc Natl Acad Sci USA* 104: 1248–1253.
37. Engelse MA, Laurens N, Verloop RE, Koolwijk P, van Hinsbergh VWM (2008) Differential gene expression analysis of tubule forming and non-tubule forming endothelial cells: CDC42GAP as a counter-regulator in tubule formation. *Angiogenesis* 11: 153–167.
38. Richardson RT, Yamasaki N, O'Rand MG (1994) Sequence of a rabbit sperm zona pellucida binding protein and localization during the acrosome reaction. *Dev Biol* 165: 688–701.
39. Wen Y, Richardson RT, Widgren EE, O'Rand MG (2001) Characterization of Sp17: a ubiquitous three domain protein that binds heparin. *Biochem J* 357: 25–31.
40. Gupta G, Sharma R, Chattopadhyay TK, Gupta SD, Ralhan R (2007) Clinical significance of sperm protein 17 expression and immunogenicity in esophageal cancer. *Int J Cancer* 120: 1739–1747.
41. Li FQ, Liu Q, Han YL, Wu B, Yin HL (2010) Sperm protein 17 is highly expressed in endometrial and cervical cancers. *BMC Cancer* 10: 429.
42. Straughn JM, Shaw DR, Guerrero A, Bhoola SM, Racelis A, et al. (2004) Expression of sperm protein 17 (Sp17) in ovarian cancer. *Int J Cancer* 108: 805–811.
43. Adams J, Kelso R, Cooley L (2000) The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol* 10: 17–24.
44. Xue F, Cooley L (1993) kelch encodes a component of intercellular bridges in Drosophila egg chambers. *Cell* 72: 681–693.
45. Braybrook C, Warry G, Howell G, Arnason A, Bjornsson A, et al. (2001) Identification and characterization of KLHL4, a novel human homologue of the Drosophila Kelch gene that maps within the X-linked cleft palate and Ankyloglossia (CPX) critical region. *Genomics* 72: 128–136.
46. Fonseca MI, Carpenter PM, Park M, Palmari G, Nelson EL, et al. (2001) ClqR(P), a myeloid cell receptor in blood, is predominantly expressed on endothelial cells in human tissue. *J Leukoc Biol* 70: 793–800.
47. Petrenko O, Beavis A, Klaine M, Kittappa R, Godin I, et al. (1999) The molecular characterization of the fetal stem cell marker AA4. *Immunity* 10: 691–700.
48. Birkenbach M, Josefsen K, Yalamanchili R, Lenoir G, Kieff E (1993) Epstein-Barr virus-induced genes: first lymphocyte-specific G protein-coupled peptide receptors. *J Virol* 67: 2209–2220.
49. Gatto D, Paus D, Basten A, Mackay CR, Brink R (2009) Guidance of B cells by the orphan G protein-coupled receptor EBI2 shapes humoral immune responses. *Immunity* 31: 259–269.
50. Pereira JP, Kelly LM, Xu Y, Cyster JG (2009) EBI2 mediates B cell segregation between the outer and centre follicle. *Nature* 460: 1122–1126.
51. Rosenkilde MM, Binned-Jensen T, Andersen H, Holst PJ, Kledal TN, et al. (2006) Molecular pharmacological phenotyping of EBI2. An orphan seven-transmembrane receptor with constitutive activity. *J Biol Chem* 281: 13199–13208.
52. Grosskurth SE, Bhattacharya D, Wang Q, Lin JJ-C (2008) Emergence of Xin demarcates a key innovation in heart evolution. *PLoS ONE* 3: e2857.
53. Choi S, Gustafson-Wagner EA, Wang Q, Harlan SM, Sinn HW, et al. (2007) The intercalated disk protein, mXinalpha, is capable of interacting with beta-catenin and bundling actin filaments [corrected]. *J Biol Chem* 282: 36024–36036.
54. Avruch J, Xavier R, Bardeesy N, Zhang X-F, Praskova M, et al. (2009) Rassf family of tumor suppressor polypeptides. *J Biol Chem* 284: 11001–11005.
55. Eckfeld K, Hesson L, Vos MD, Bieche I, Latif F, et al. (2004) RASSF4/AD037 is a potential ras effector/tumor suppressor of the RASSF family. *Cancer Res* 64: 8688–8693.

56. Kanamoto T, Mizuhashi K, Terada K, Minami T, Yoshikawa H, et al. (2009) Isolation and characterization of a novel plasma membrane protein, osteoblast induction factor (obif), associated with osteoblast differentiation. *BMC Dev Biol* 9: 70.
57. Kumar CC, Mohan SR, Zavodny PJ, Narula SK, Leibowitz PJ (1989) Characterization and differential expression of human vascular smooth muscle myosin light chain 2 isoform in nonmuscle cells. *Biochemistry* 28: 4027–4035.
58. Licht AH, Nübel T, Feldner A, Jurisch-Yaksi N, Marcello M, et al. (2010) Junb regulates arterial contraction capacity, cellular contractility, and motility via its target Myl9 in mice. *J Clin Invest* 120: 2307–2318.
59. Jalagadugula G, Mao G, Kaur G, Goldfinger LE, Dhanasekaran DN, et al. (2010) Regulation of platelet myosin light chain (MYL9) by RUNX1: implications for thrombocytopenia and platelet dysfunction in RUNX1 haploinsufficiency. *Blood* 116: 6037–6045.
60. Topczewska JM, Topczewski J, Szostak A, Solnica-Krezel L, Hogan BLM (2003) Developmentally regulated expression of two members of the Nrarp family in zebrafish. *Gene Expr Patterns* 3: 169–171.
61. Phng LK, Potente M, Leslie JD, Babbage J, Nyqvist D, et al. (2009) Nrarp coordinates endothelial Notch and Wnt signaling to control vessel density in angiogenesis. *Dev Cell* 16: 70–82.
62. Lee SJ, Chan TH, Chen TC, Liao BK, Hwang PP, et al. (2008) LPA1 is essential for lymphatic vessel development in zebrafish. *FASEB J* 22: 3706–3715.
63. Rodriguez F, Vacaru A, Overvoorde J, den Hertog J (2008) The receptor protein-tyrosine phosphatase, Dep1, acts in arterial/venous cell fate decisions in zebrafish development. *Dev Biol* 324: 122–130.
64. Thisse C, Thisse B (2008) High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat Protoc* 3: 59–69.
65. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 621–628.