Stem Cell Reports, Volume *1*

Supplemental Information

A Cdx4-Sall4 Regulatory Module Controls the Transition from Mesoderm Formation to Embryonic Hematopoiesis

Elizabeth J. Paik, Shaun Mahony, Richard M. White, Emily N. Price, Anthony DiBiase, Bilguujin Dorjsuren, Christian Mosimann, Alan J. Davidson, David Gifford, and Leonard I. Zon

Inventory of supplemental information

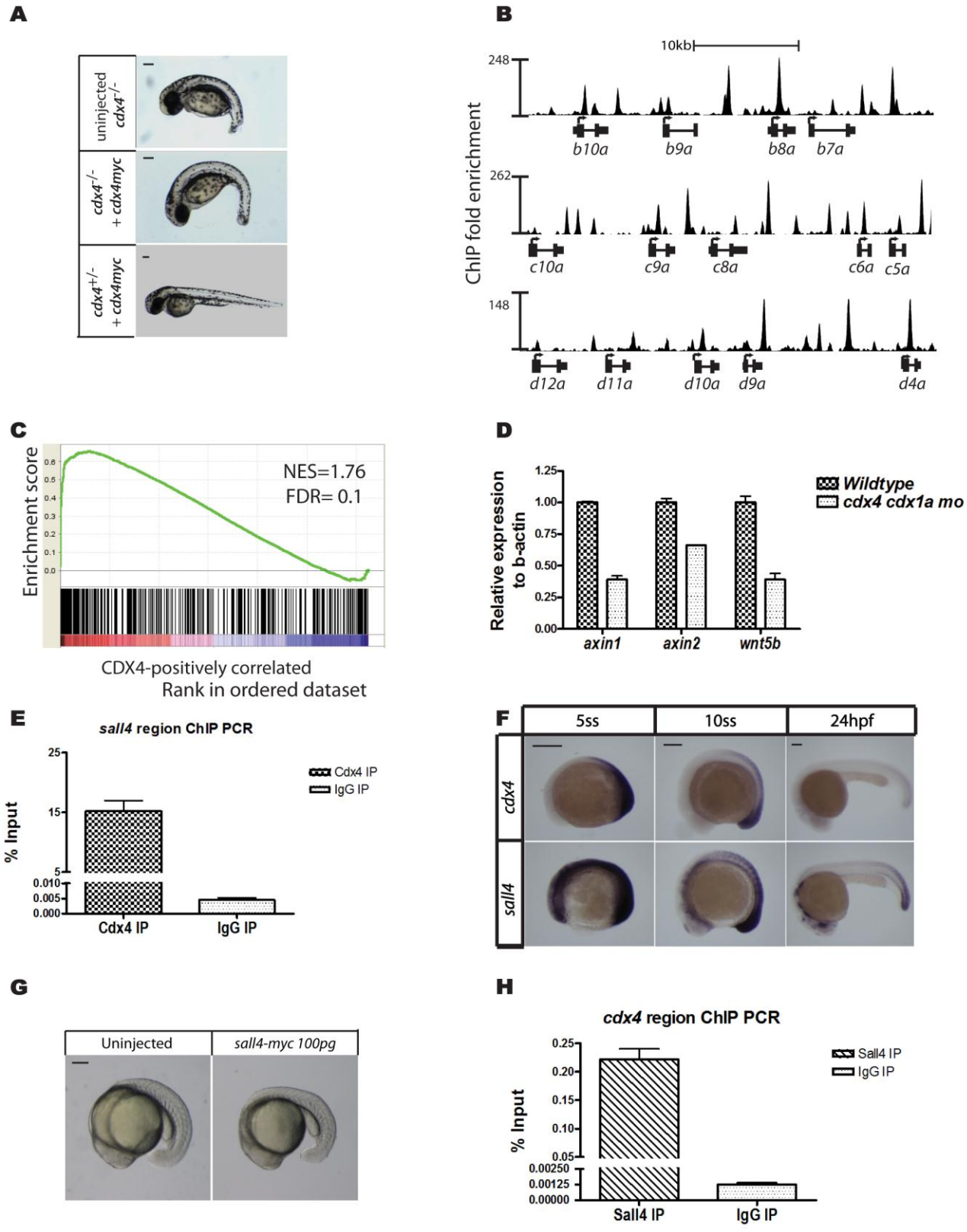
A
uninjected cdx4-/-
cdx4-/- + cdx4myc
cdx4+/- + cdx4myc
B
10kb
248
b10a
b9a
b8a
b7a
262
c10a
c9a
c8a
c6a
c5a
148
d12a
d11a
d10a
d9a
d4a
ChIP fold enrichment
C
Enrichment score
NES=1.76
FDR= 0.1
CDX4-positively correlated
Rank in ordered dataset
D
Relative expression to b-actin
Wildtype
cdx4 cdx1a mo
axin1
axin2
wnt5b
E
sall4 region ChIP PCR
% Input
Cdx4 IP
IgG IP
F
5ss
10ss
24hpf
cdx4
sall4
G
Uninjected
sall4-myc 100pg
H
cdx4 region ChIP PCR
% Input
Sall4 IP
IgG IP

Figure S1. Sall4 is a Cdx4 target gene, related to Figure 1.

A. *cdx4myc* mRNA rescues *cdx4*$^{-/-}$. Top: uninjected *cdx4*$^{-/-}$ embryo at 24hpf. Middle: *cdx4*$^{-/-}$ embryo with 50pg *cdx4myc* mRNA. Bottom: *cdx4*$^{+/-}$ embryo with 50pg cdx4myc mRNA. Scale bar = 200um. B. Gene track of the *hox ba, ca,* and *da* loci showing Cdx4 binding at genomic regions along the x axis and the total number of reads on the y axis. The genomic scale is shown on the top. C. GSEA of the Cdx4-bound genes using the ranked probe sets (16,000 probes) from the *cdx4/cdx1a* double morphant microarray list. Cdx4-bound genes are enriched in the downregulated gene list (NES=1.76, FDR=0.1). D. Real-time quantitative PCR results of the Wnt pathway genes - *axin1, axin2,* and *wnt5* - in wild-type and *cdx4/cdx1a* double morphants. E. Cdx4 ChIP-PCR result of the *sall4* locus. F. 100pg *sall4myc* mRNA does not cause morphological defects at 20-somite stage. Scale bar=200um G. *cdx4* and *sall4* WISH comparison from the 5-somite stage to 24hpf stage. Scale bar=200um H. Sall4 ChIP-PCR result of the *cdx4* locus.

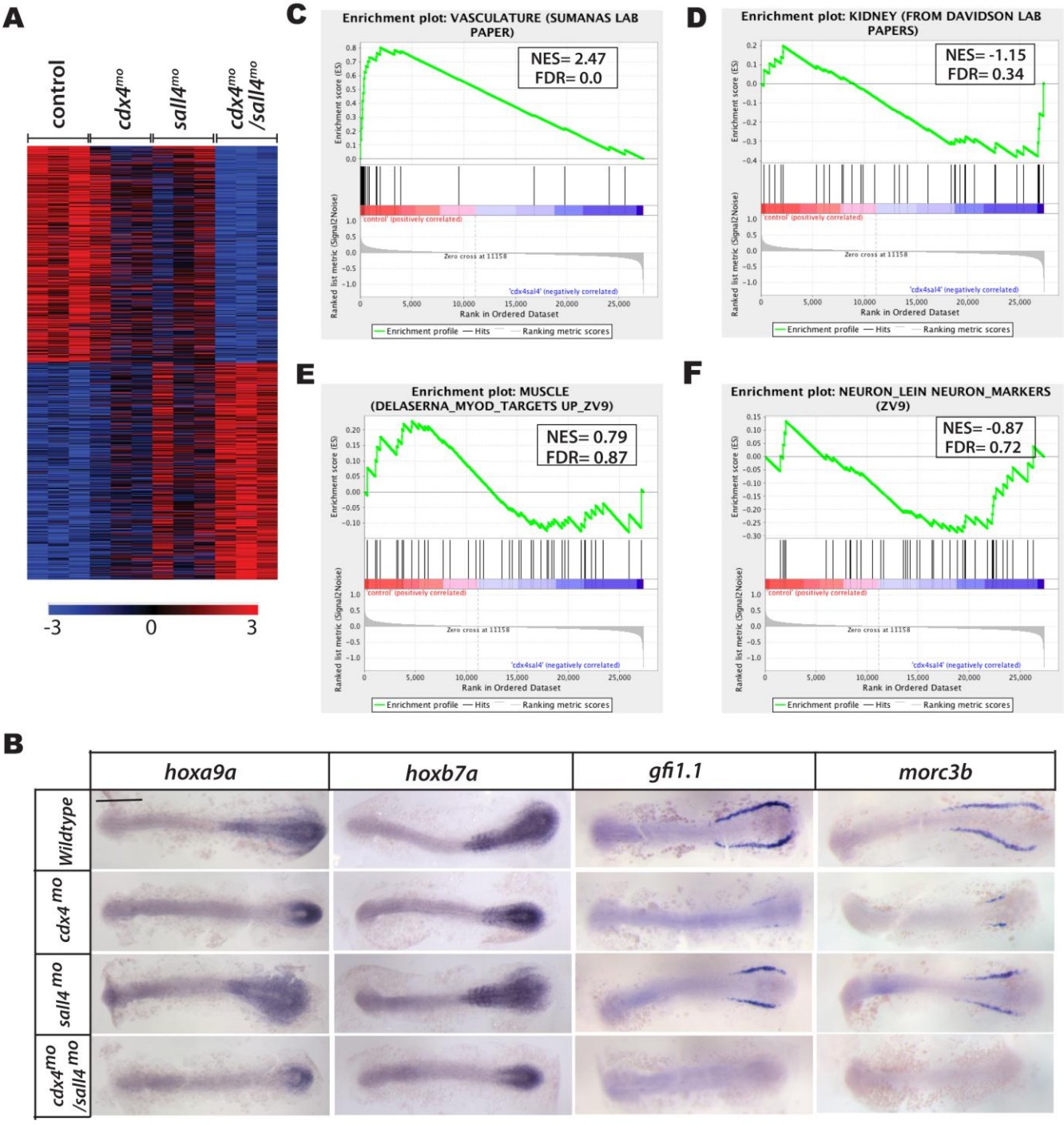

Figure S2. Cdx4 and Sall4 co-regulate downstream target genes, related to Figure 2.

A. Heatmap demonstrating the top 1000 genes that show cooperation between the *cdx4* and *sall4* knockdown at the 3-somite stage. Genes only moderately affected in either single

morphant show significantly greater effects in the 3-somite stage *cdx4/sall4* morphants. B. WISH of *hoxa9a, hoxb7a, gfi1.1*, and *morc3b* at 10-somite stage wild-type, *cdx4* morphant, *sall4* morphant, and *cdx4/sall4* double morphant embryos. Scale bar = 200um. C-F: GSEA based comparison between the *cdx4/sall4* morphants compared to vasculature (C), kidney (D), muscle (E), and neuron (F) gene list.

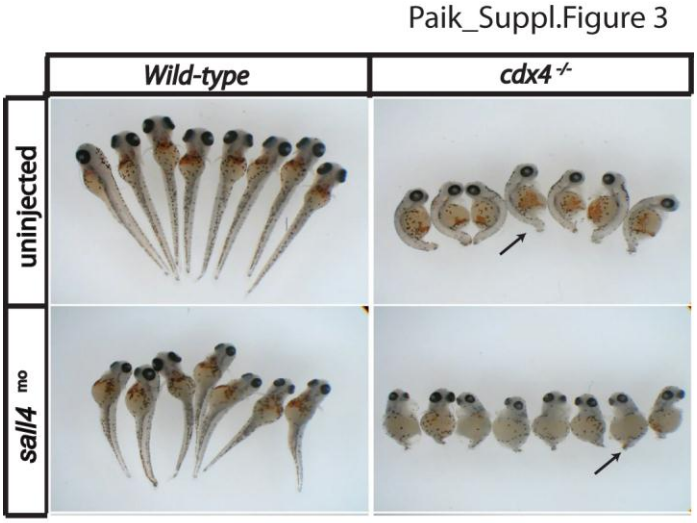

Figure S3. *sall4* cooperates with *cdx4* in zebrafish hematopoiesis, related to Figure 3. O-dianisidine staining of 3 days post fertilization (dpf) old wild-type or *cdx4*$^{-/-}$ embryos that were uninjected or *sall4*^{mo} injected. Note the shorter tails in the *cdx4*$^{-/-}$*;sall4* ^{mo} embryos. Scale bar = 200um.

Paik_Suppl.Figure 4

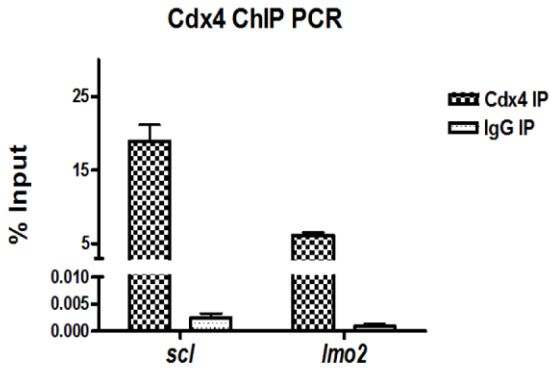

Figure S4. *scl* and *lmo2* loci are bound by Cdx4, related to Figure 4.

Q-PCR of Cdx4 ChIP DNA and IgG ChIP DNA. The bars show relative enrichment +/− SEM.

Paik_Suppl.Figure 5

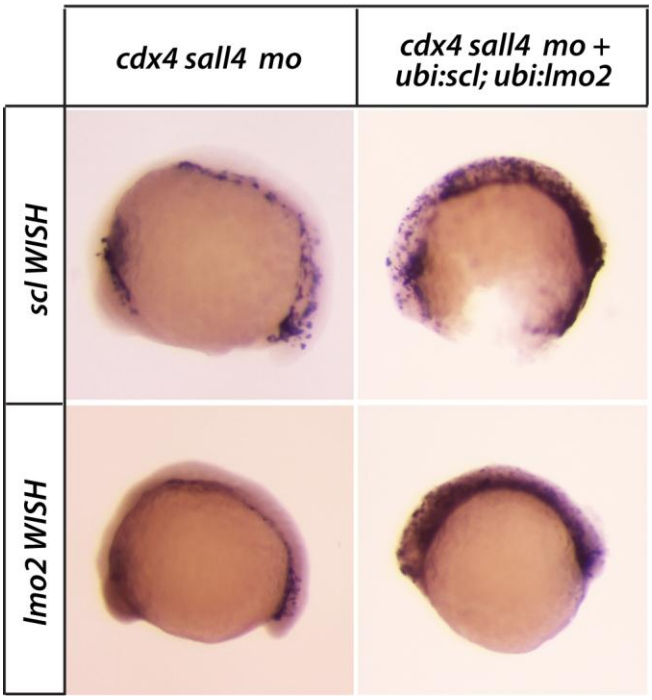

Figure S5. *scl* and *lmo2* transgenes are expressed ubiquitously, related to Figure 5.

WISH of scl and lmo2 in cdx4/sall4 double morphants with no transgenes (left) and double morphants with transgenes (right) at 15 somite stage.

Table S1. Primer sequence information

Name	Sequence
zscl_ChIP_FOR	GCAAATTCATCCCAAGAGGA
zscl_ChIP_REV	GACATTCCCAGCCAATCAGT
zlmo2_ChIP_FOR	TACATGCGCGAATACCAGTC
zlmo2_ChIP_REV	GAAATGTTTGCCAGCCAAGT
zcdx4_ChIP_FOR	TCGTCGCGTTGGTACTGTAG
zcdx4_ChIP_REV	TCAGACTCCGTGCGTATTTG
zsall4_ChIP_FOR	AATTCTTTCGCACCCTCAAA
zsall4_ChIP_REV	ACCAATTGGCACATCCTGTT

Table S2. DNA injection scheme

	ubi:scl	ubi:lmo2	ubi:lacZ	ubi:mcherry	Tol2 mRNA	Total DNA
ubi:scl injection	4ng/ul	-	8ng/ul	4ng/ul	2.5ng/ul	16ng/ul
ubi:scl ubi:lmo2 injection	4ng/ul	4ng/ul	4ng/ul	4ng/ul	2.5ng/ul	16ng/ul
Control plasmid injection	-	-	12ng/ul	4ng/ul	2.5ng/ul	16ng/ul

Supplemental Experimental Procedures

ChIP.

For each ChIP experiment, ~2000 embryos were carefully staged. 500 embryos were then collected, transferred into PBS containing 20mM Na-butyrate, protease inhibitors (Sigma). Embryos were then transferred to 5ml syringe fitted with 21G needle. After leaving ~500ul buffer above the embryos, embryos were forced through the needle by pushing the pistons to

dissociate the embryos. The resulting lysed extract was collected in 2ml eppendorf tubes, and crosslinked by adding formaldehyde solution to final concentration of 1.1%. Crosslinking was conducted for 15 minutes at room temperature while samples were mixed using nutator. After crosslinking, glycine was added to 0.125M to quench the formaldehyde, and incubated on ice for 5 minutes. Tubes were then centrifuged at 470g for 10 minutes at 4C to sediment the cells. The cells were then washed twice with 500ul ice-cold PBS/Na-butyrate/protease inhibitor solution, and flash-frozen and kept at -80C till the ChIP experiment.

ChIP-seq analysis

In GPS, the scaling ratio between IP and control channels was estimated using the median ratio of all 10Kbp windows along the genome. The GPS binding model was initialized to the default and iteratively updated over up to 3 training rounds, and the shape deviation threshold for final peaks was set to -0.2. In this study, we require that reported peaks contain a ChIP-seq enrichment level that is significantly greater than 1.5 times the control channel read count over the same region with *p*-value <0.001 as tested using a Binomial test.

Microarray analysis

Expression values for each chip were computed using the Affymetrix Expression console software suite. Fold-change for each condition (*cdx4*^{mo}*, sall4*^{mo} , or *cdx4*^{mo}*/sall4*^{mo}) was calculated against the control embryo expression levels. The average fold change for each gene was then determined to be up or down regulated compared to control. To identify genes cooperatively affected in the double morphant (compared to the single morphants), we calculated a "cooperation score" defined as (FC[*cdx4*^{mo}*/sall4*^{mo}])/(FC[*cdx4*^{mo}]/FC[*sall4*^{mo}]). This cooperation score was then rank ordered, which identified genes in the double morphant

that were affected to a greater degree than either single morphant alone. The top 500 or 1000 genes most cooperatively affected in the double morphants were then tested for significance using false-discovery rate analysis of Benjamini-Hochberg, and genes with a q value <0.05 were considered significant.

GSEA

Enrichment of the *cdx4/sall4* double morphant gene signature in hematopoietic signatures.

Genes were defined as bound by Cdx4 if they were located within 10Kbp of Cdx4 ChIP-seq peaks. For GSEA analysis, bound gene list was converted into Affymetrix Zebrafish Gene Array v1.0 probe identifiers using NetAffx (www.affymetrix.com), yielding 381 probes. This probe list was then used as input to GSEA v2.0 (www.broadinstitute.org/gsea) for query into the microarray dataset representing *cdx4/cdx1a* morphant data. Enrichment was assessed using the normalized enrichment score and a false-discovery rate of <0.25.

Enrichment of the *cdx4/sall4* double morphant gene signature in hematopoietic signatures.

The microarray signatures of hematopoietic cells in the zebrafish were previously described (Weber et al., 2005). Briefly, fli1a-GFP positive cells were isolated from 10 somite stage zebrafish embryos, and underwent microarray analysis along with the GFP-negative population. Genes significantly up or downregulated in the fli1a-GFP population (compared to the GFP-negative remainder of the embryo) were calculated. Similarly, microarrays were performed on 5 somite stage *scl* morphants or wild-type siblings to generate genes significantly up or downregulated in the *scl* morphants. The endothelial gene set was taken

from the published results (Wong et al., 2009), and the kidney gene set was taken by combining the published WISH results (O'Brien et al., 2011; Wingert and Davidson, 2011; Wingert et al., 2007). Muscle and neuron gene sets were taken from the Broad institute MSigDB. As these two gene sets were created in mouse, gene sets were first converted to zebrafish genes by Ensembl Biomart. Then, each gene was probed in Zfin.org for its expression in developing zebrafish embryos. Only the genes that had expression in their respective organs were taken to generate the final gene list for GSEA.

DNA injection.

Vectors *ubi:scl* (*pCM228*), *ubi:lmo2* (*pCM227*), and *ubi:lacZ* (*pCM258*) are multisite gateway assemblies. *pCM228*: *pENTR5'_ubi* (*pCM206*), *pENTR/D_scl* (*pCM209*, full-length zebrafish ORF from PCR amplification), Tol2kit *#302* (*p3E_SV40polyA*), and *#394* (*pDestTol2A2*) (*ubi:scl*); *pCM227*: *pENTR5'_ubi* (*pCM206*), *pENTR/D_lmo2* (*pCM208*, full-length zebrafish ORF from PCR amplification), Tol2kit *#302* (*p3E_SV40polyA*), and *#394* (*pDestTol2A2*) (*ubi:lmo2*); *pCM258*: *pENTR5'_ubi* (*pCM206*), *pENTR/D_lacZ* (*pCM257*, full-length *lacZ* ORF from PCR amplification), Tol2kit *#302* (*p3E_SV40polyA*), and *#394* (*pDestTol2A2*) (*ubi:lacZ*). *ubi:mCherry* was described previously (Mosimann et al. 2011).