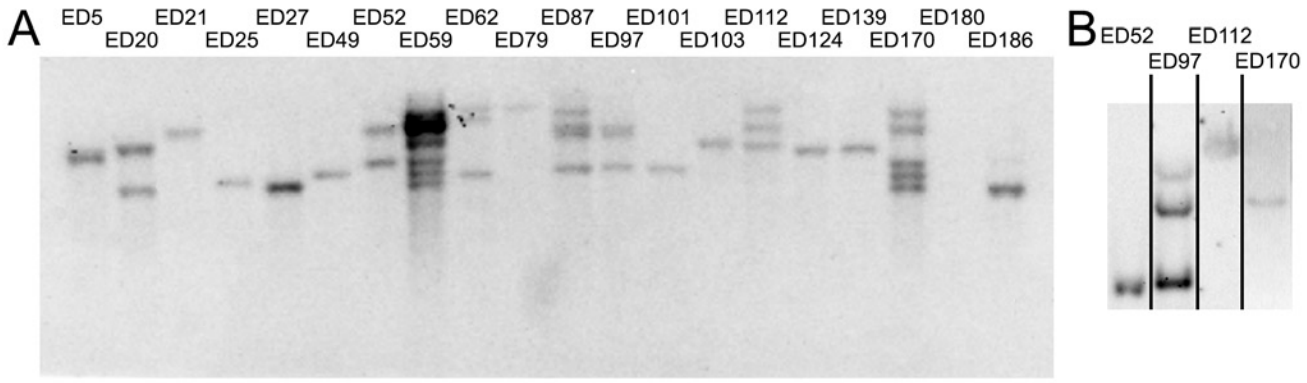
Bessa *et al.*

A mobile insulator system to detect and disrupt cis-regulatory landscapes in vertebrates.

SUPPLEMENTAL FIGURES

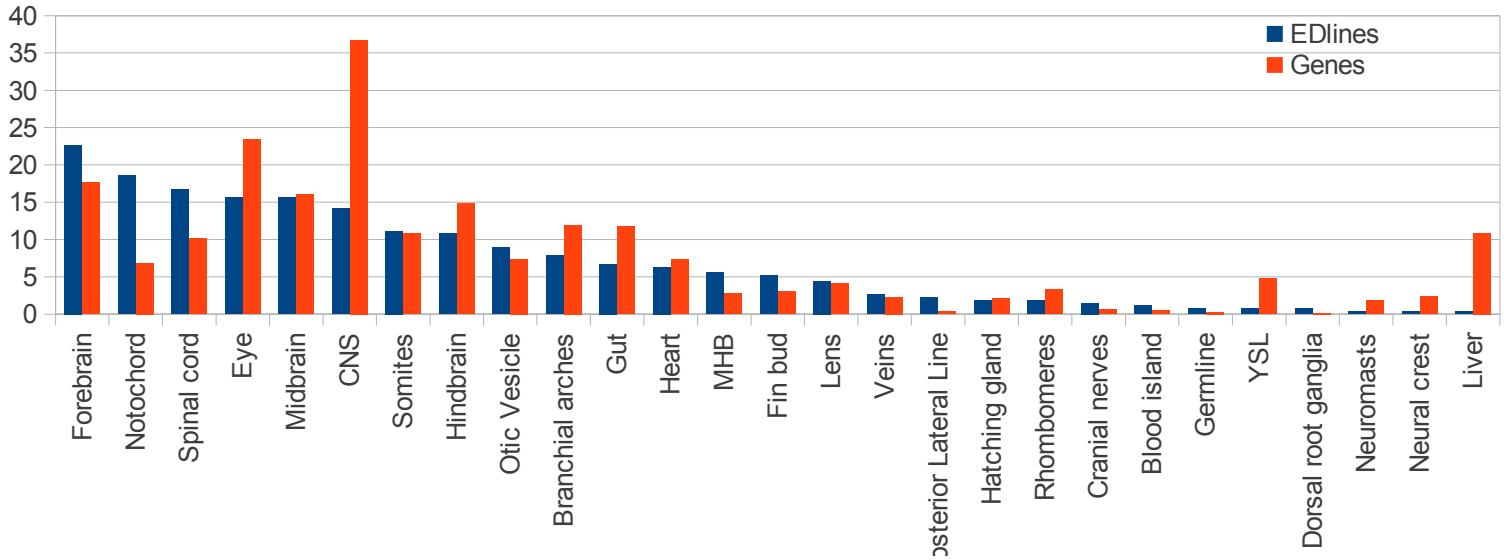
Supplemental Figure 1



Supplemental Figure 1 – Souther blot for 19 ED lines.

(A) Southern blot performed for 19 ED lines using genomic DNA extracted from a single F2 animal per line. Eleven of these animals presented single insertions. (B) A second Southern blot was performed for some lines that presented more than one insertion, using genomic DNA extracted from F4 animals. Three of these animals (ED52, ED112 and ED170) presented single insertions.

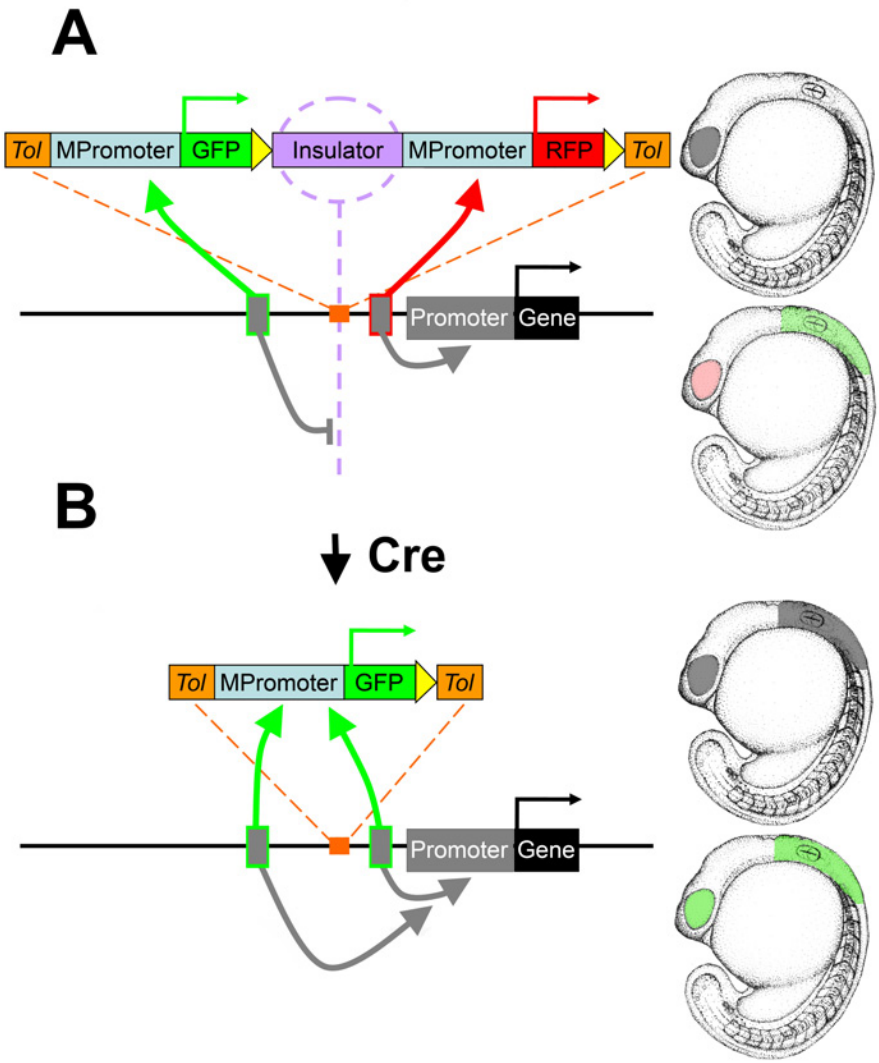
Supplemental Figure 2



Supplemental Figure 2 – Comparison of tissue specific expression of ED lines and genes.

This analysis has been done by comparing the percentage of ED lines and genes that show expression in a given tissue. A paired t-test for these two groups did not show statistically significant differences (Ttest: p=0.36) indicating that in general the data from the ED transposon reflects the expression sites of genes during embryonic development. Analyzing each anatomical region per separate we observed that there are anatomical regions that show higher bias when comparing both groups, however only a minority show differences higher than four-fold (Liver 29.2; YSL 6.4; Neural crest 6.4; Dorsal root ganglia 5.7; Posterior lateral line 5.6; Neuromasts 4.9). It should be noted that the limited number of ED lines expressed in those tissues reduces the statistical significance of these differences. In addition, these differences could be also due to the technical differences for detection of gene expression in both groups (fluorescence vs in situ hybridization). Finally, genes with known expression patterns might not be completely randomly distributed; for instances genes associated to exhaustively studied tissues might be over-represented (e.g. CNS; almost 180.000 entries in Pubmed), while genes associated to tissues not as well known might be under-represented (e.g. notochord; almost 3.000 entries in Pubmed). Data on gene tissue specif expression was extracted from zfin.org (Bradford et al. 2011).

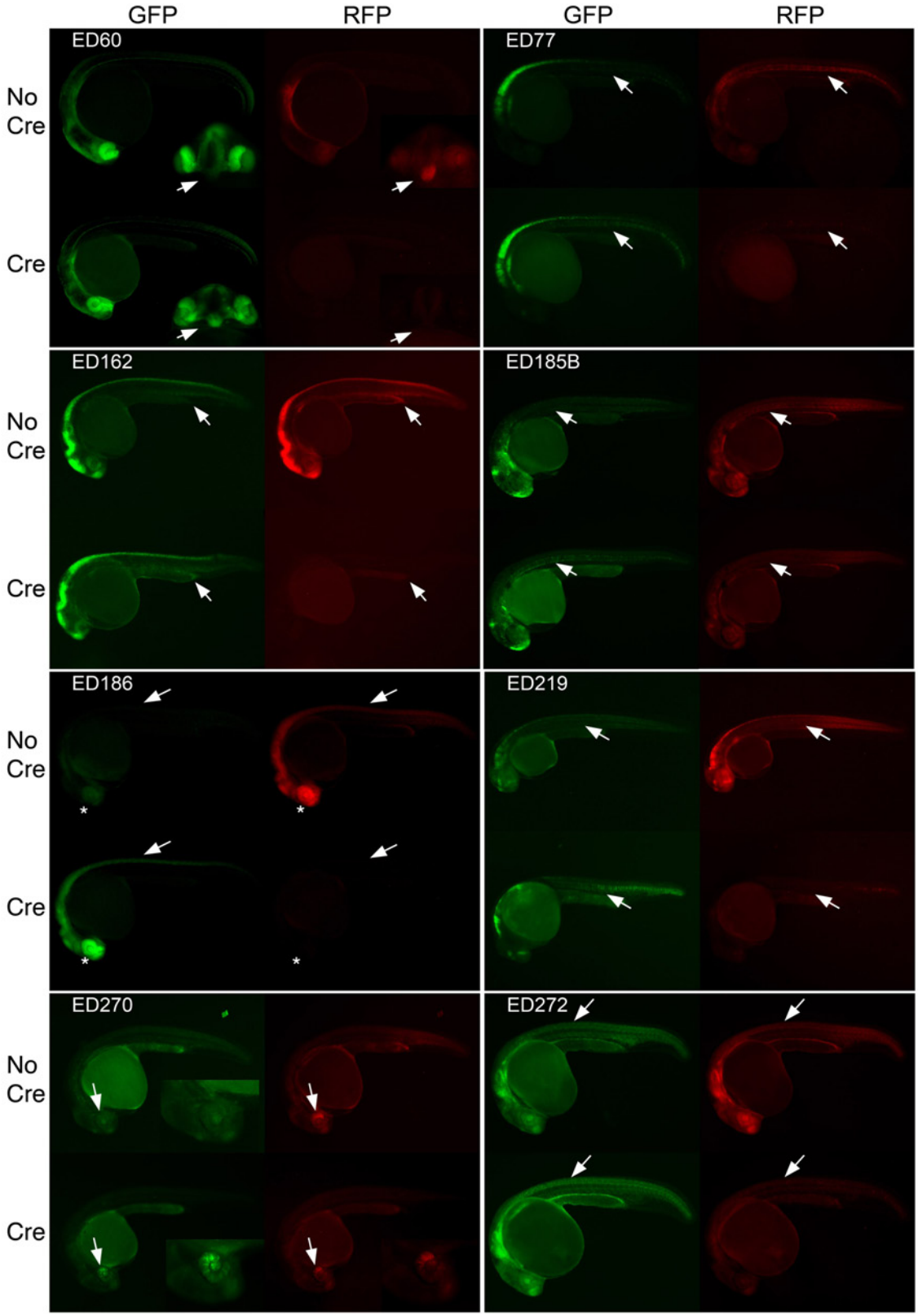
Supplemental Figure 3



Supplemental Figure 3 – Cre mediated excision of the ED's Insulator/RFP cassette.

The *loxP* flanked cassette (yellow triangles) that includes the Insulator and the RFP enhancer trap (A) can be excised by the activity of Cre recombinase (B). In this case, the dowstream enhancer (**a**; red box) is now able to interact with the GFP minimal promoter (B), resulting in a shift of RFP expression (A; lower embryo) to GFP expression in the eye (B; lower embryo). In addition, the upstream enhancer (A; green box) can now interact with the promoter of the nearby gene (B), rescuing the loss of expression in the hindbrain that results from the insulator activity (A and B; higher embryos).

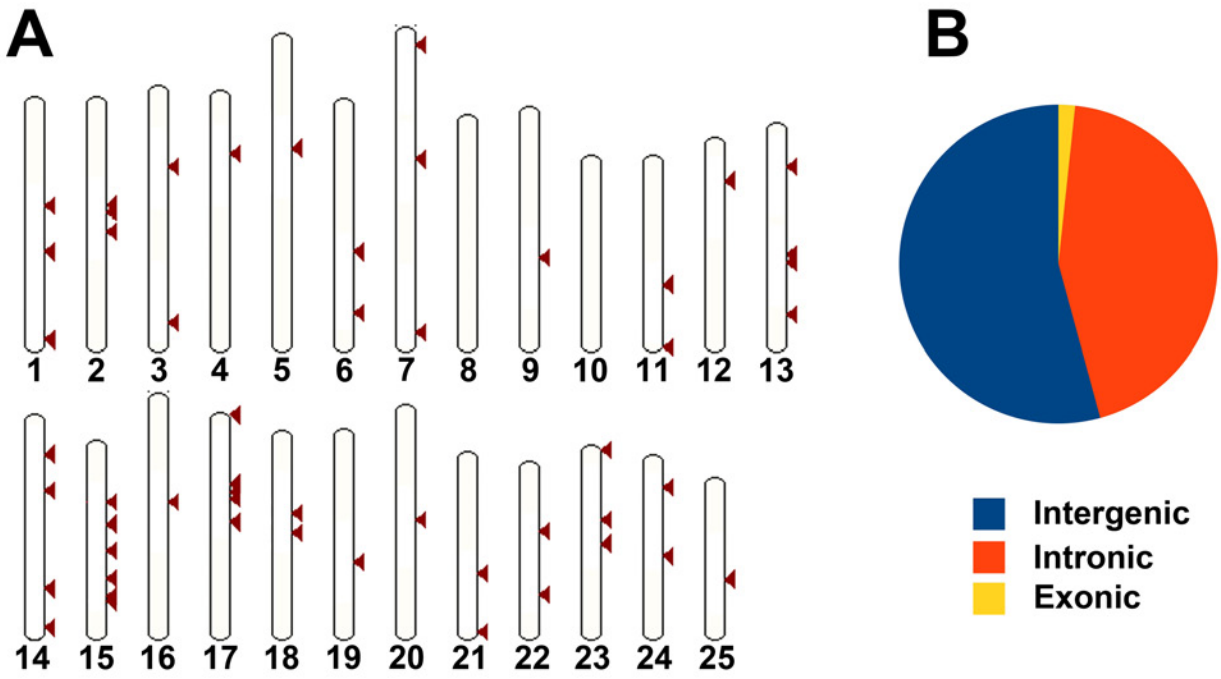
Supplemental Figure 4



Supplemental Figure 4 - Eight ED lines injected with Cre recombinase.

Each line is documented for GFP (first column and third column) and RFP (second column and fourth column) expression, in a Cre injected (second, fourth, sixth and eighth rows; Cre) and not injected (first, third, fifth and seventh rows; No Cre) background. The line ED60 shows a shift of RFP to GFP expression in the forebrain (arrow), when comparing the control (No Cre) with the Cre injected background (Cre). The line ED77 shows a shift of RFP (No Cre) to GFP (Cre) expression in the notochord (arrow). In ED162 a shift of RFP (No Cre) to GFP (Cre) expression is observed in the posterior region of yolk syncytial layer (arrow). The ED185B line shows a shift of RFP (No Cre) to GFP (Cre) expression in the notochord (arrow). ED186 line shows a shift of RFP (No Cre) to GFP (Cre) expression in the central nervous system (arrow) and eye (asterisk). ED219 line shows a shift of RFP (No Cre) to GFP (Cre) expression in the notochord (arrow). In ED270 a shift of RFP (No Cre) to GFP (Cre) expression is observed in the posterior retina (arrow). ED272 line shows a shift of RFP (No Cre) to GFP (Cre) expression in the spinal cord (arrow).

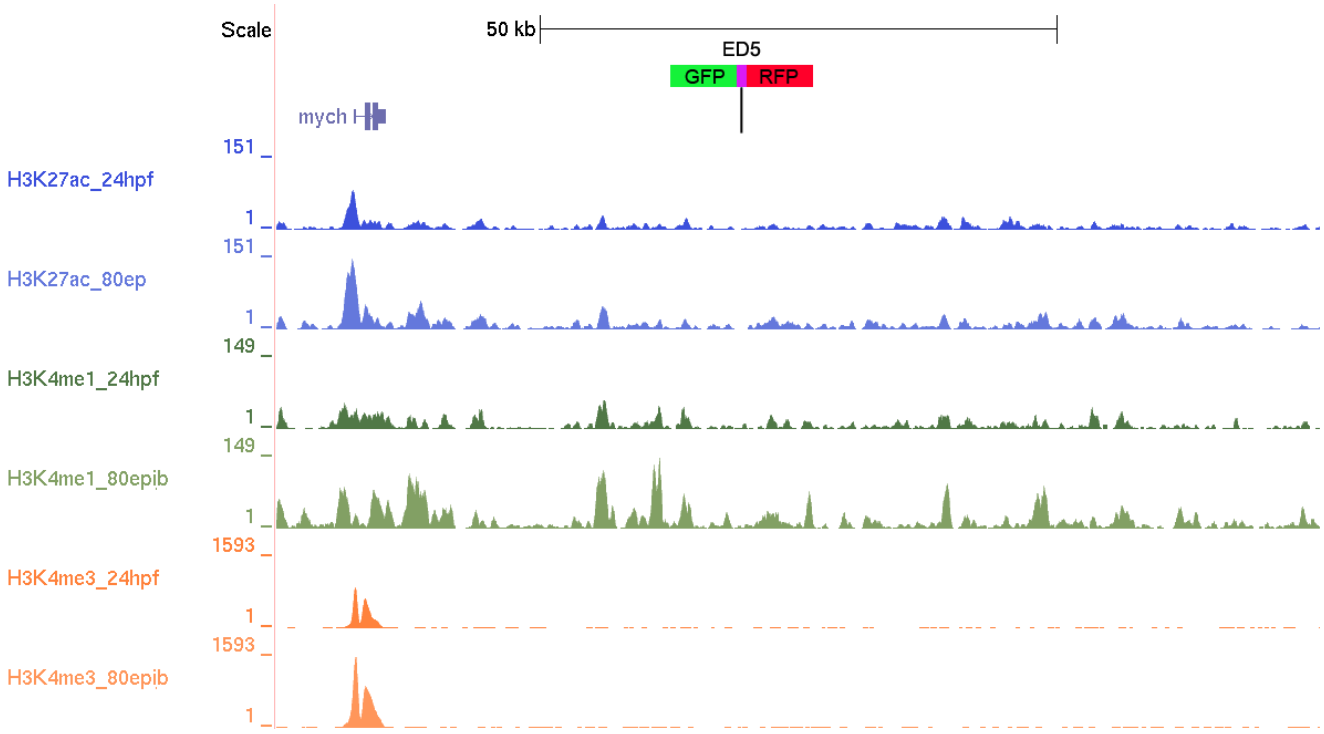
Supplemental Figure 5



Supplemental Figure 5 - ED insertions on the zebrafish genome. (A) Distribution of ED mapped insertions (red triangles) on the zebrafish genome. (B) Distribution of insertions between genes (intergenic, blue; 54%), in introns (orange; 44%) and in exons (yellow; 2%).

Supplemental Figure 6

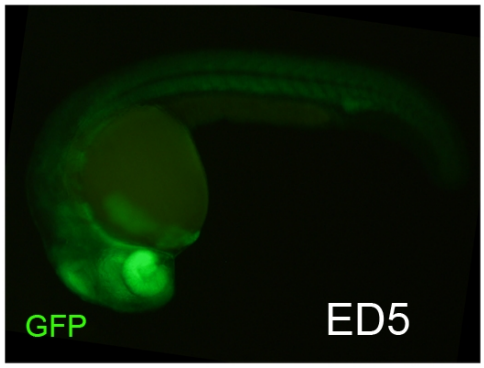
ED5



Genomic Landscape:
chr6:50395584-50496447

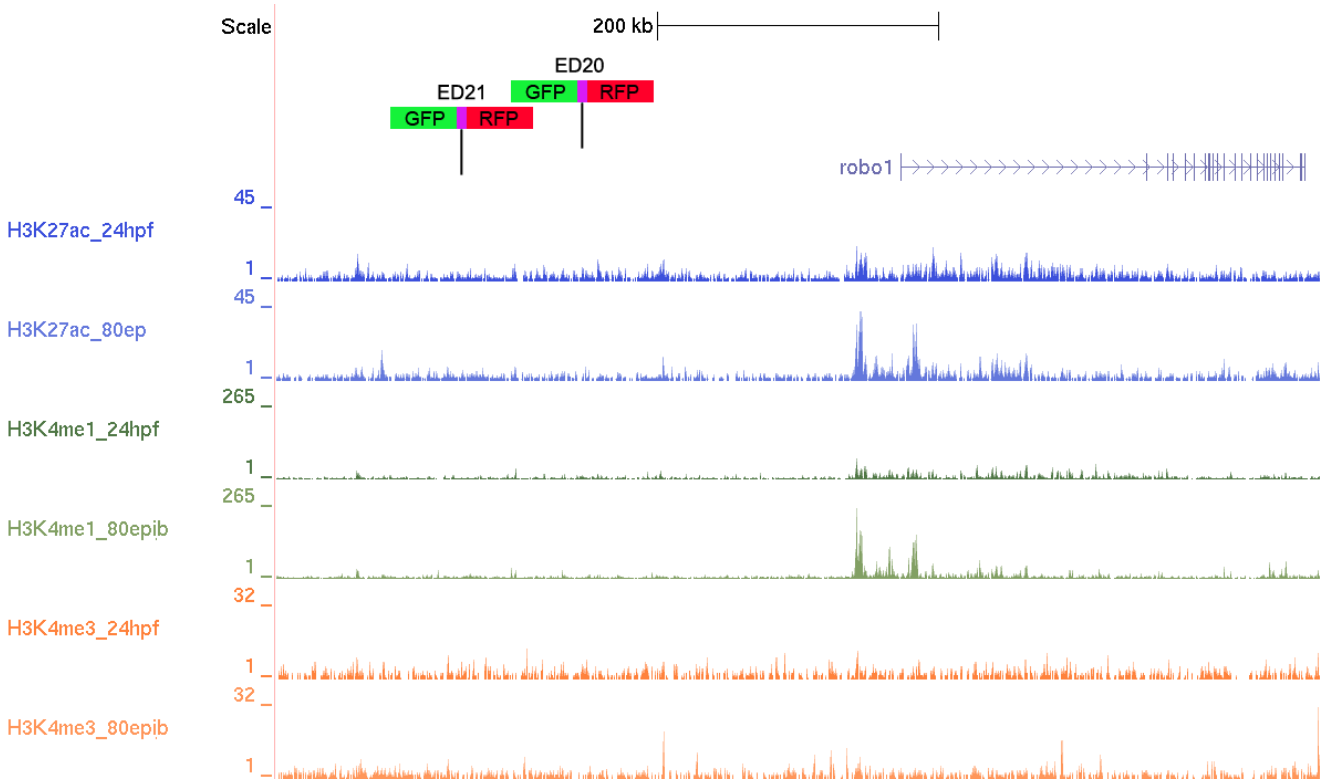
Size of Landscape
100863bp

Associated Gene:
mych



mych

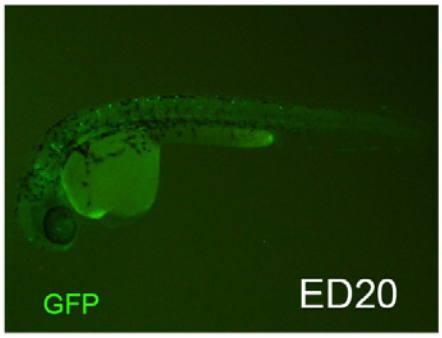
ED20 And ED21

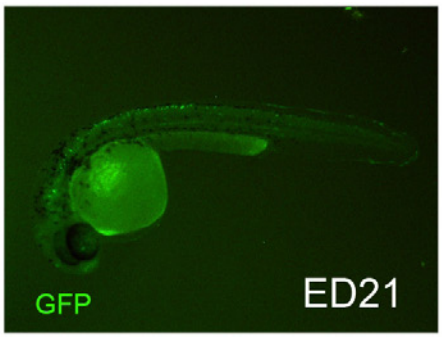


Genomic Landscape:
chr15:37061399-37803435

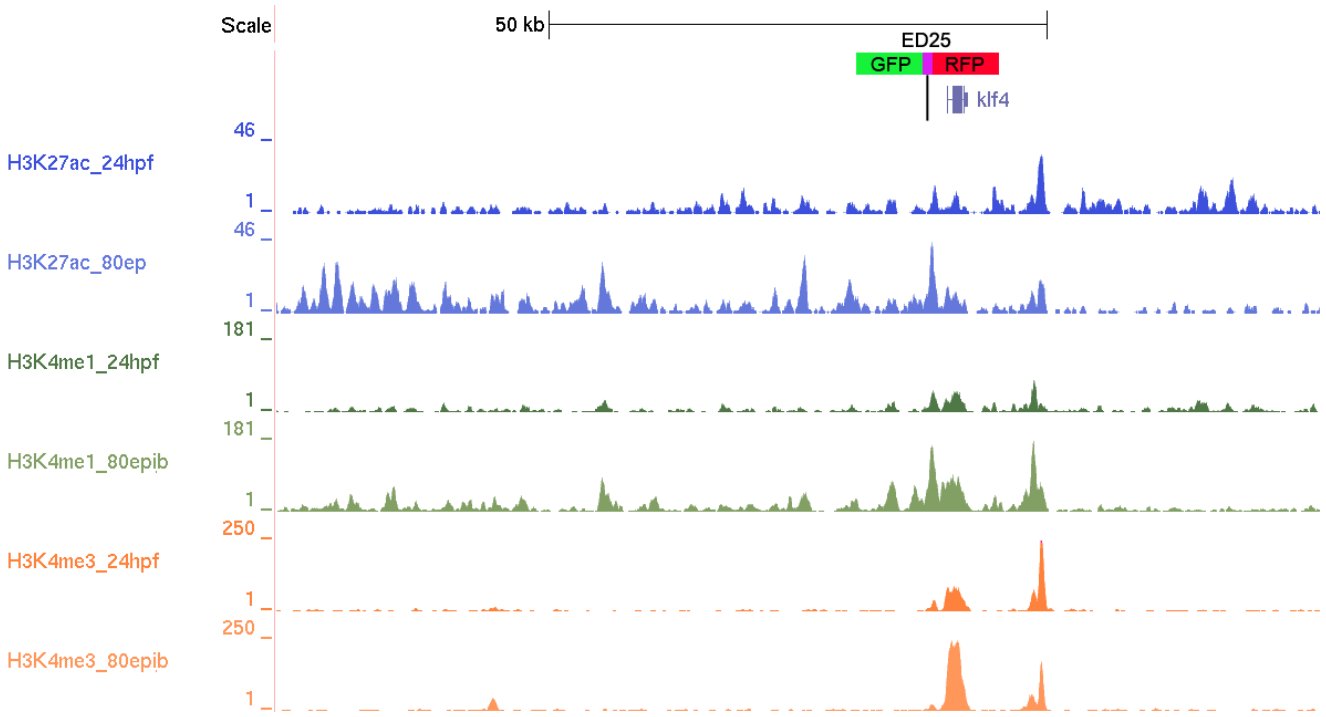
Size of Landscape
742036 bp

Associated Gene:
robo1





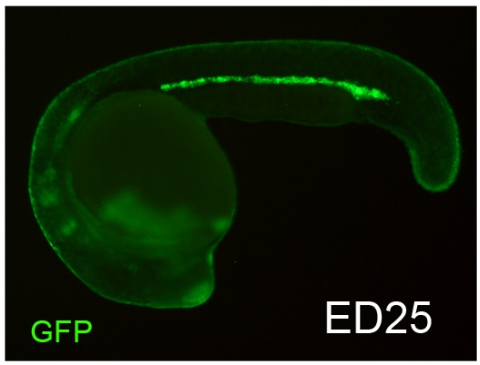
ED25

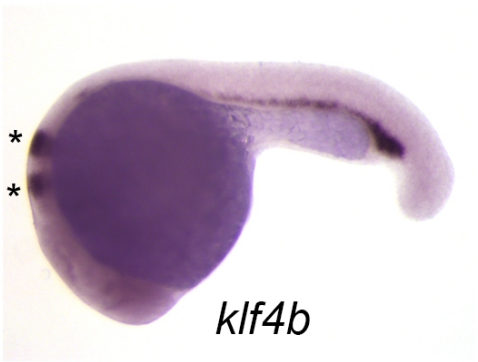


Genomic Landscape:
chr2:32977192-33081797

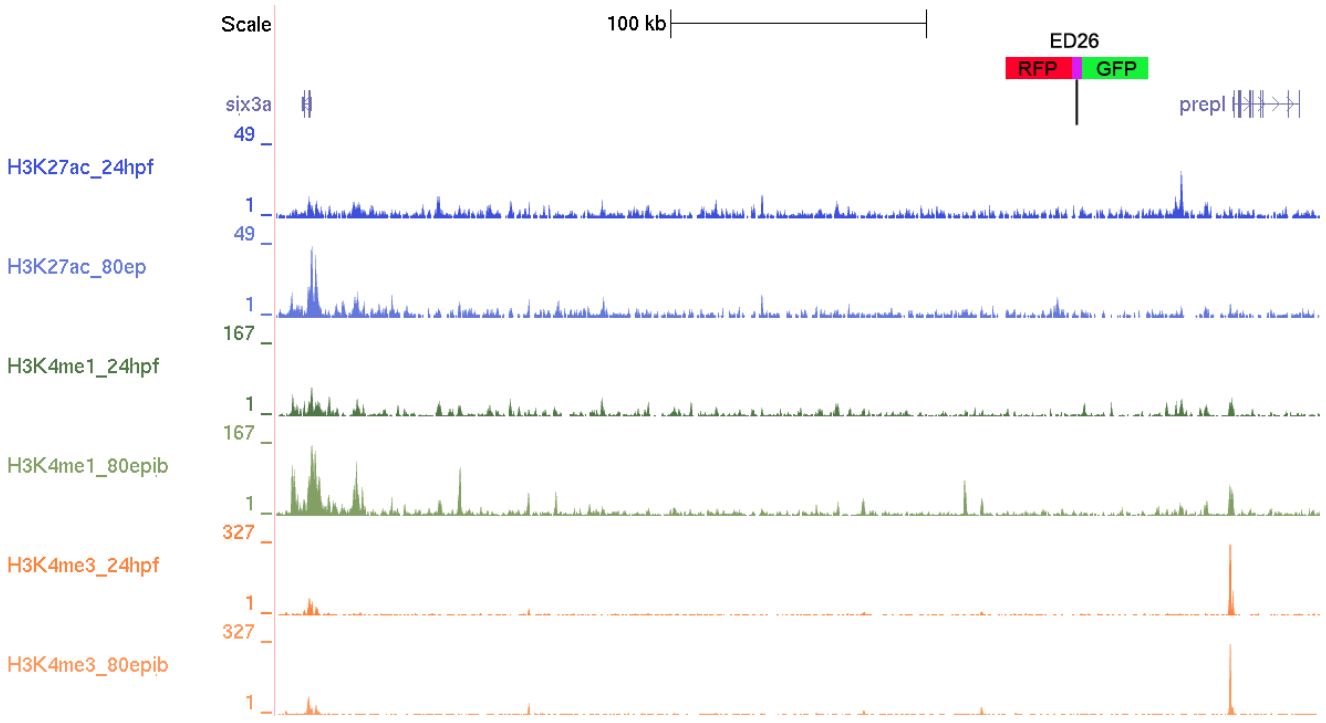
Size of Landscape
104605 bp

Associated Gene:
klf4b





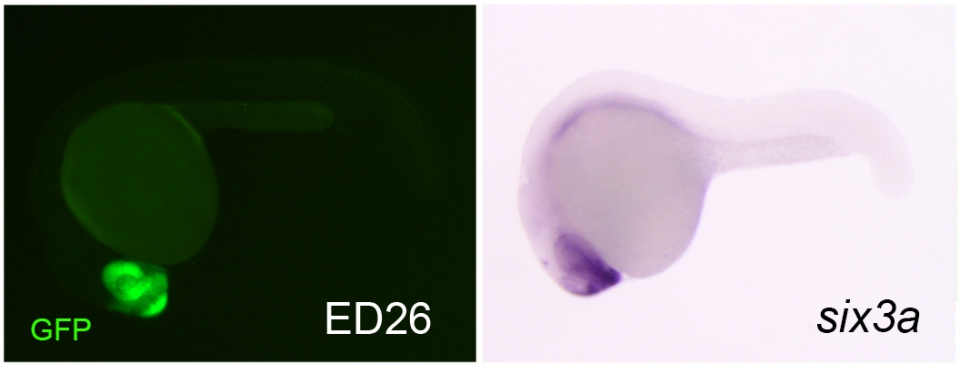
ED26



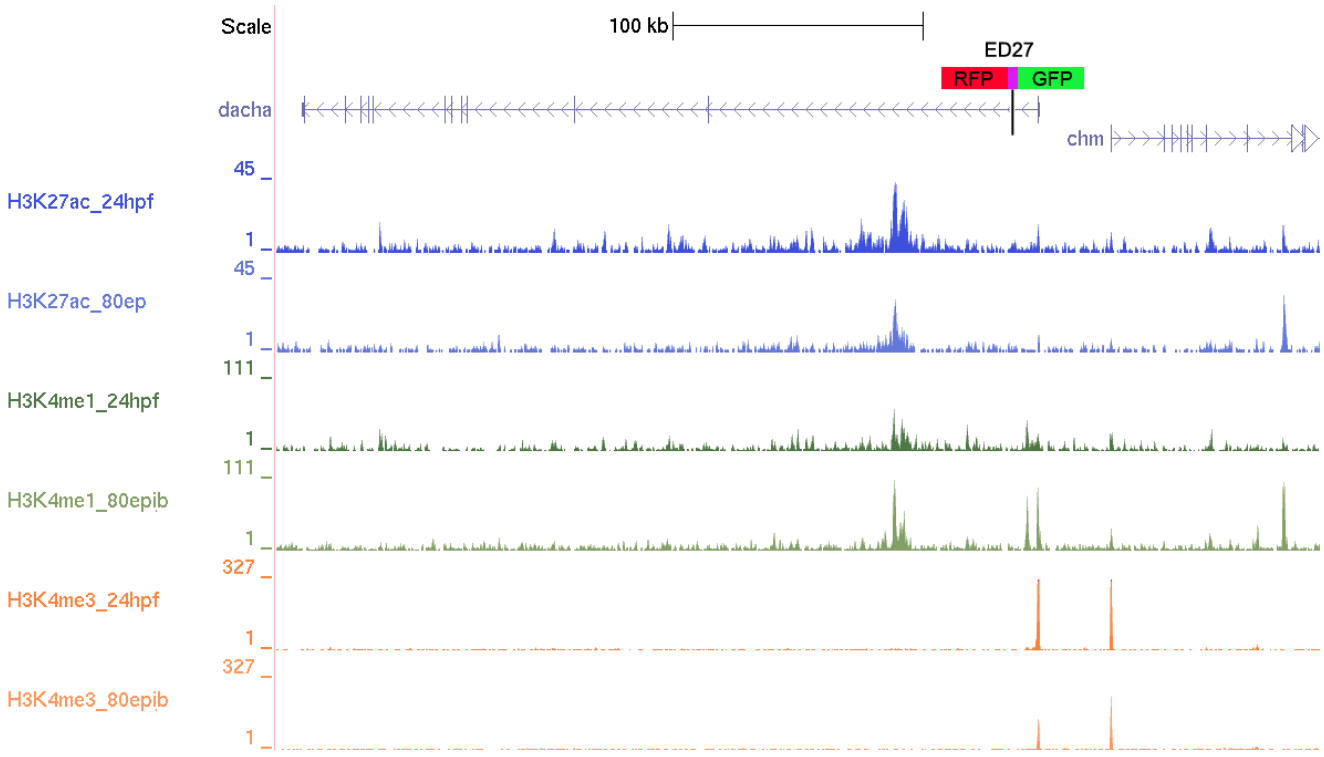
Genomic Landscape:
chr13:9813235-10220458

Size of Landscape
407223 bp

Associated Gene:
six3a



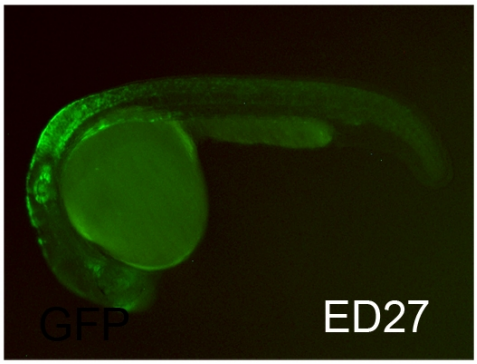
ED27



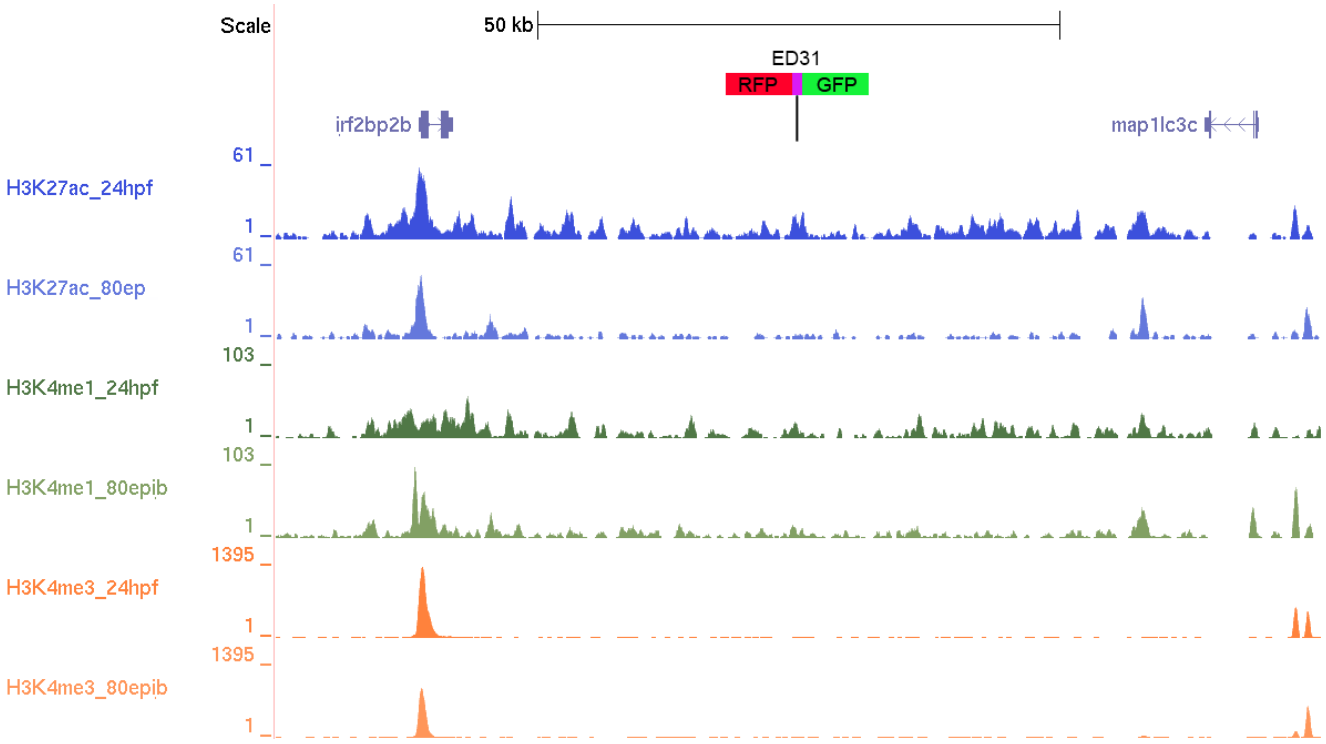
Genomic Landscape:
chr21:33320279-33735792

Size of Landscape
415513 bp

Associated Gene:
dacha



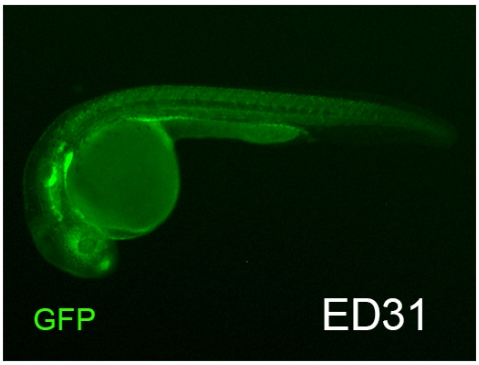
ED31



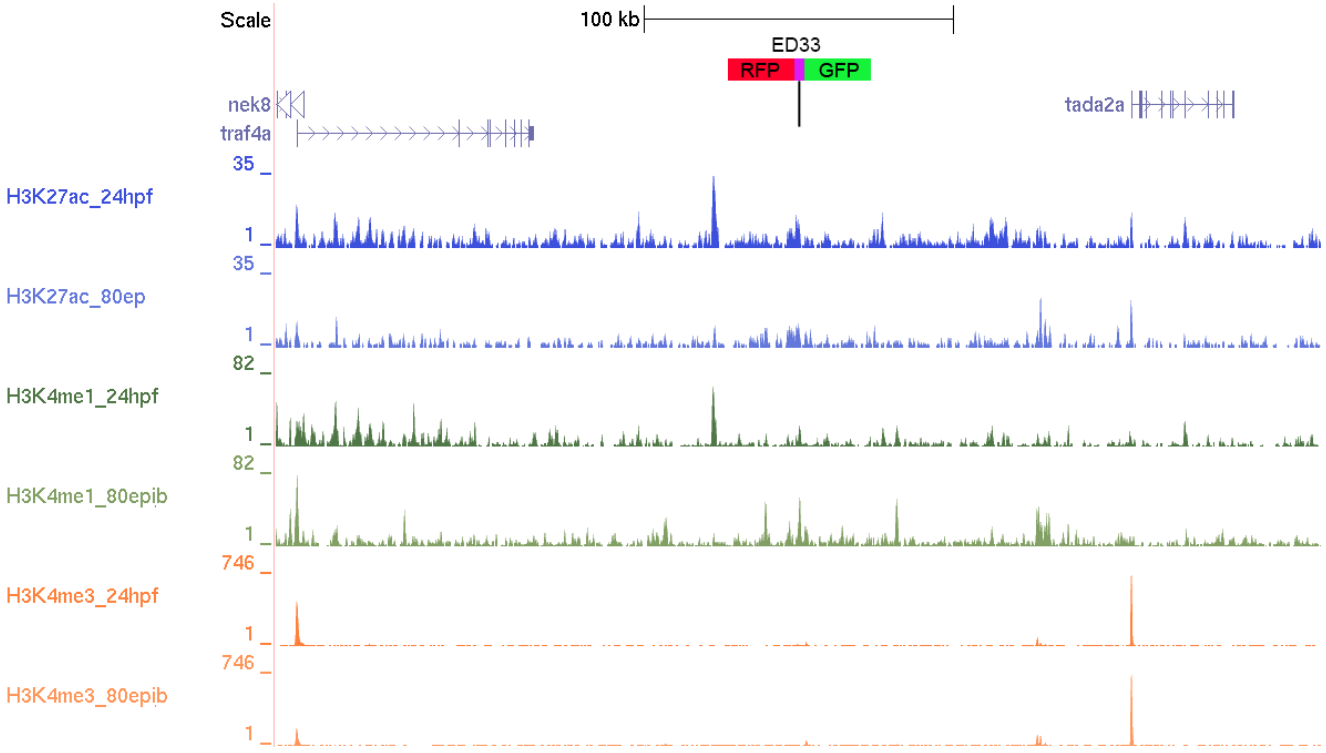
Genomic Landscape:
chr11:45555315-45655324

Size of Landscape
100009 bp

Associated Gene:
irf2bp2b



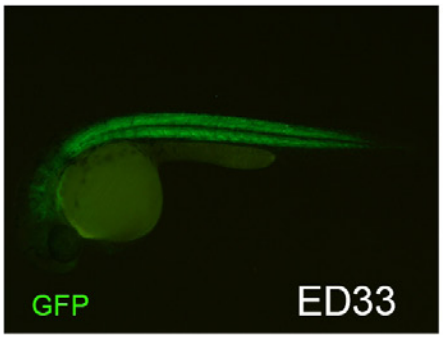
ED33

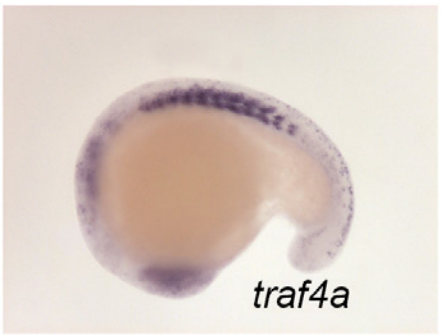


Genomic Landscape:
chr15:14528990-14866507

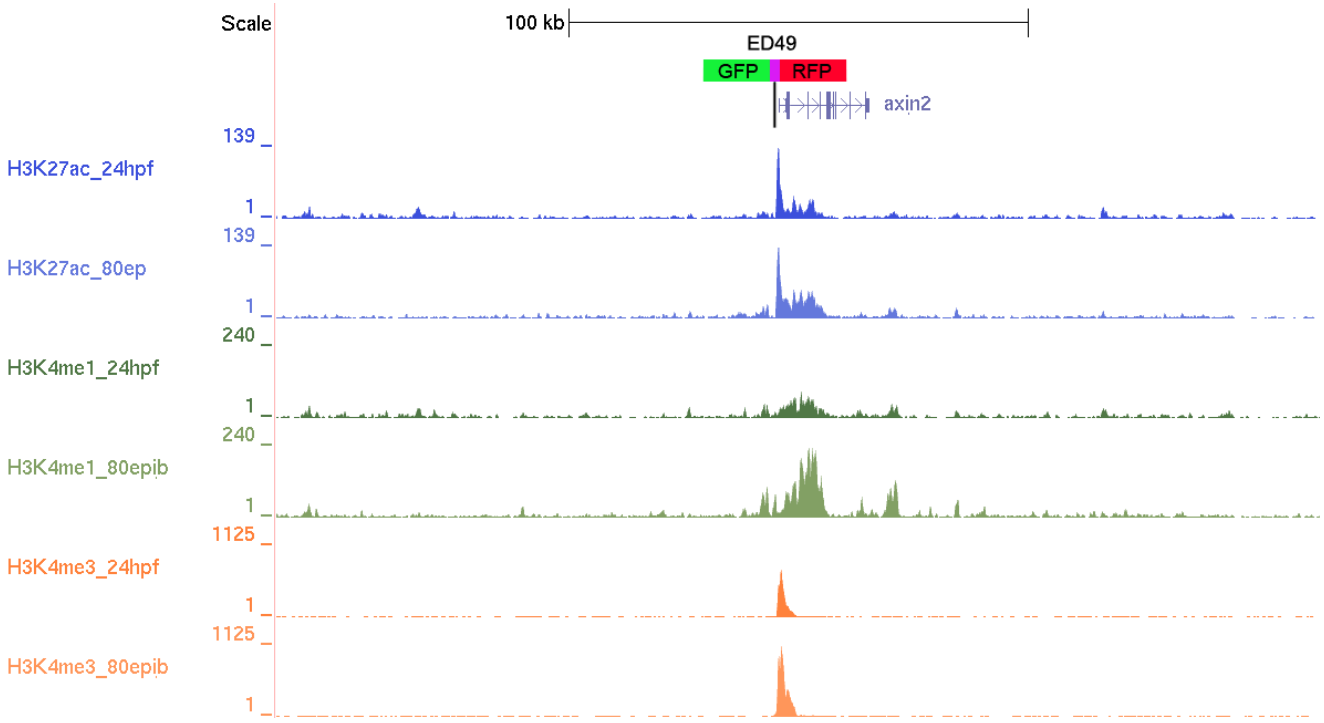
Size of Landscape
337517bp

Associated Gene:
traf4a





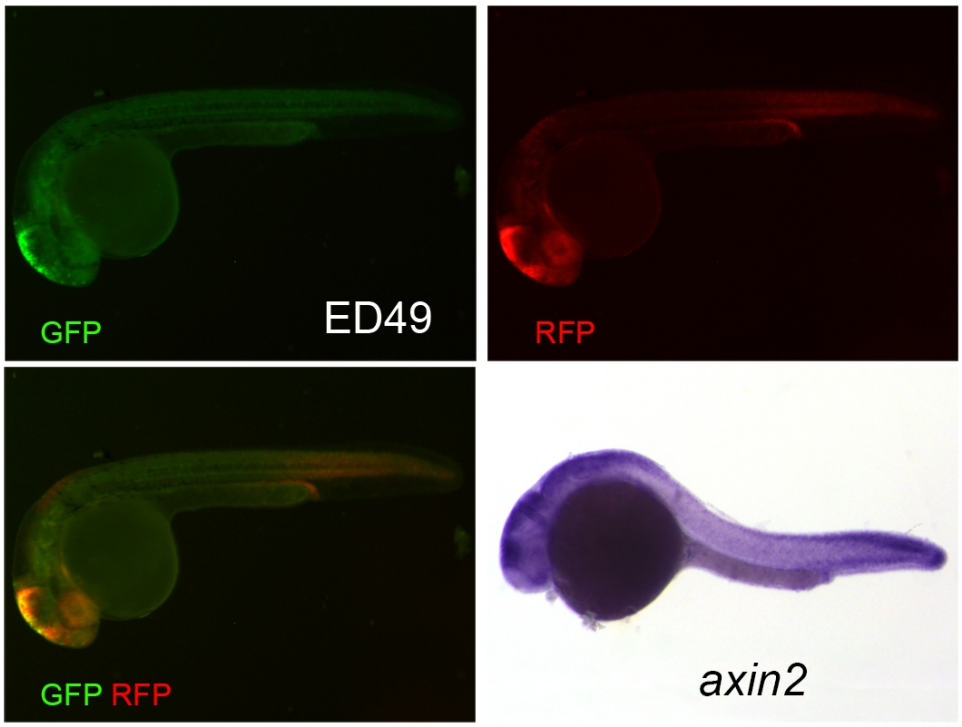
ED49



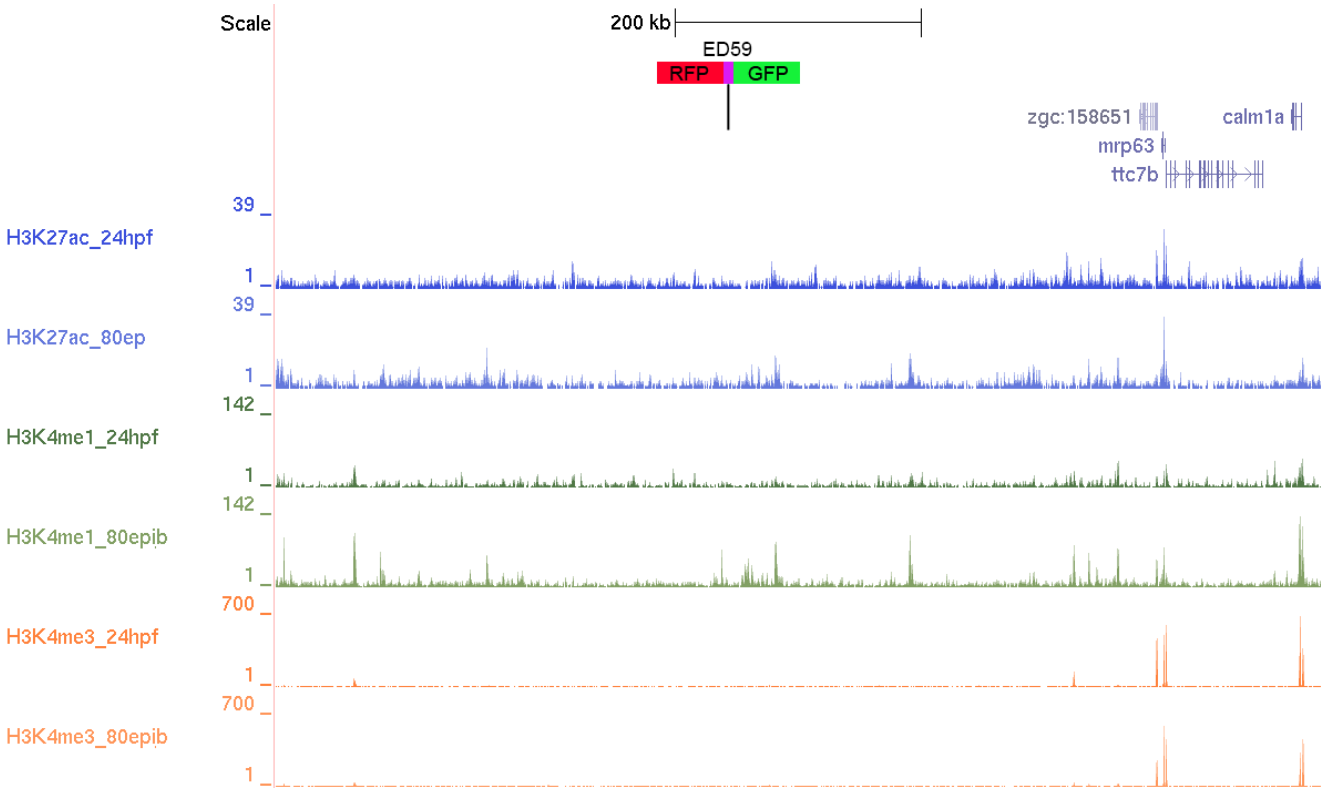
Genomic Landscape:
chr3:56507796-56734626

Size of Landscape
226830 bp

Associated Gene:
axin2



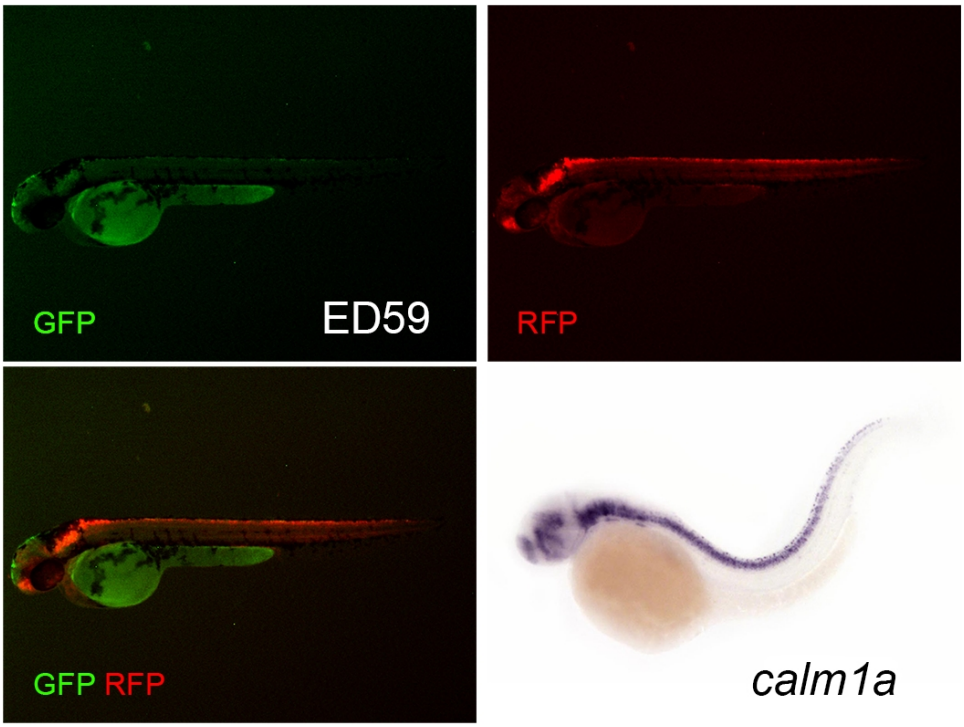
ED59



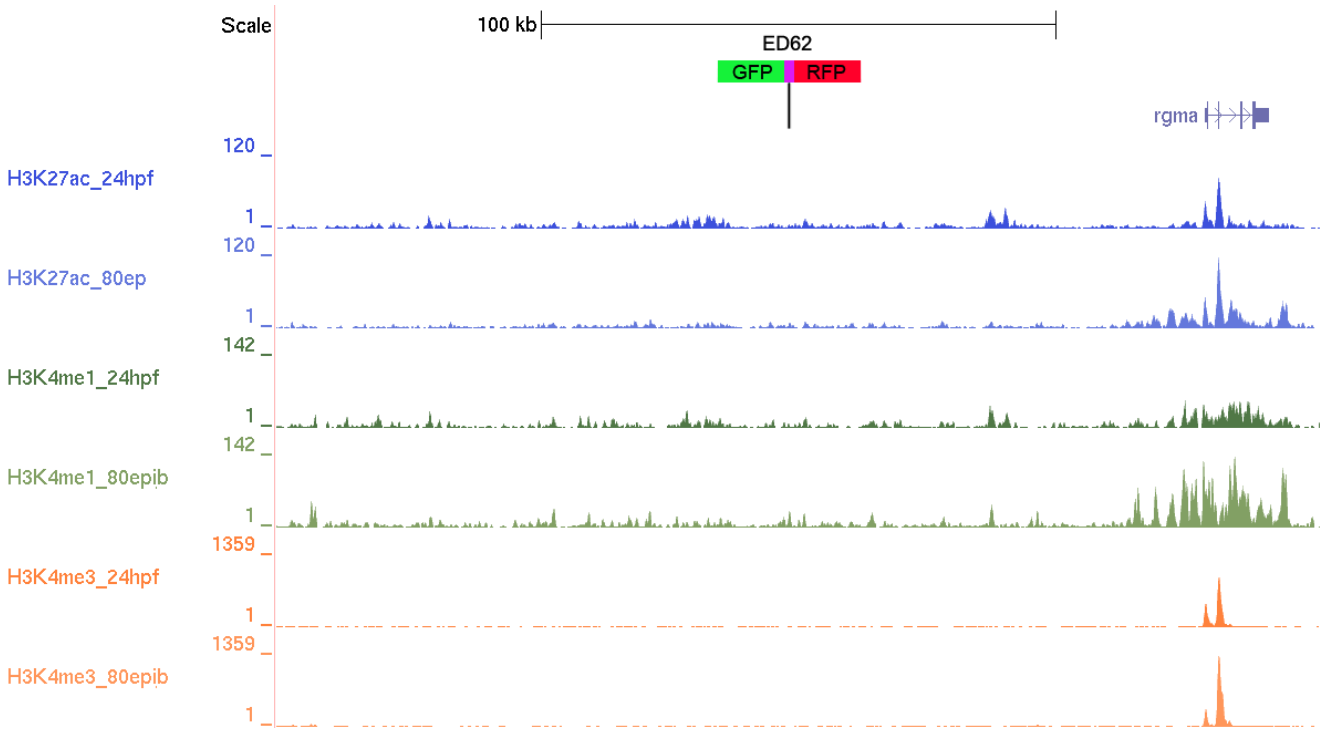
Genomic Landscape:
chr17:25732569-26579075

Size of Landscape
846506 bp

Associated Gene:
calm1a



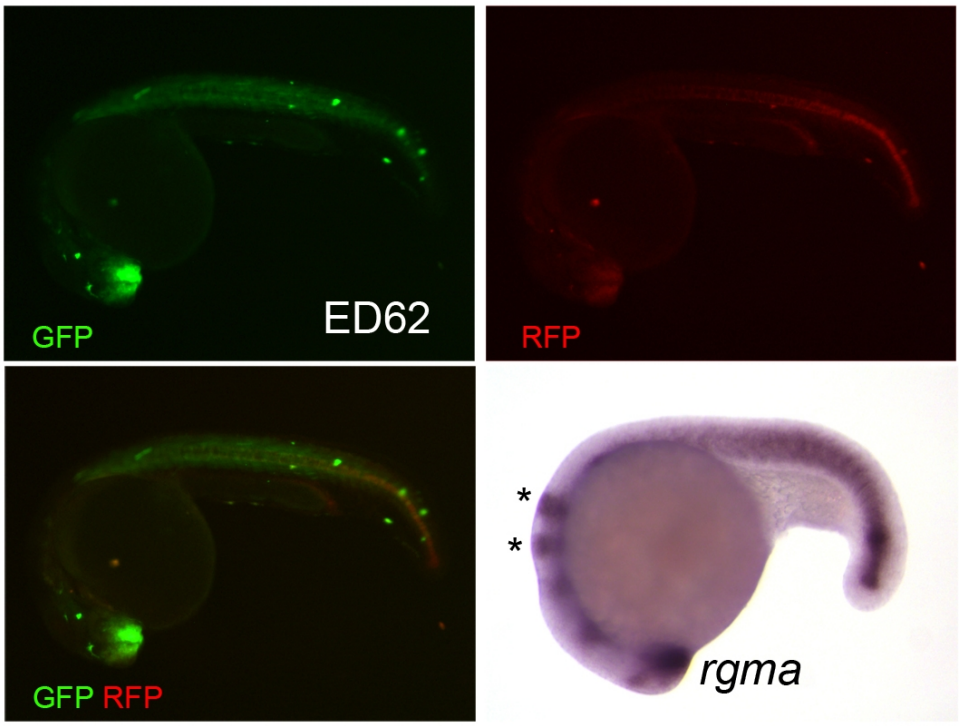
ED62



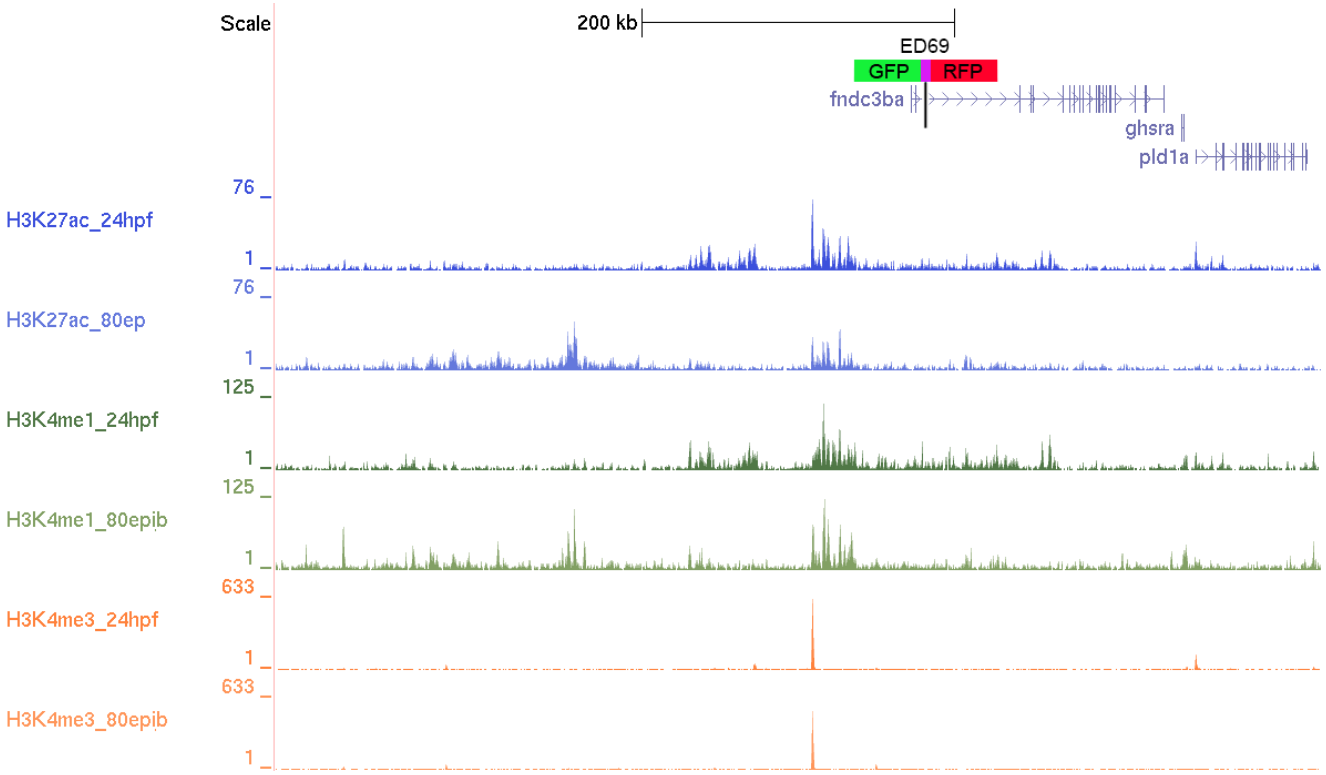
Genomic Landscape:
chr18:24522878-24725643

Size of Landscape
202765 bp

Associated Gene:
rgma



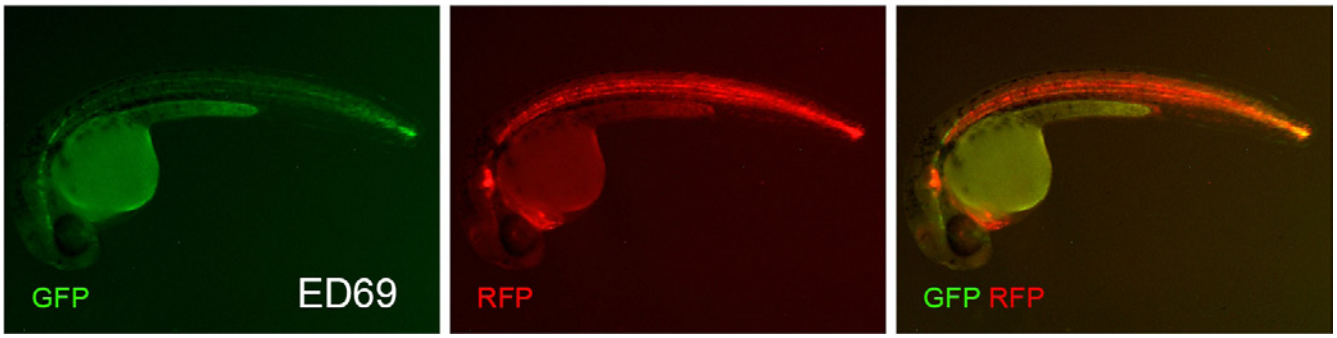
ED69



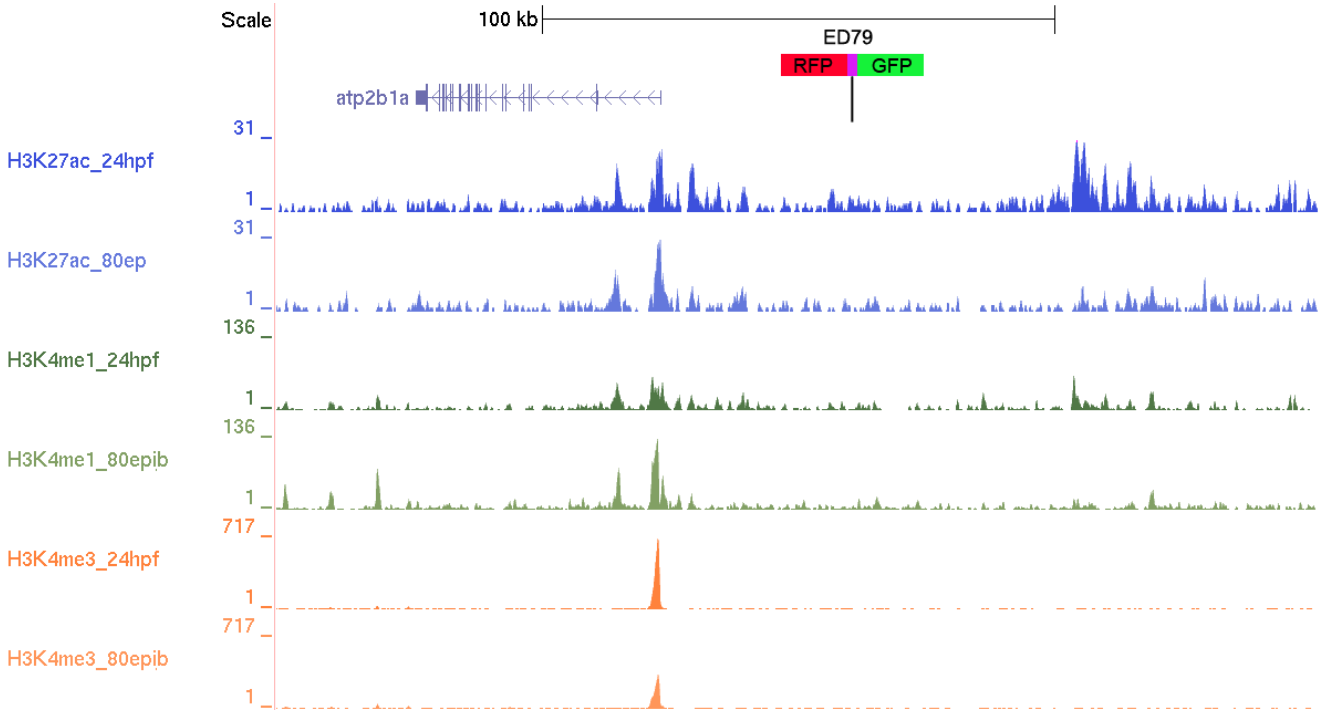
Genomic Landscape:
chr2:25121146-25788771

Size of Landscape
667625 bp

Associated Gene:
pld1a



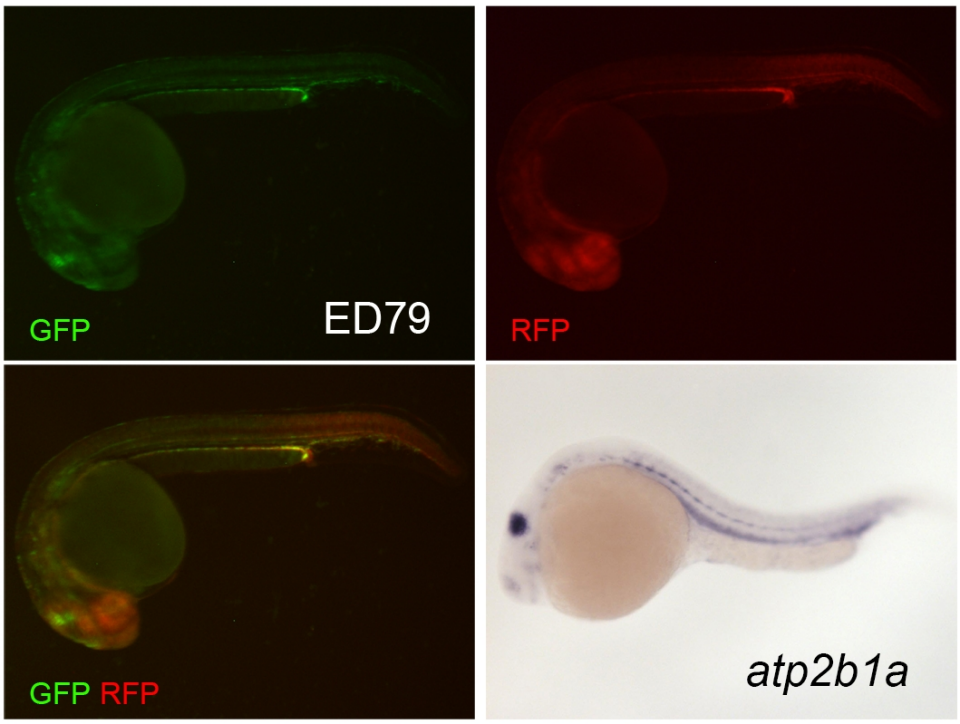
ED79



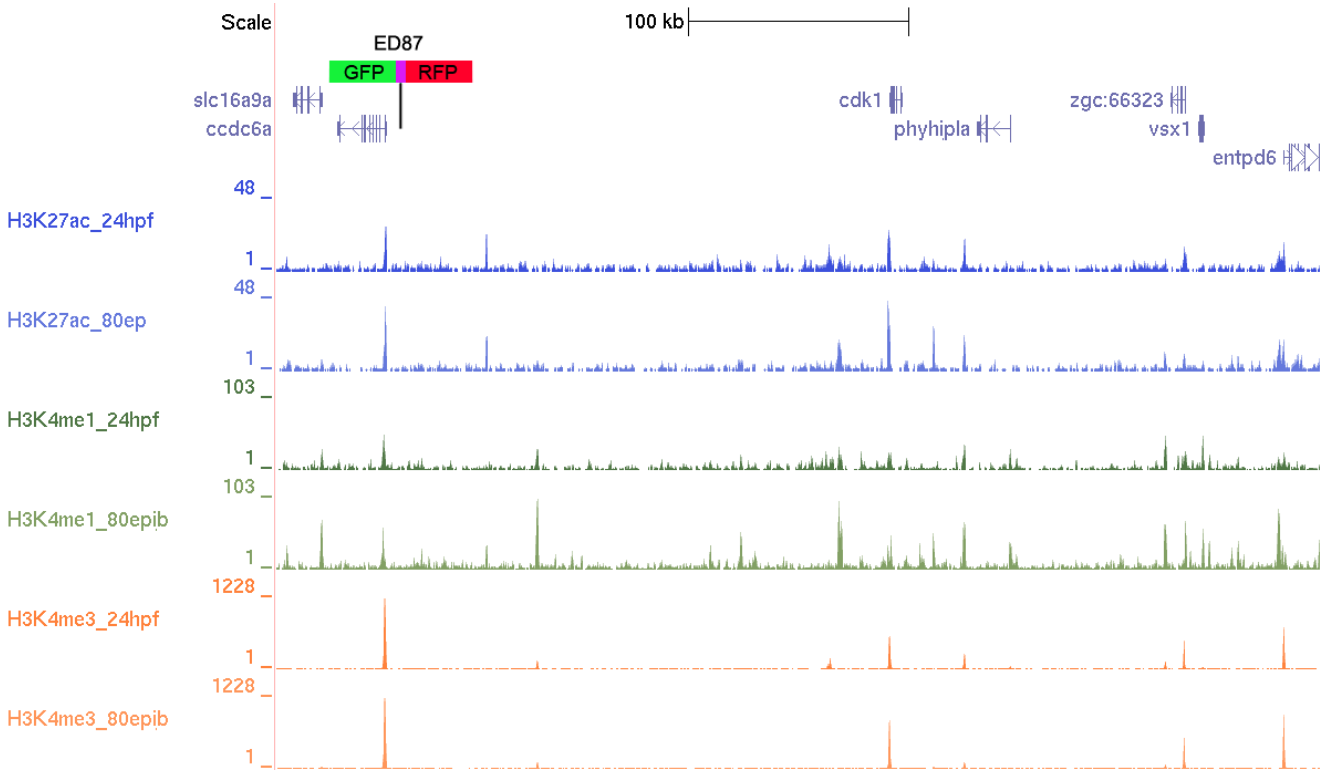
Genomic Landscape:
chr4:15114897-15318480

Size of Landscape
203583 bp

Associated Gene:
atp2b1a



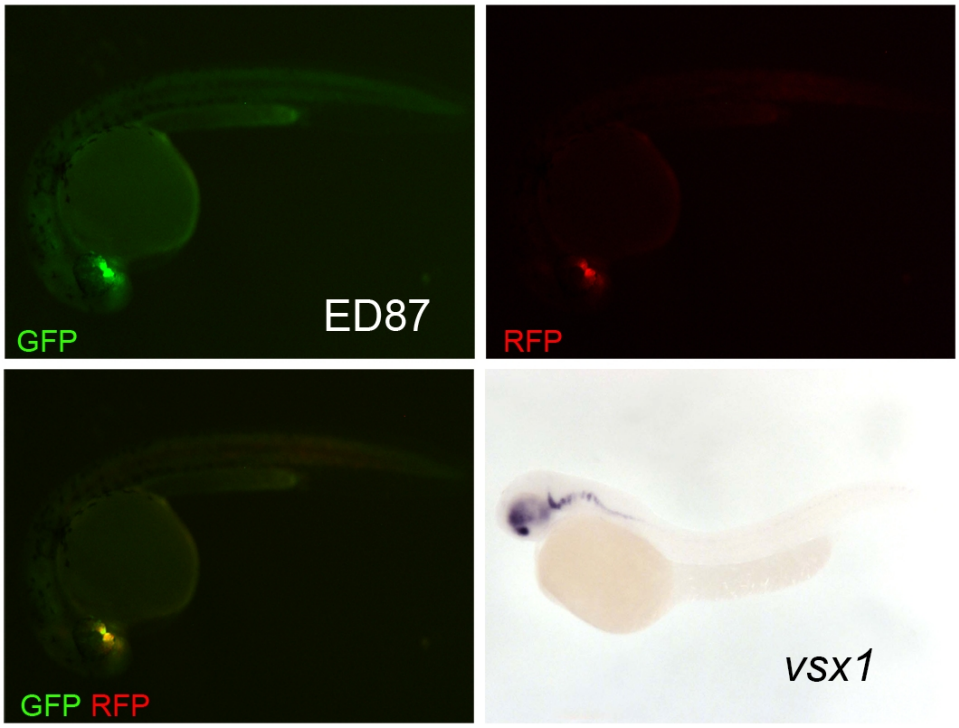
ED87



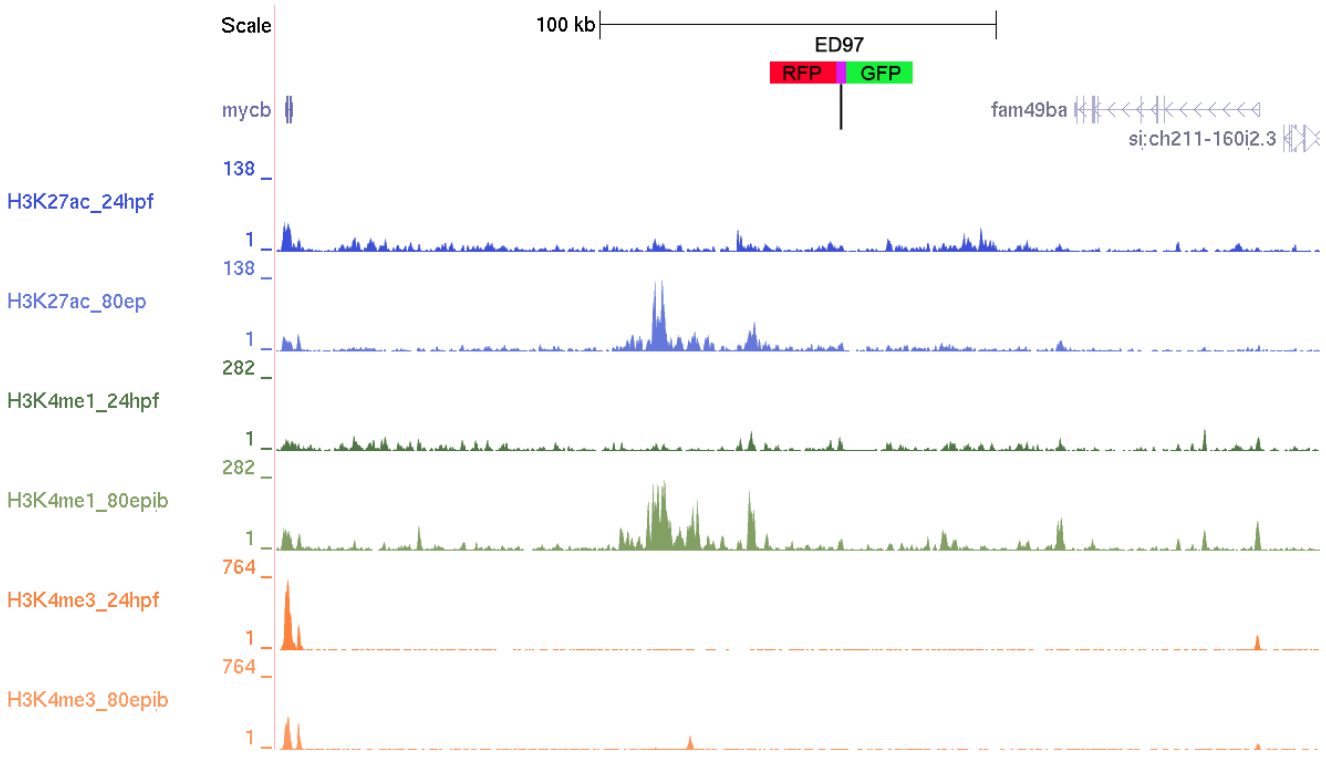
Genomic Landscape:
chr17:20476343-20948904

Size of Landscape
472561 bp

Associated Gene:
vsx1



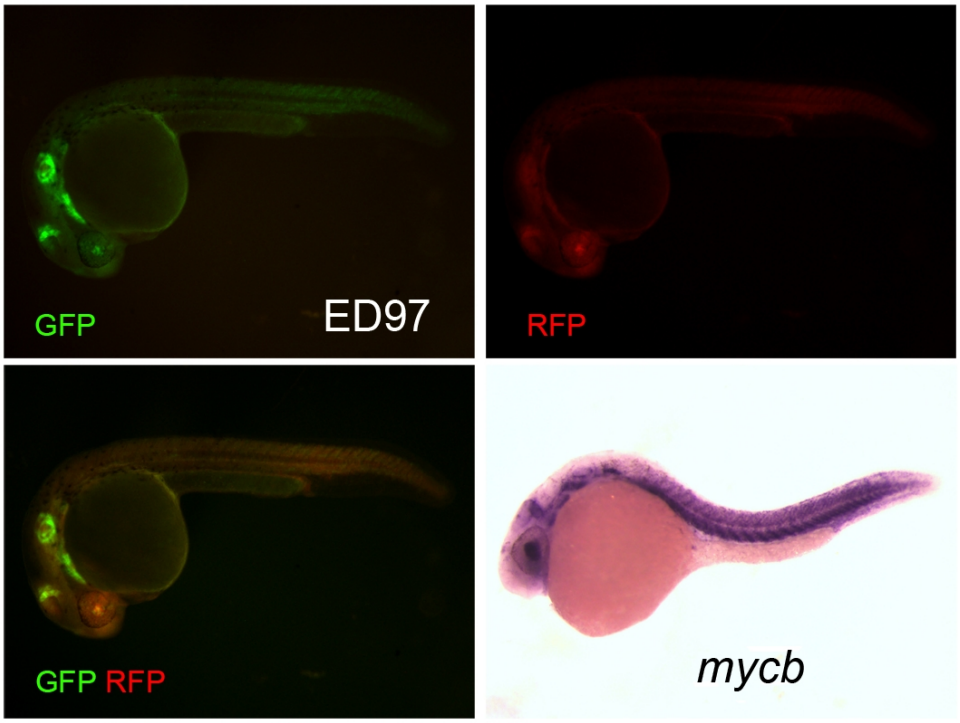
ED97



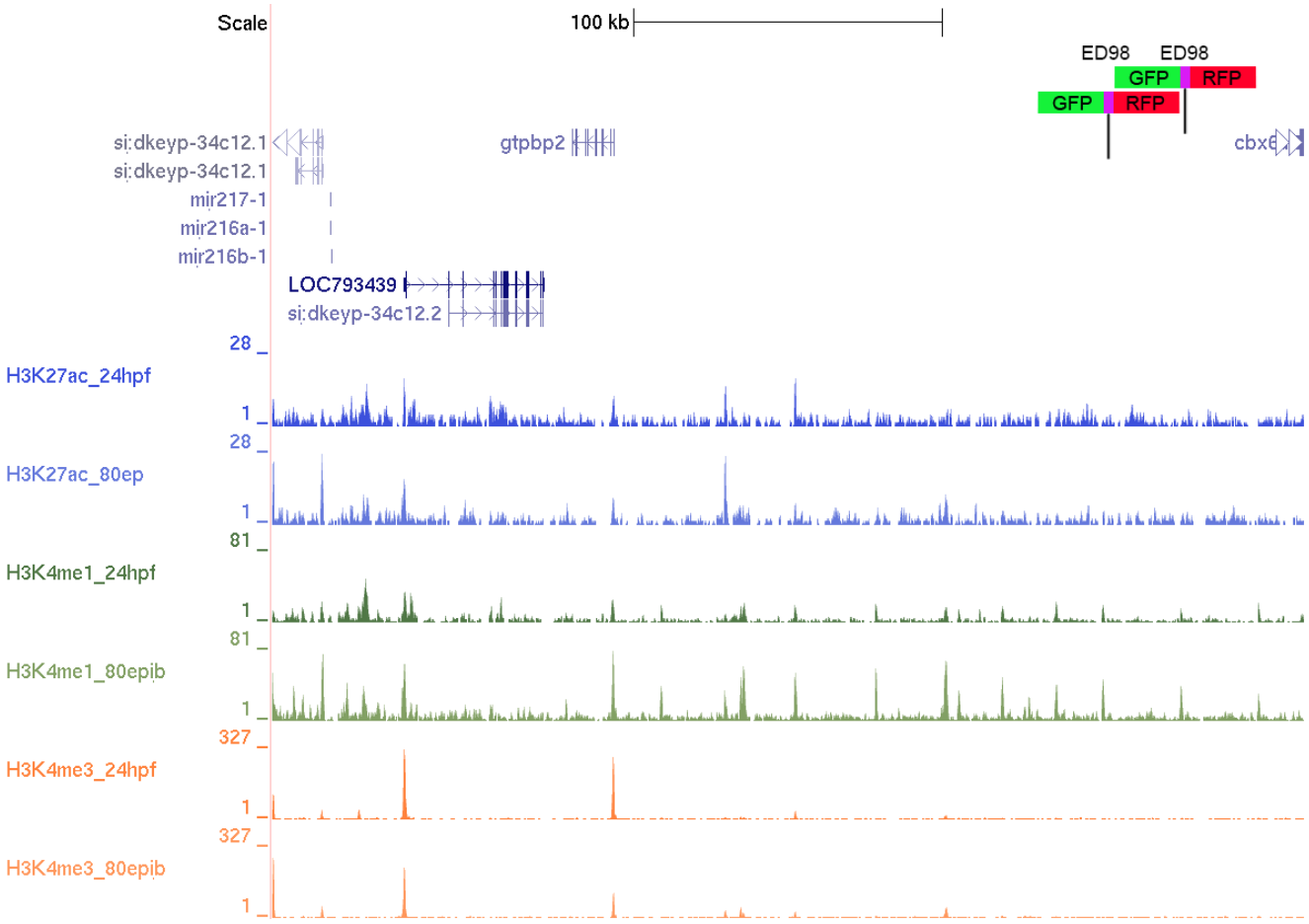
Genomic Landscape:
chr2:31732031-31995185

Size of Landscape
263154 bp

Associated Gene:
mycb



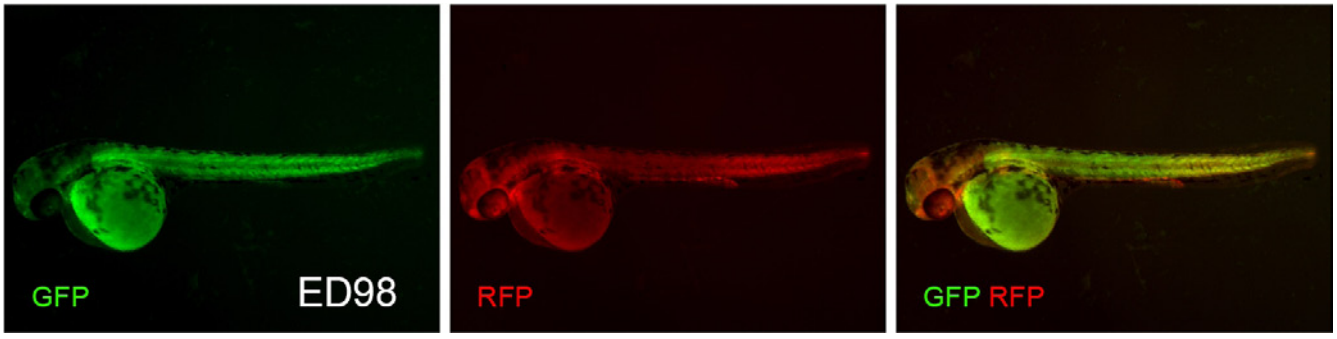
ED98



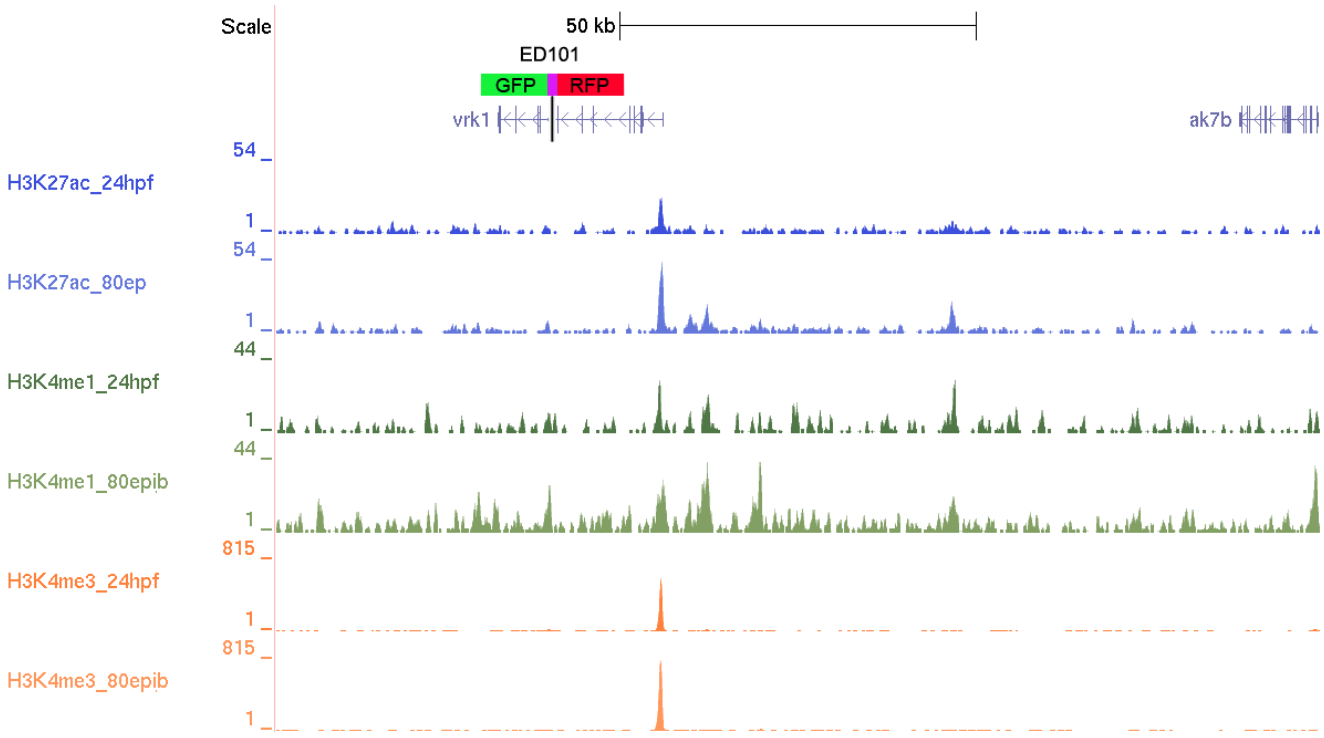
Genomic Landscape:
chr22:31317812-31652128

Size of Landscape
334316 bp

Associated Gene:
mir216a-1



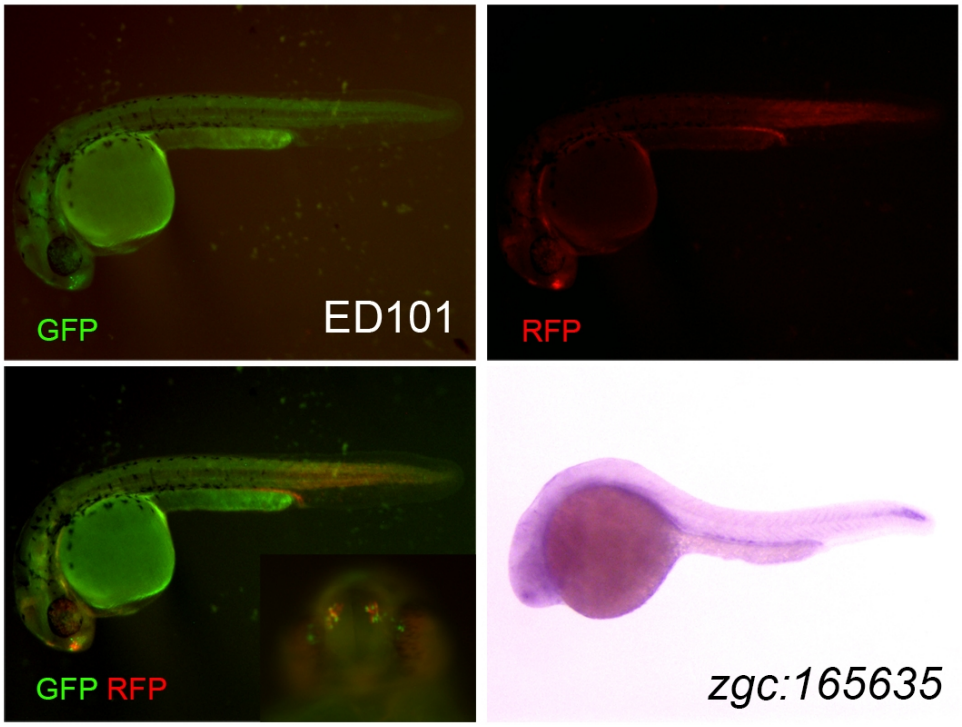
ED101



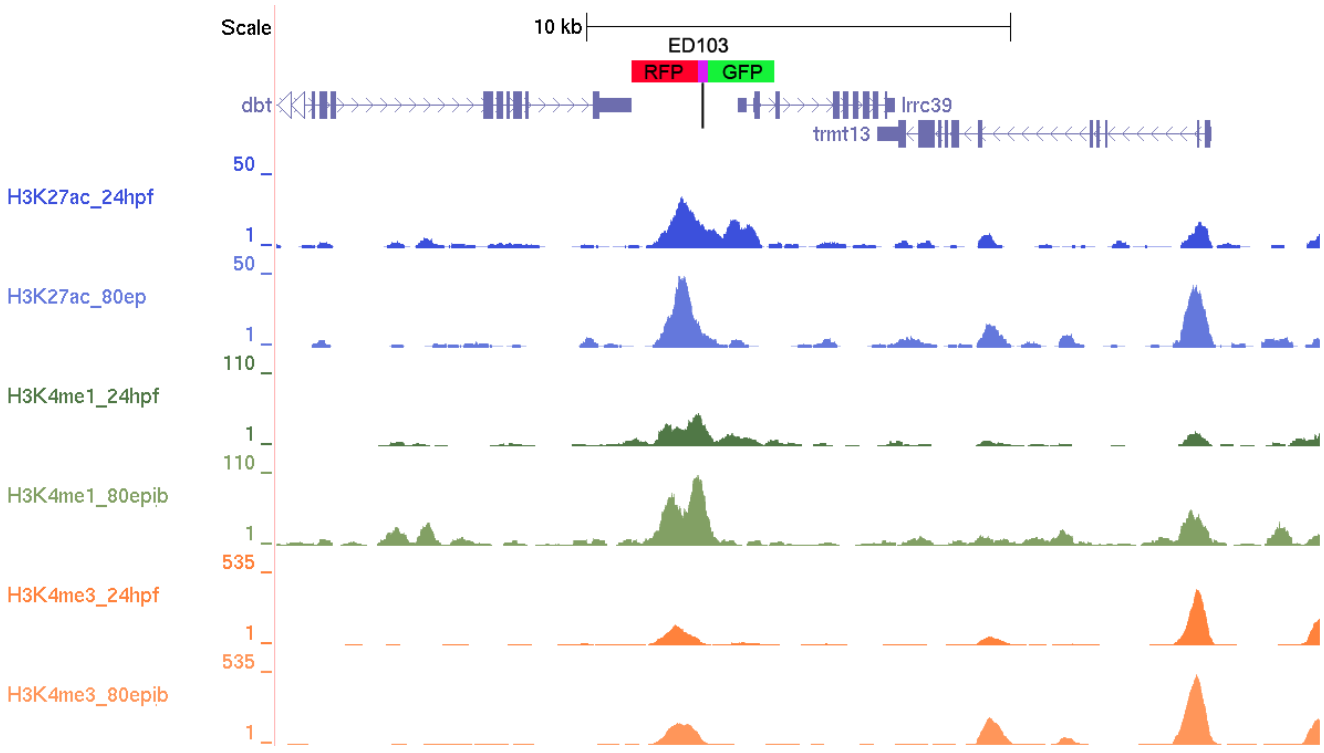
Genomic Landscape:
chr17:18573131-18719447

Size of Landscape
146316 bp

Associated Gene:
zgc:165635



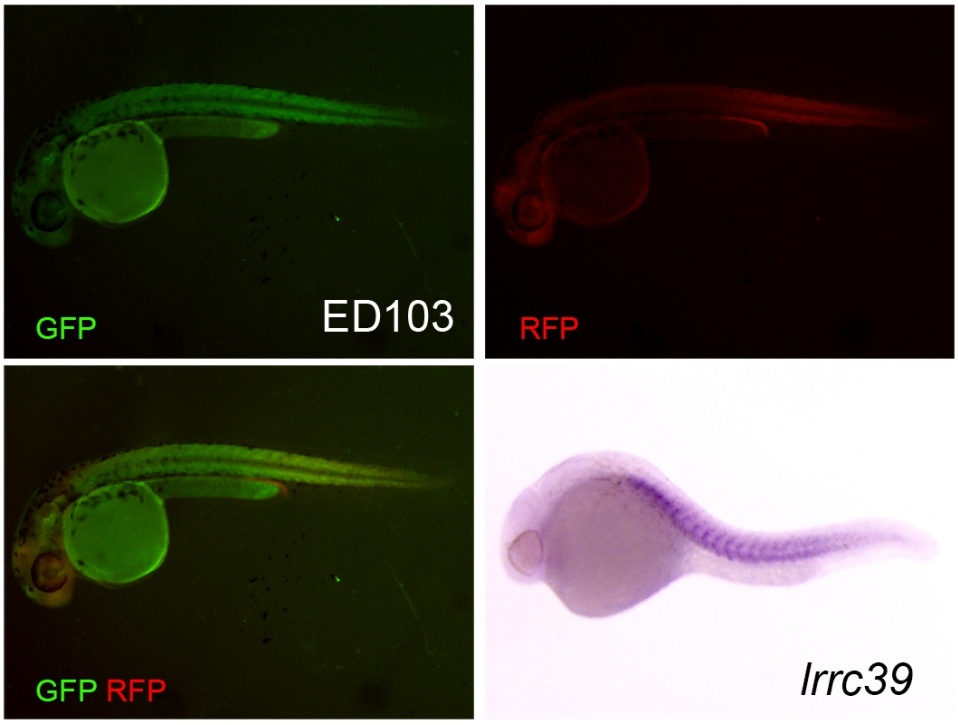
ED103



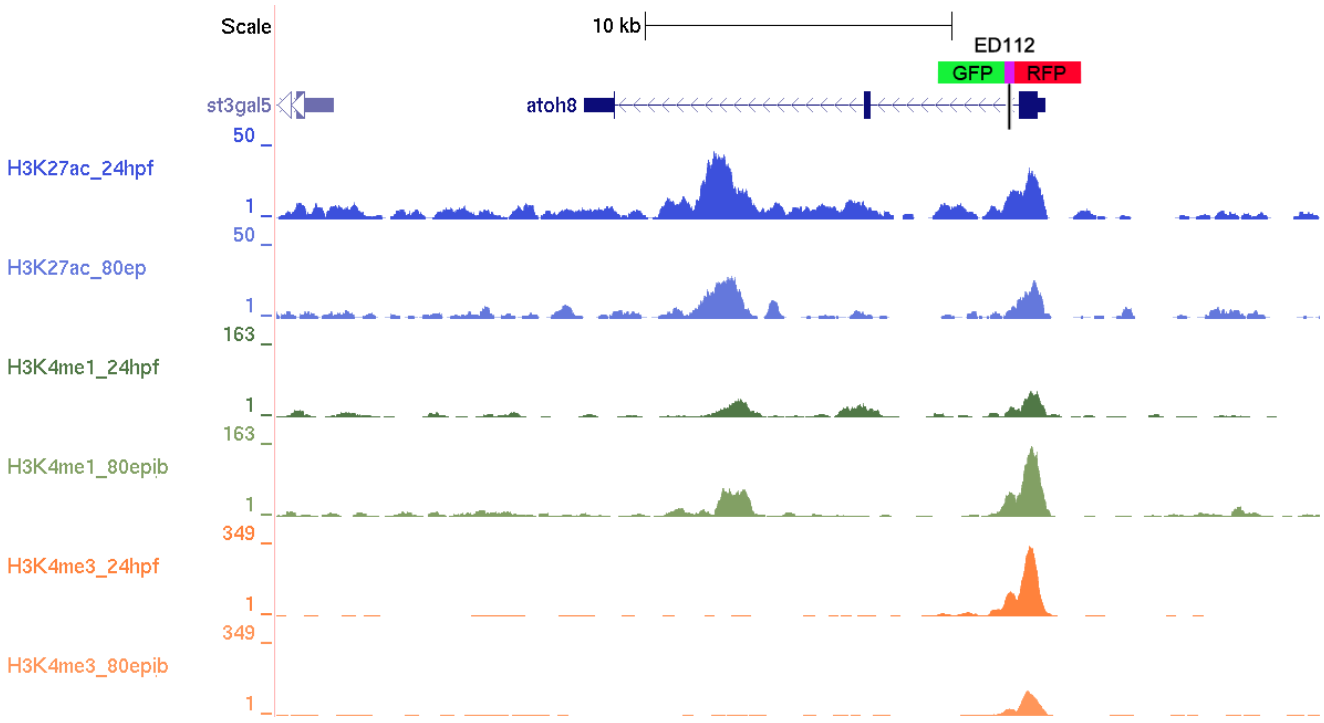
Genomic Landscape:
chr22:16511237-16535821

Size of Landscape
24584 bp

Associated Gene:
lrrc39



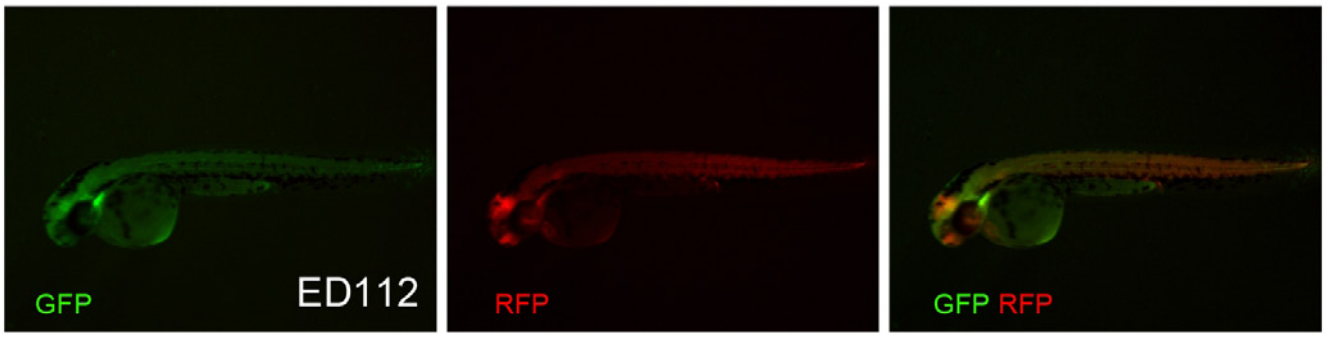
ED112



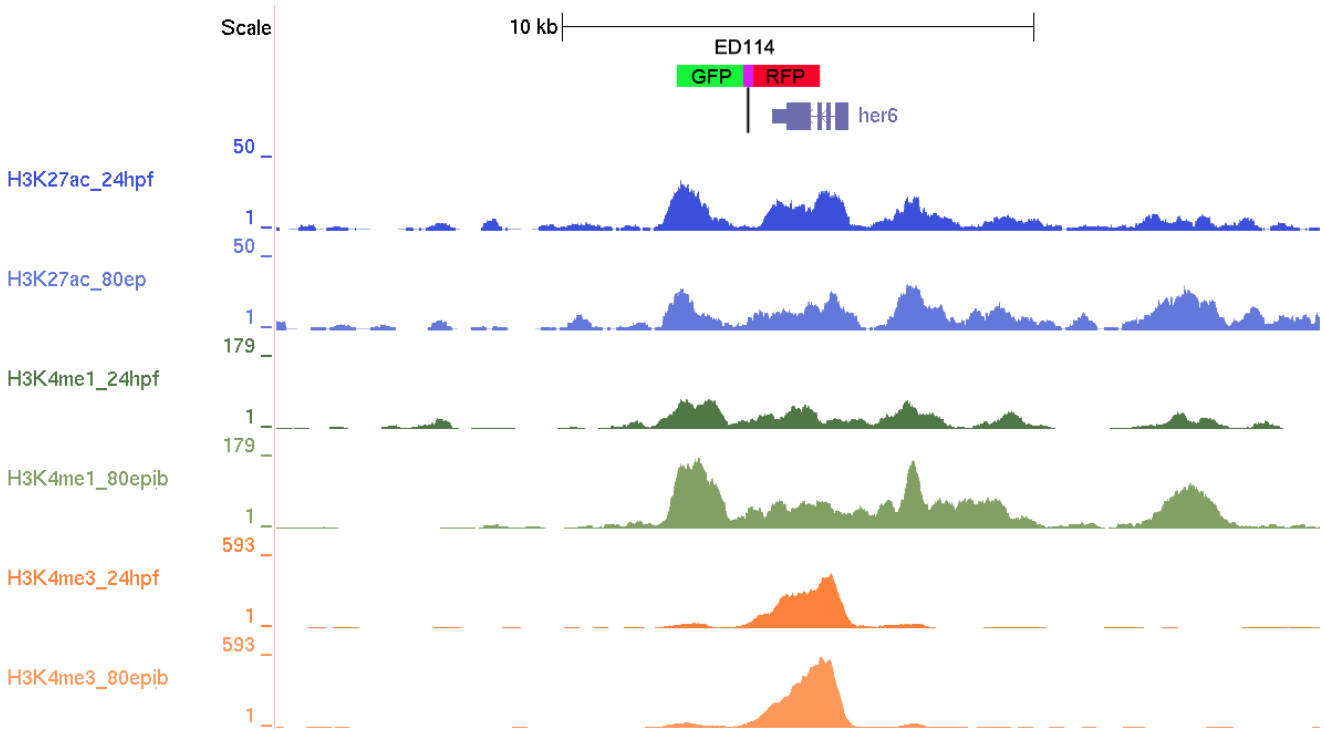
Genomic Landscape:
chr14:9669475-9703440

Size of Landscape
33965 bp

Associated Gene:
atoh8



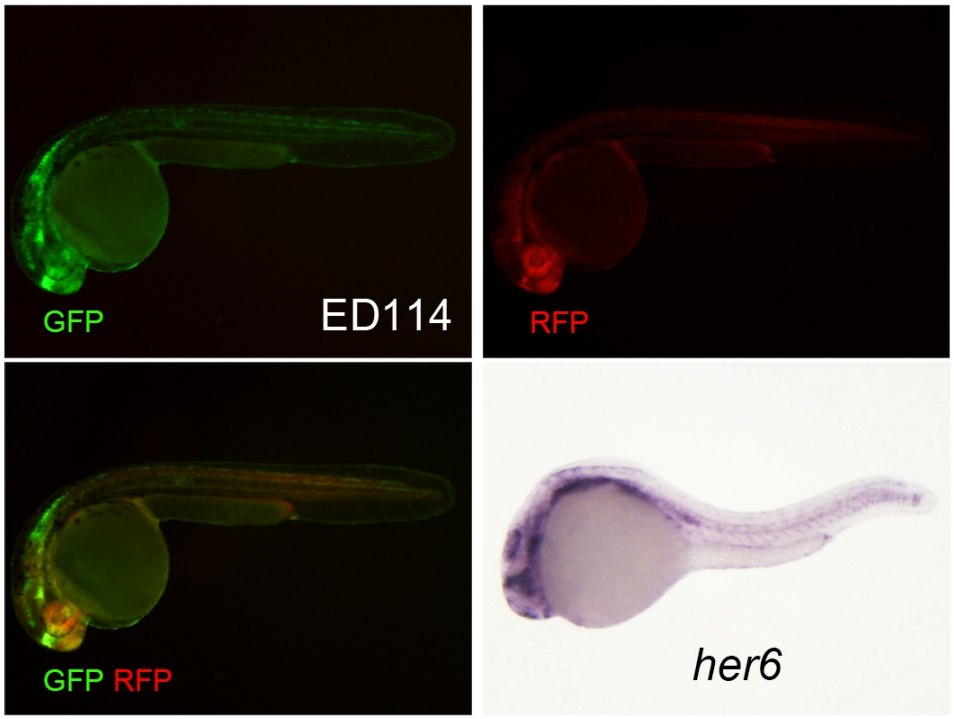
ED114



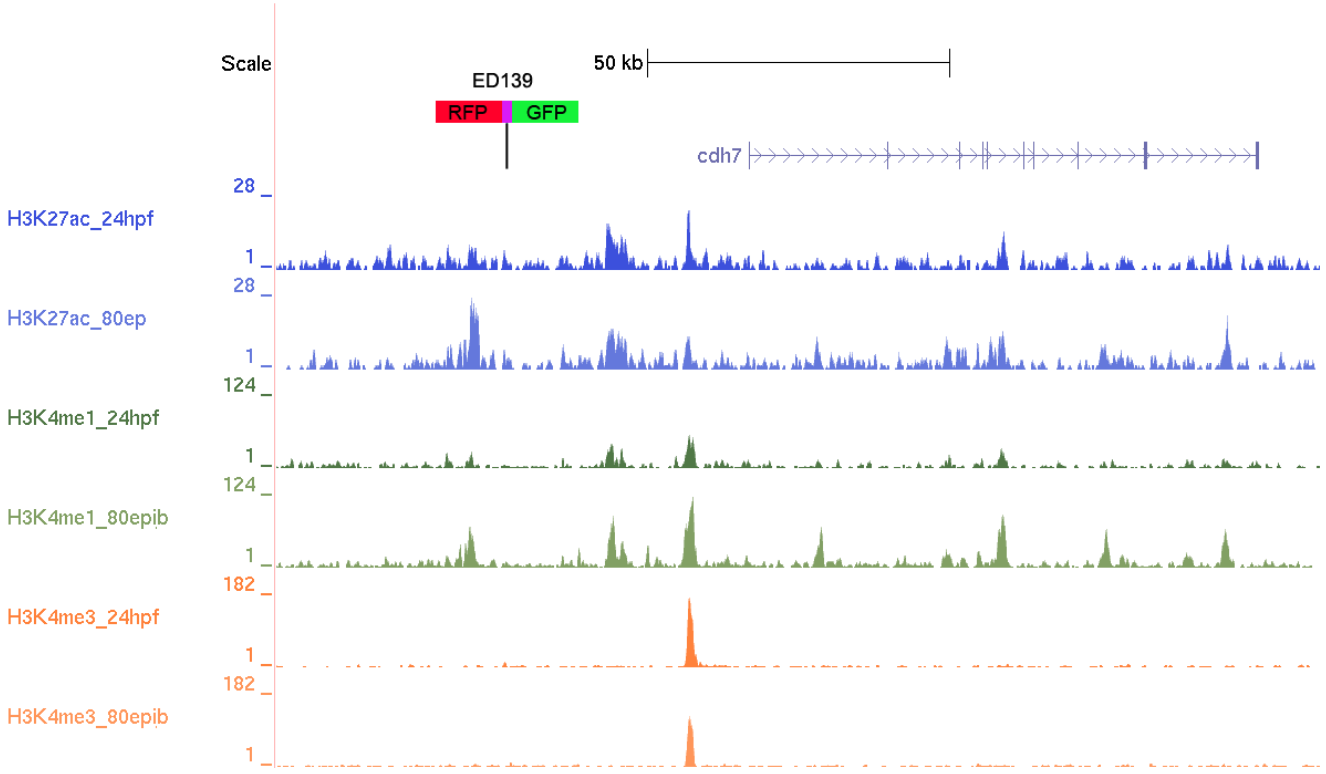
Genomic Landscape:
chr6:36319151-36341285

Size of Landscape
22134 bp

Associated Gene:
her6



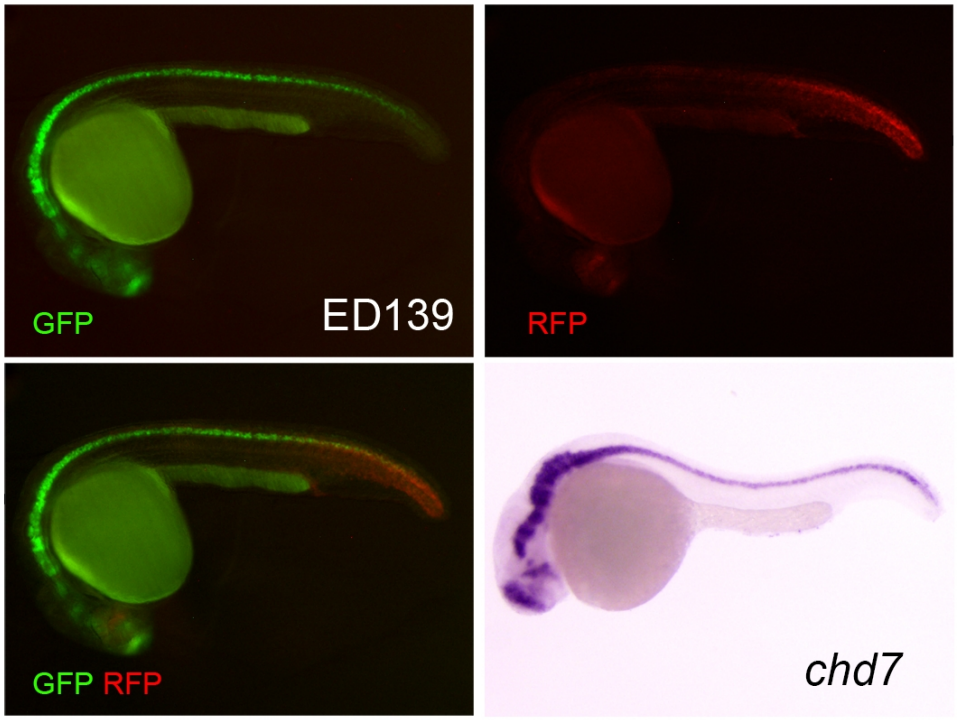
ED139



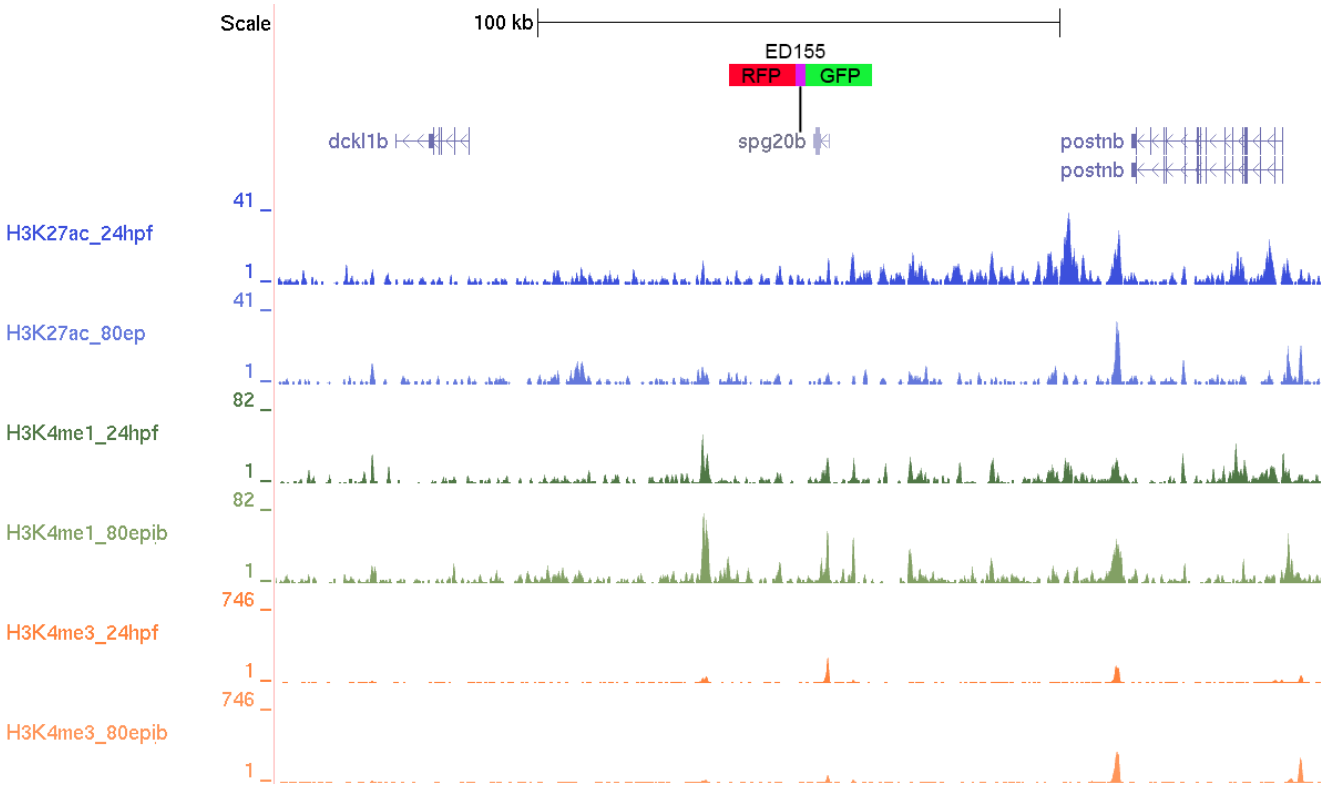
Genomic Landscape:
chr2:27080519-27252823

Size of Landscape
172304 bp

Associated Gene:
cdh7



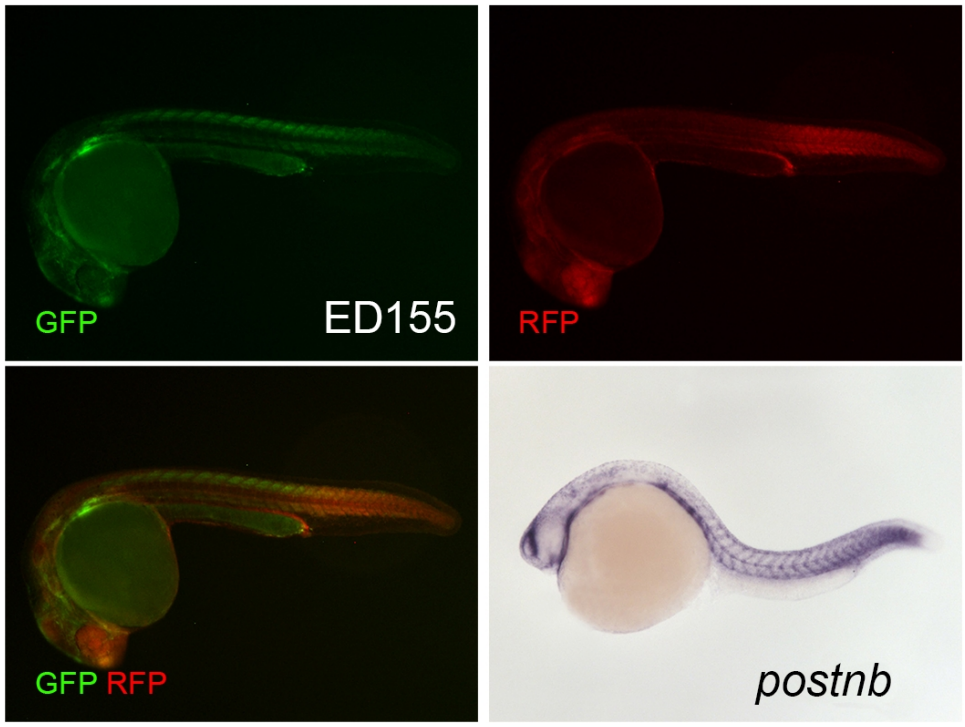
ED155



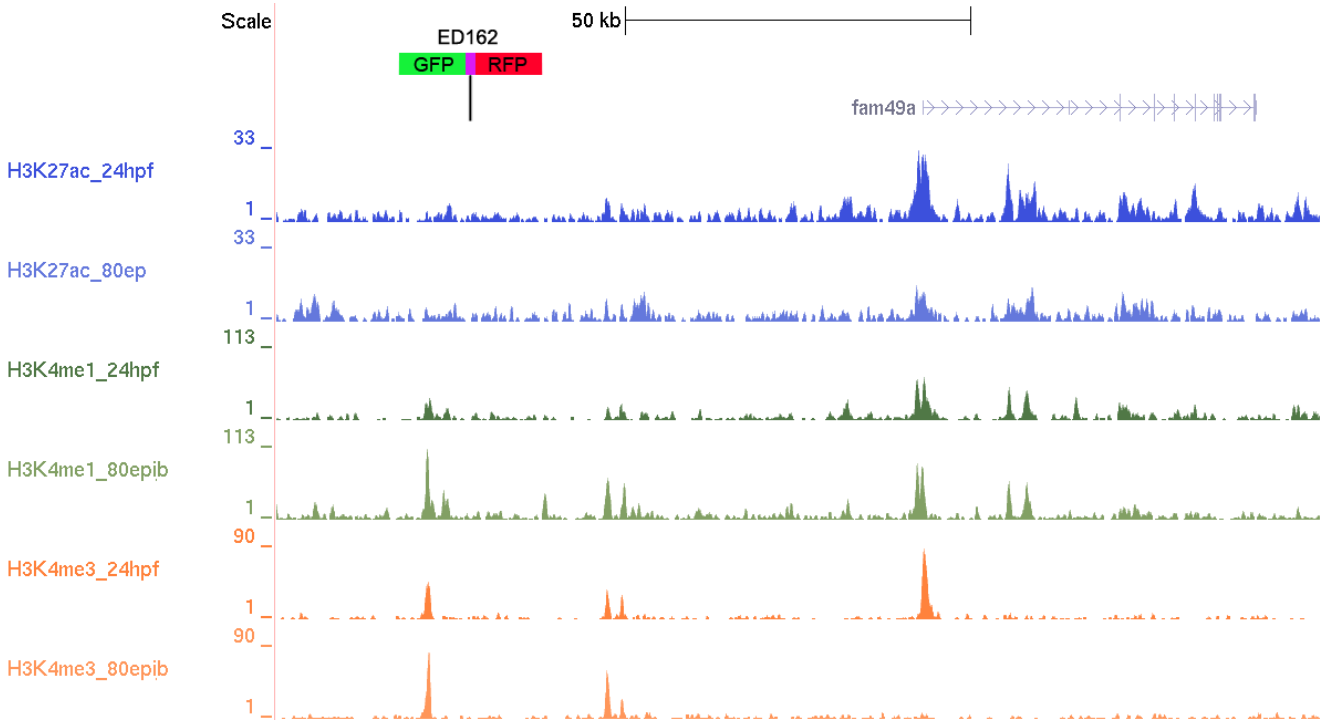
Genomic Landscape:
chr15:32970160-33170167

Size of Landscape
200007 bp

Associated Gene:
postnb



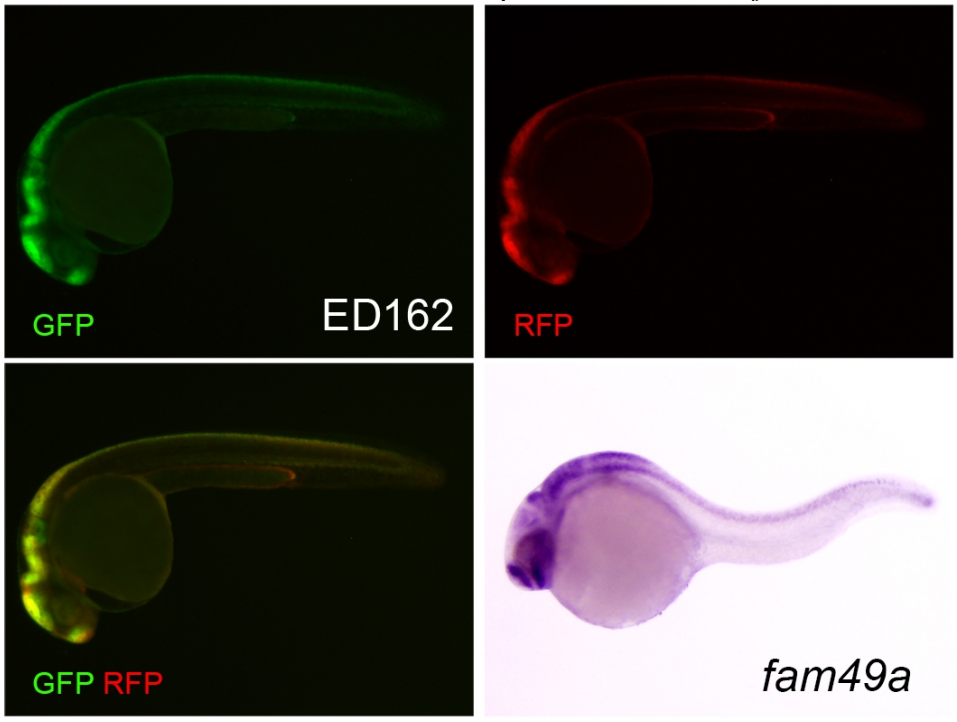
ED162



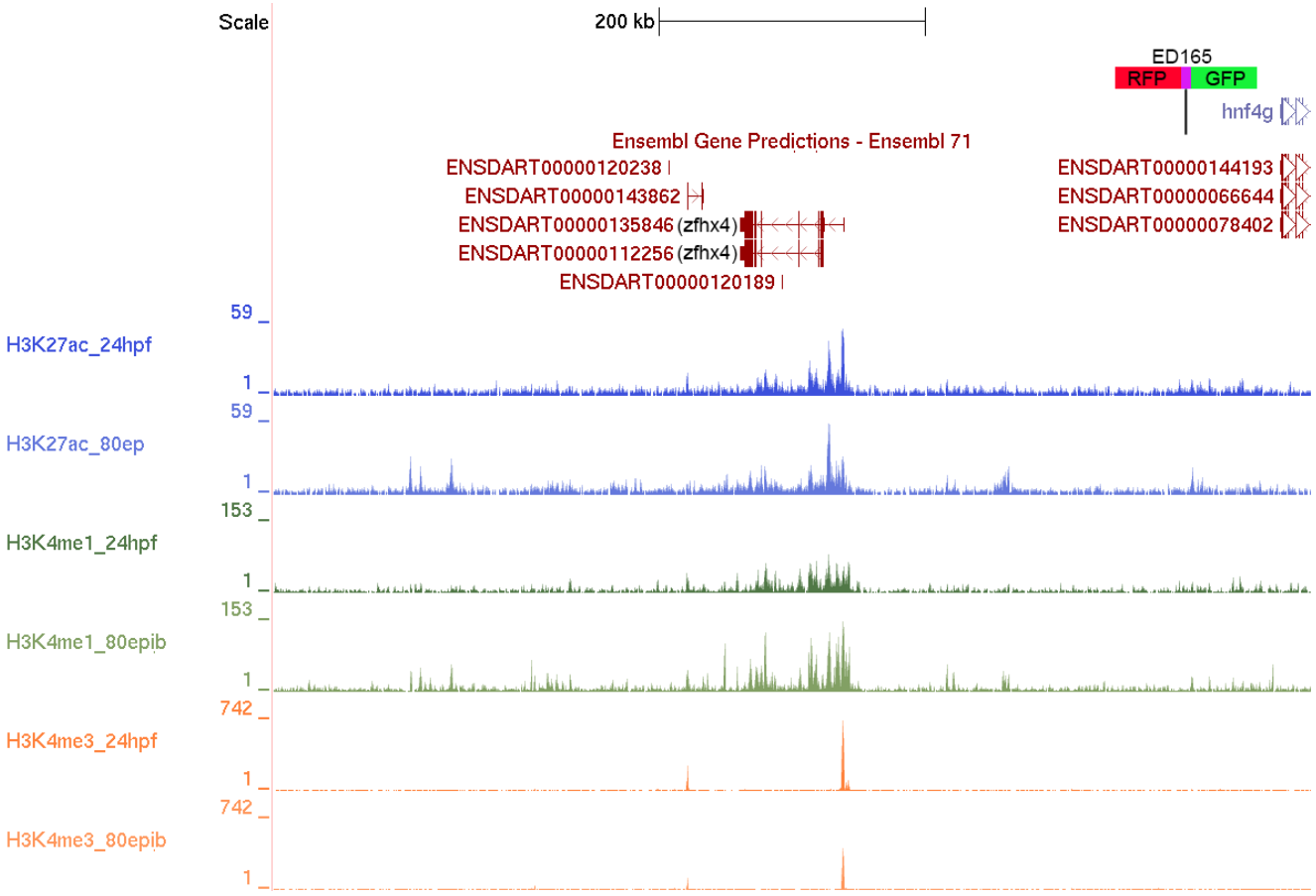
Genomic Landscape:
chr19:31785663-31936390

Size of Landscape
150727 bp

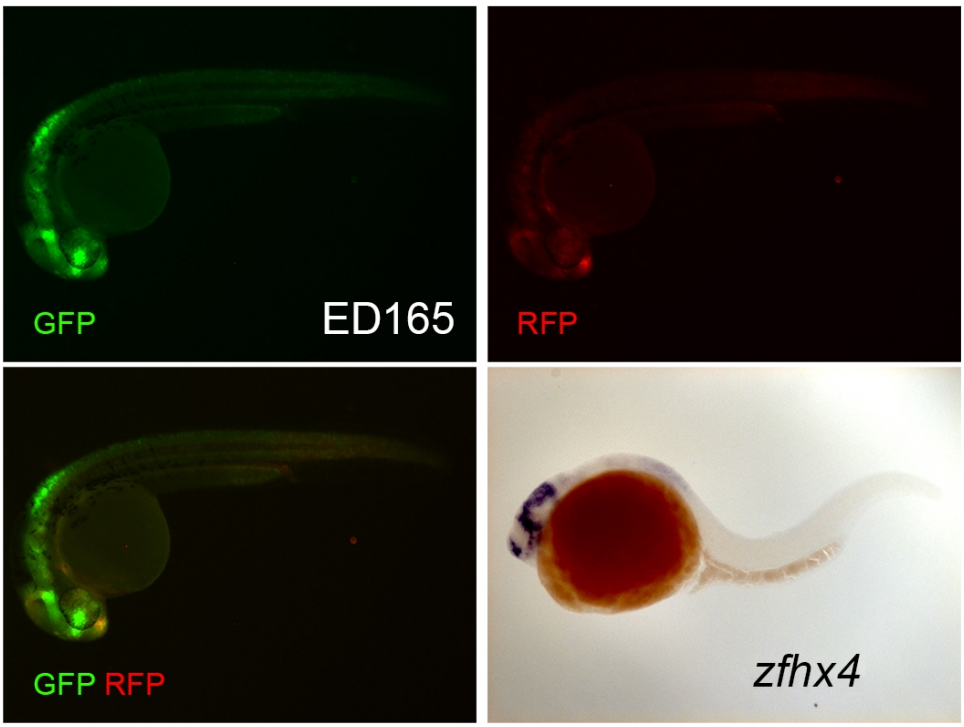
Associated Gene:
fam49a



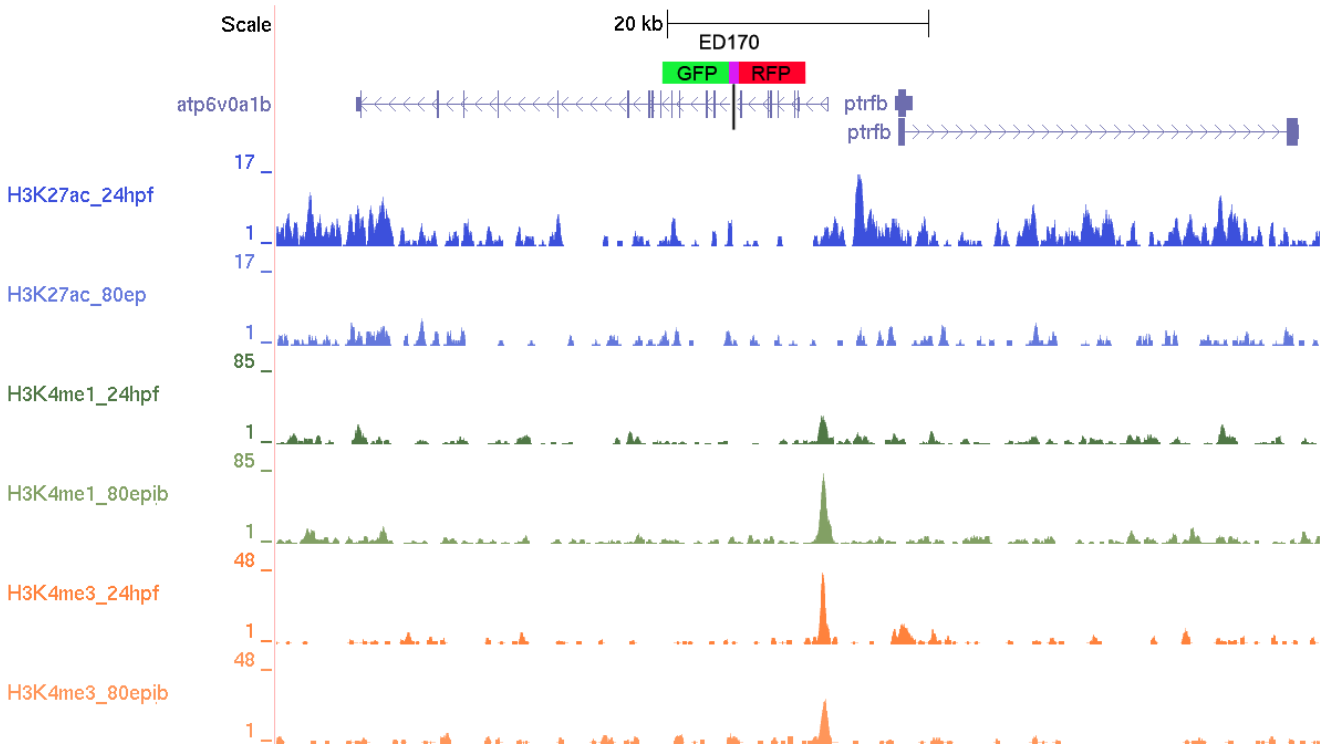
ED165



Genomic Landscape:	Size of Landscape	Associated Gene:
chr24:23499725-24278704	778979 bp	*zfhx4*



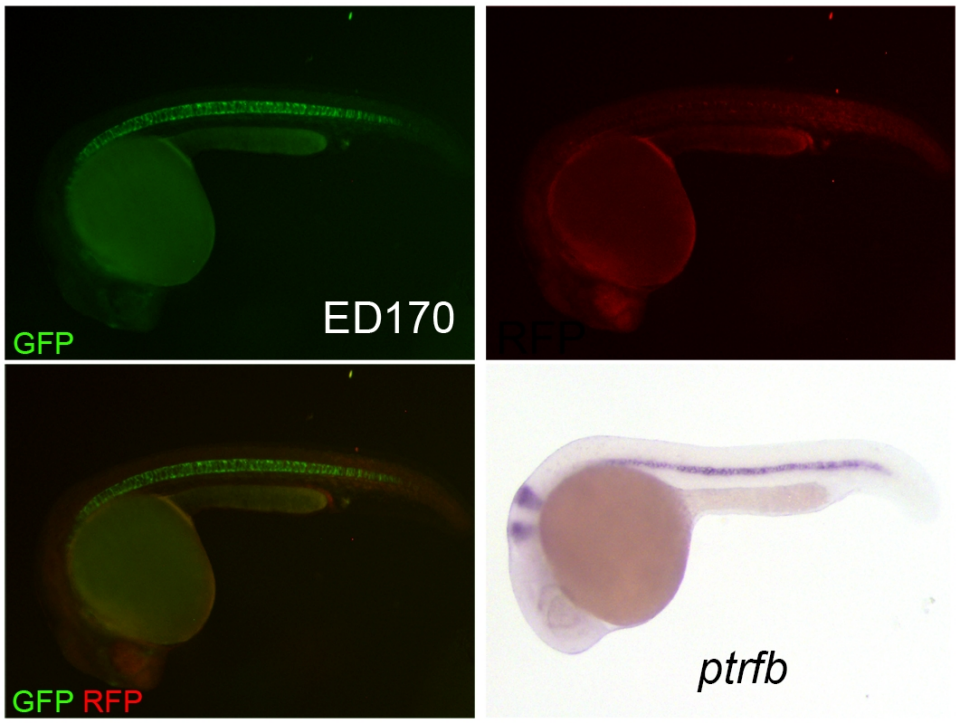
ED170



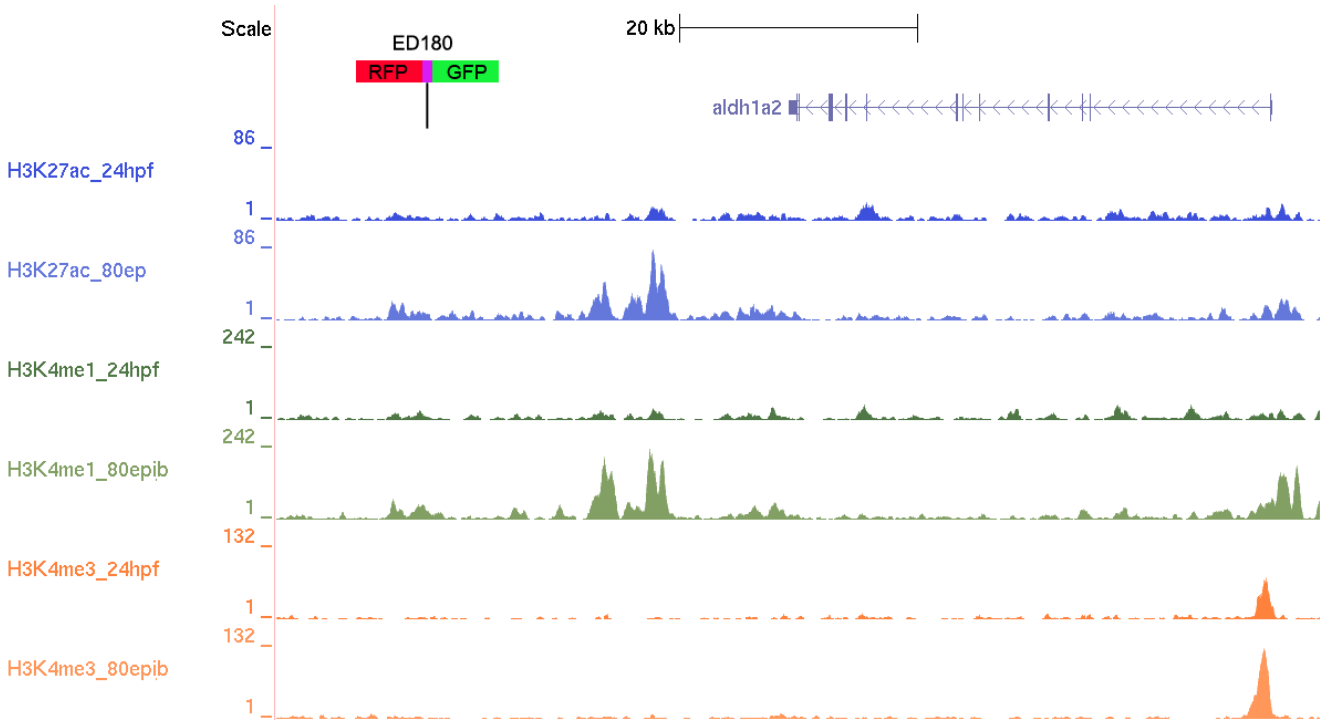
Genomic Landscape:
chr24:7703271-7783173

Size of Landscape
79902 bp

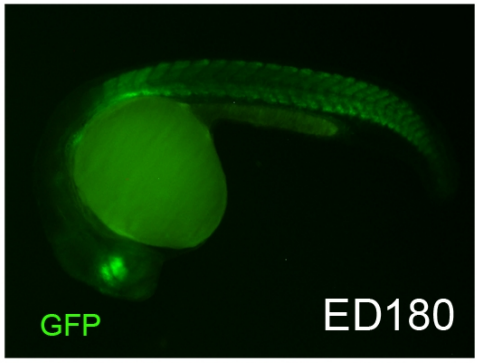
Associated Gene:
ptrfb

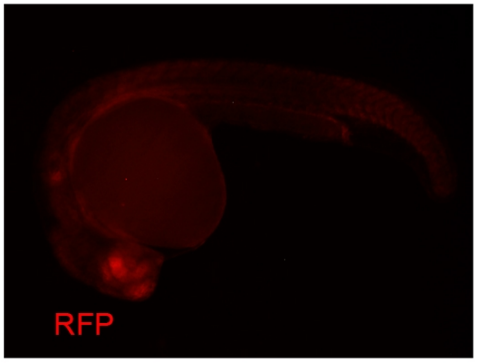


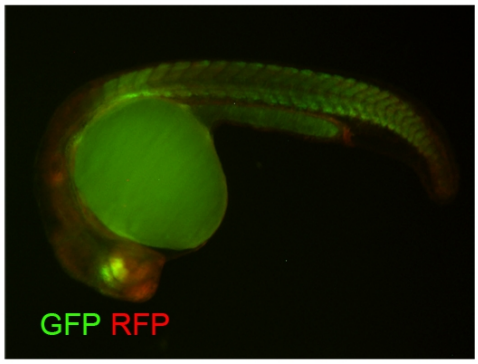
ED180

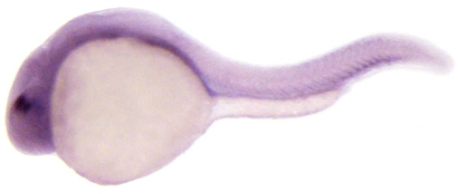


Genomic Landscape:	Size of Landscape	Associated Gene:
chr7:31620580-31708059	87479 bp	*aldh1a2*



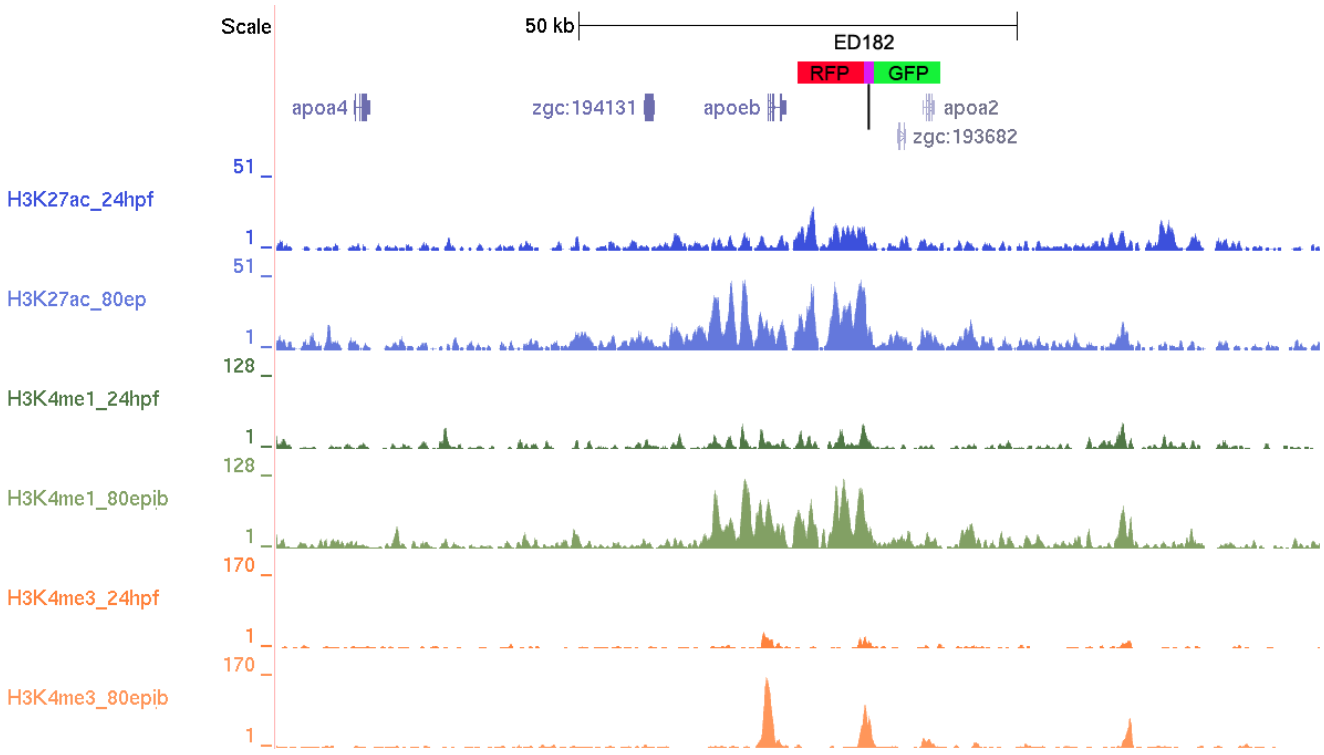






aldh1a2

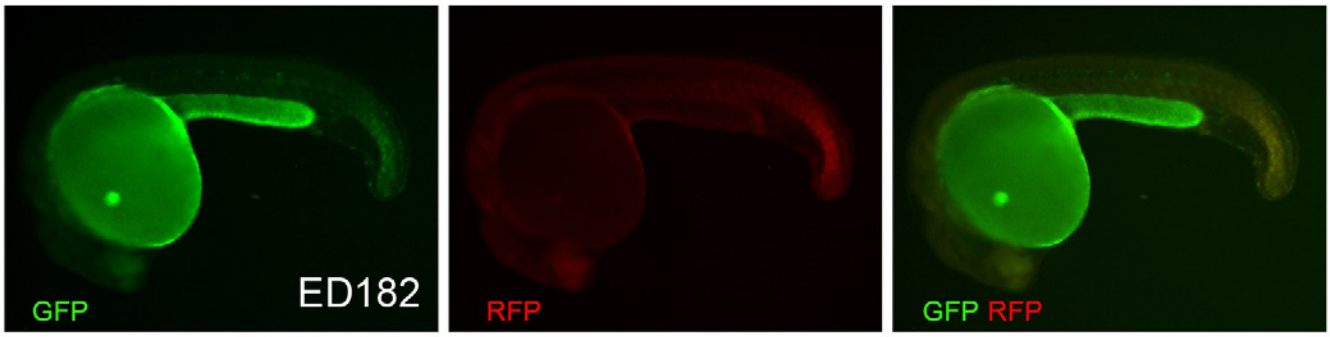
ED182

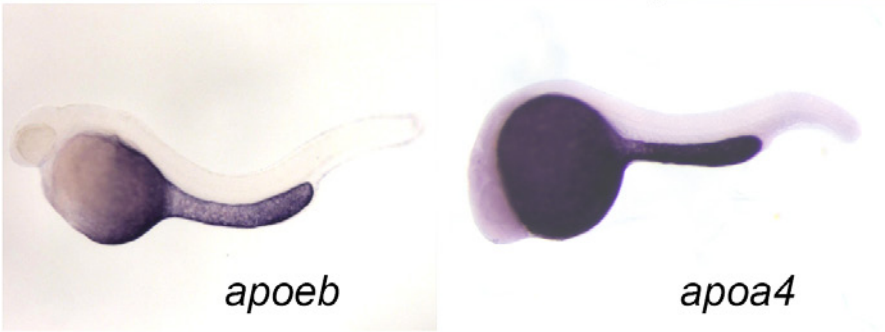


Genomic Landscape:
chr16:26144614-26263494

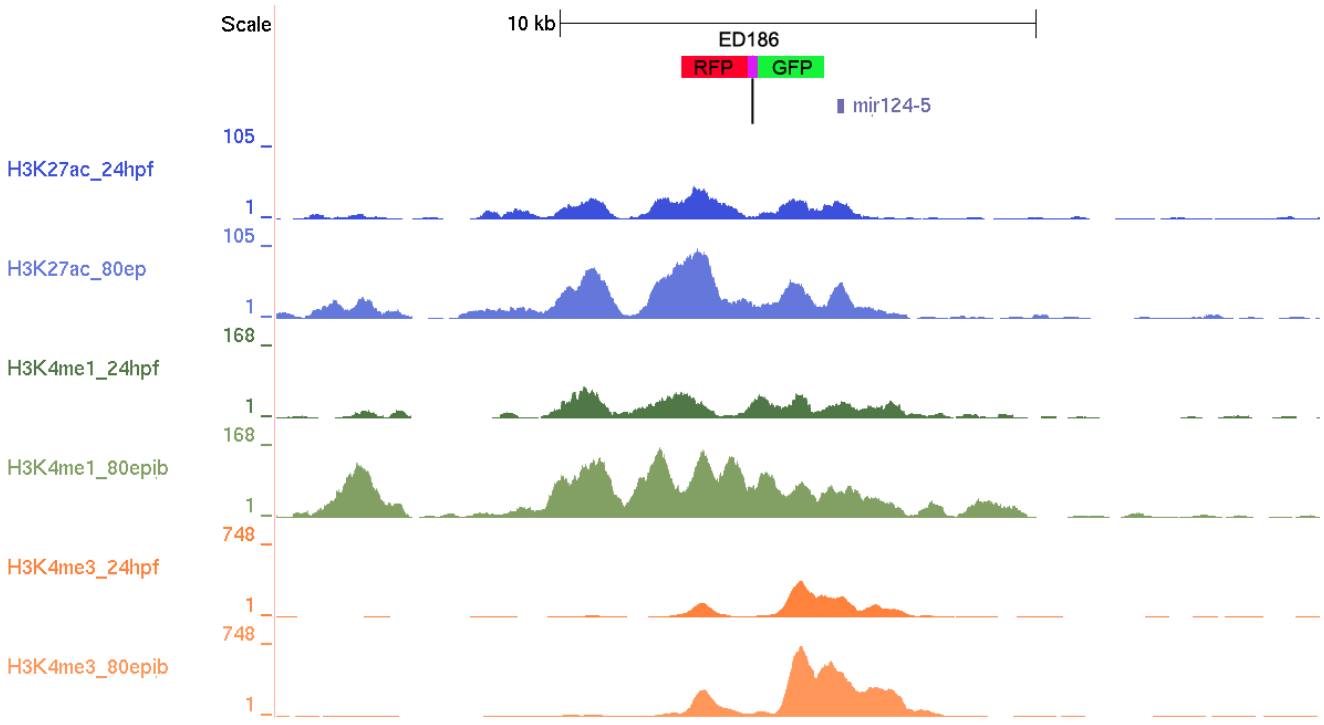
Size of Landscape
118880 bp

Associated Gene:
apoa4 and *apoeb*





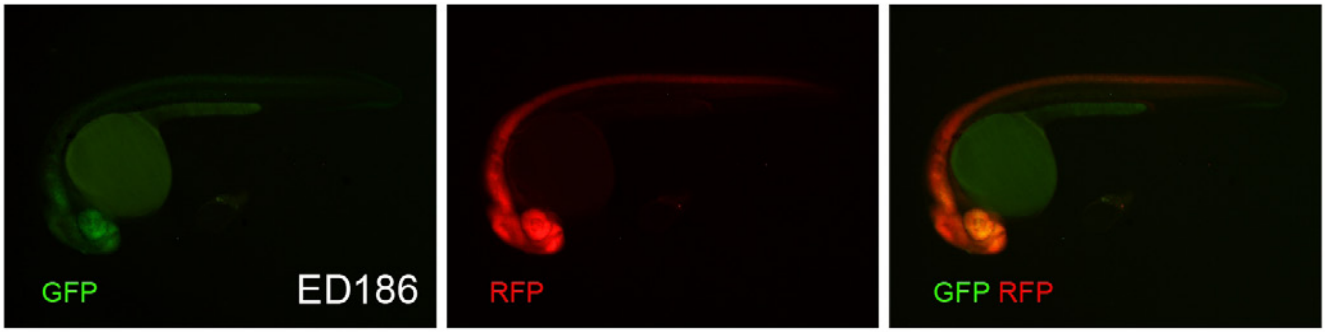
ED186



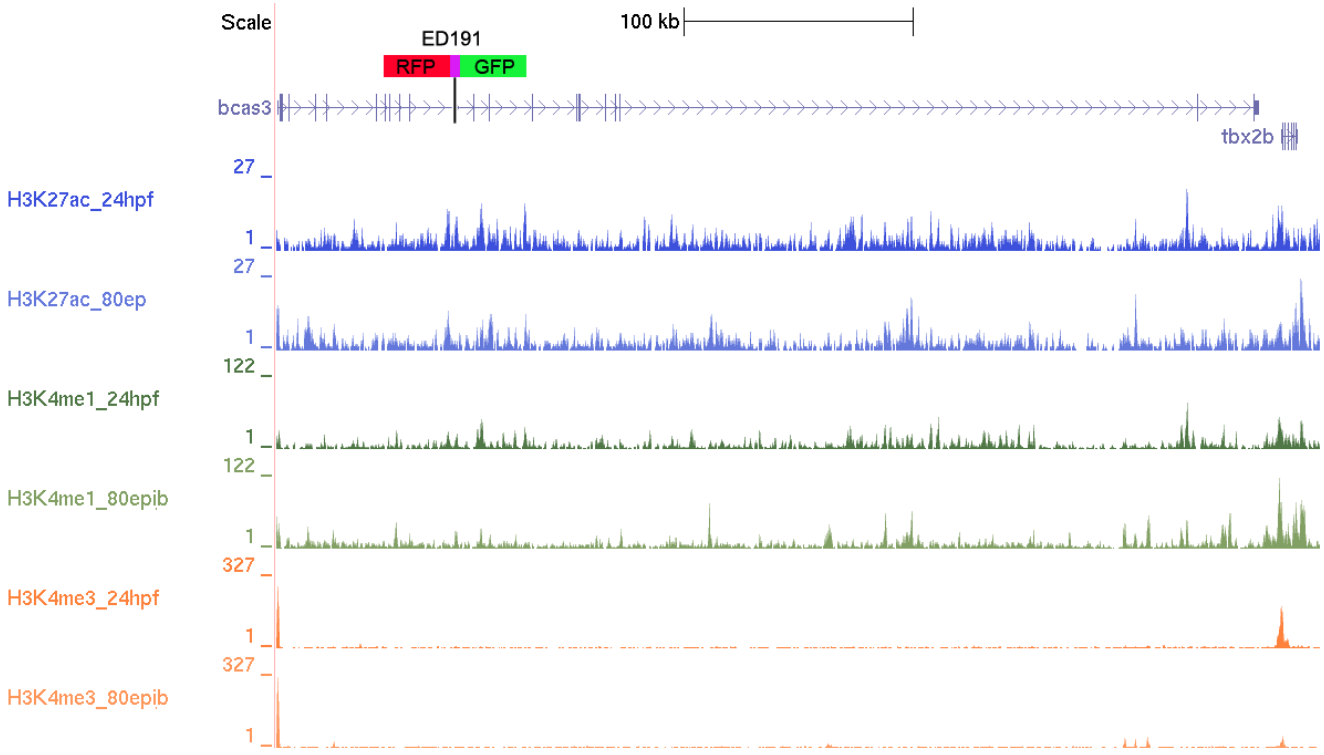
Genomic Landscape:
chr23:17942332-17964259

Size of Landscape
21927 bp

Associated Gene:
mir124-5



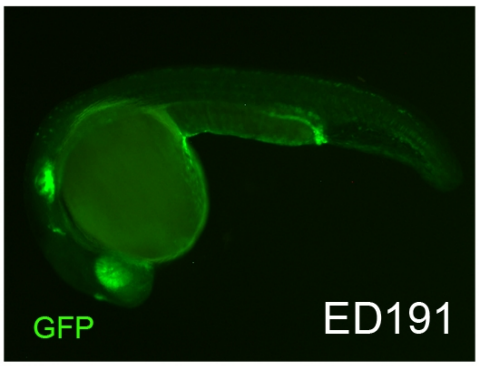
ED191

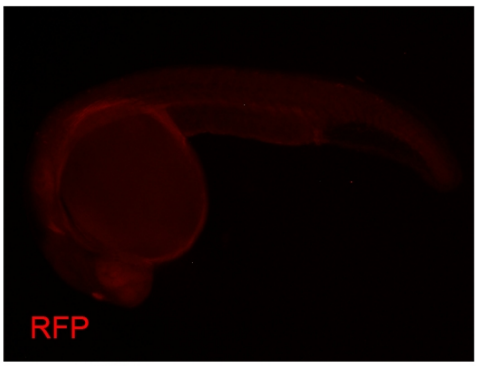


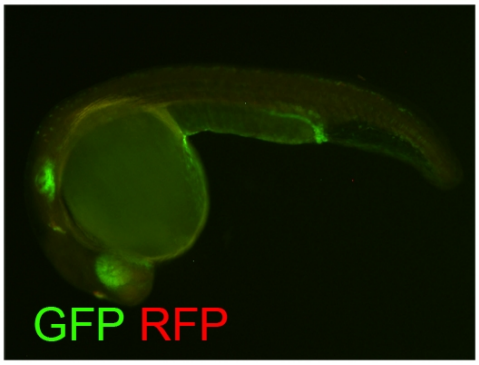
Genomic Landscape:
chr15:26276119-26730220

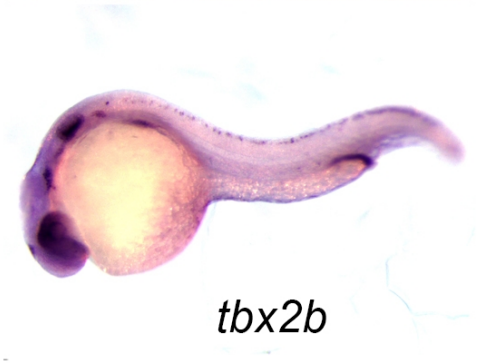
Size of Landscape
454101 bp

Associated Gene:
tbx2b

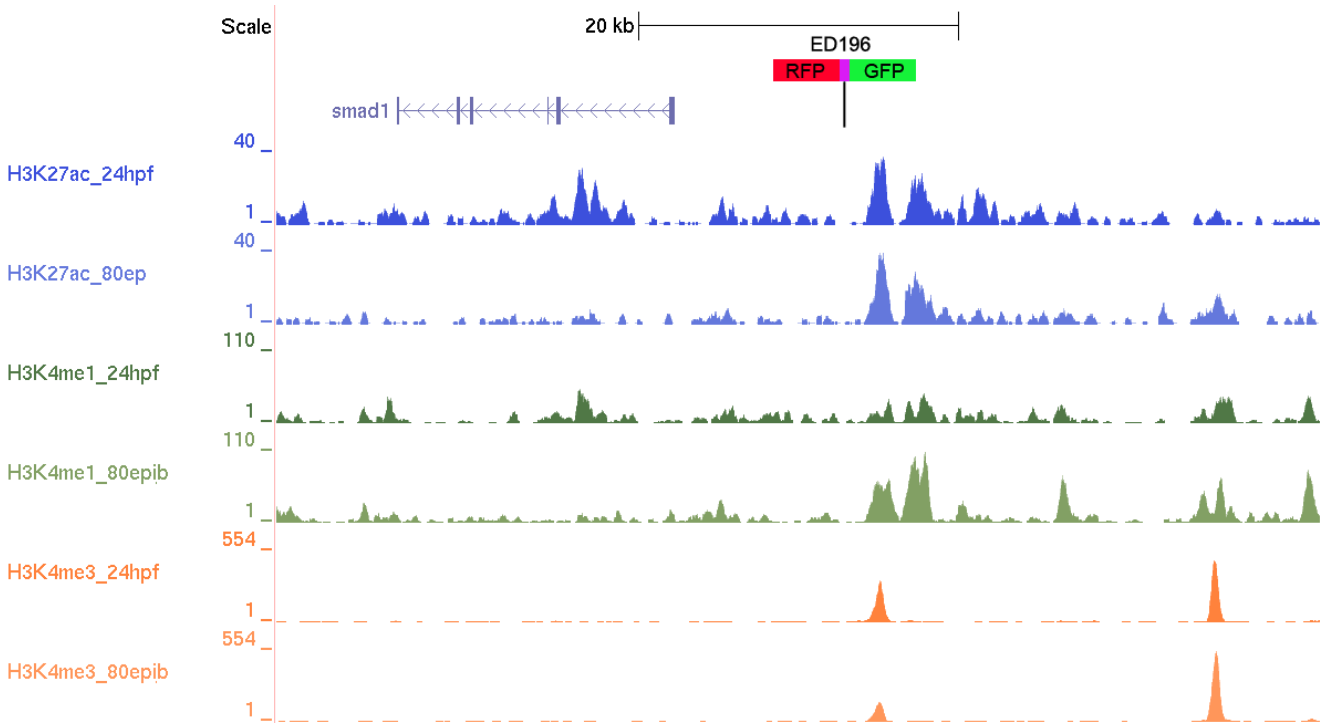








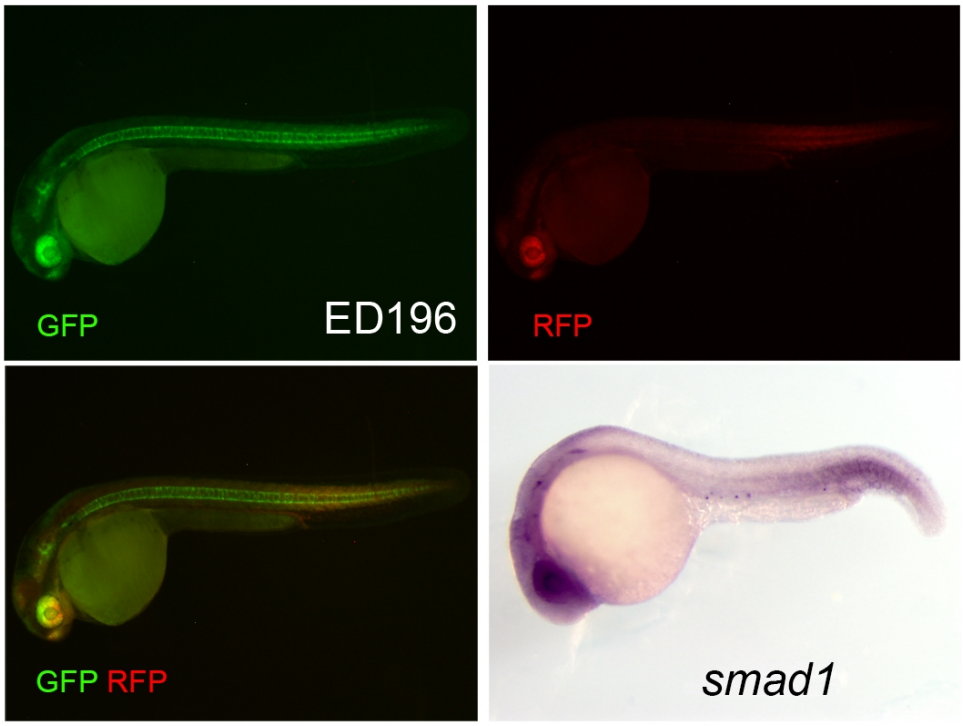
ED196



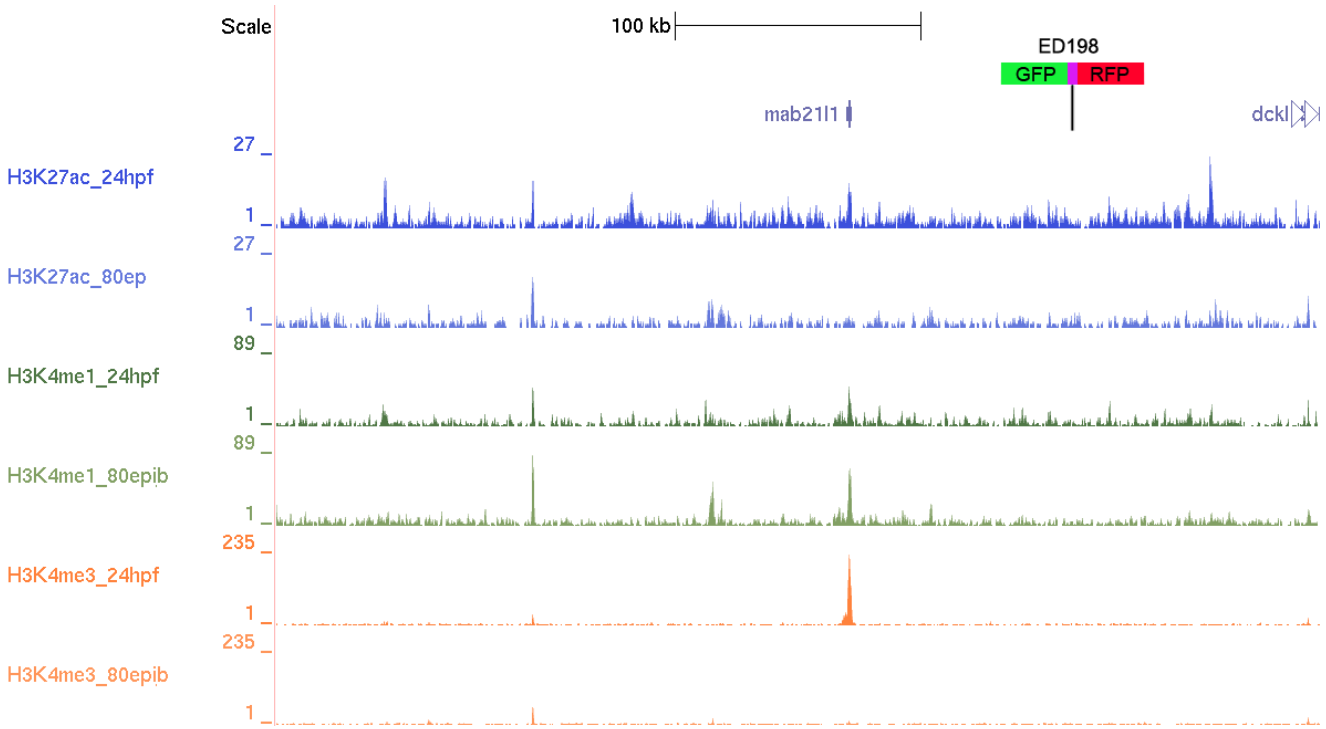
Genomic Landscape:
chr1:35599004-35664148

Size of Landscape
65144 bp

Associated Gene:
smad1



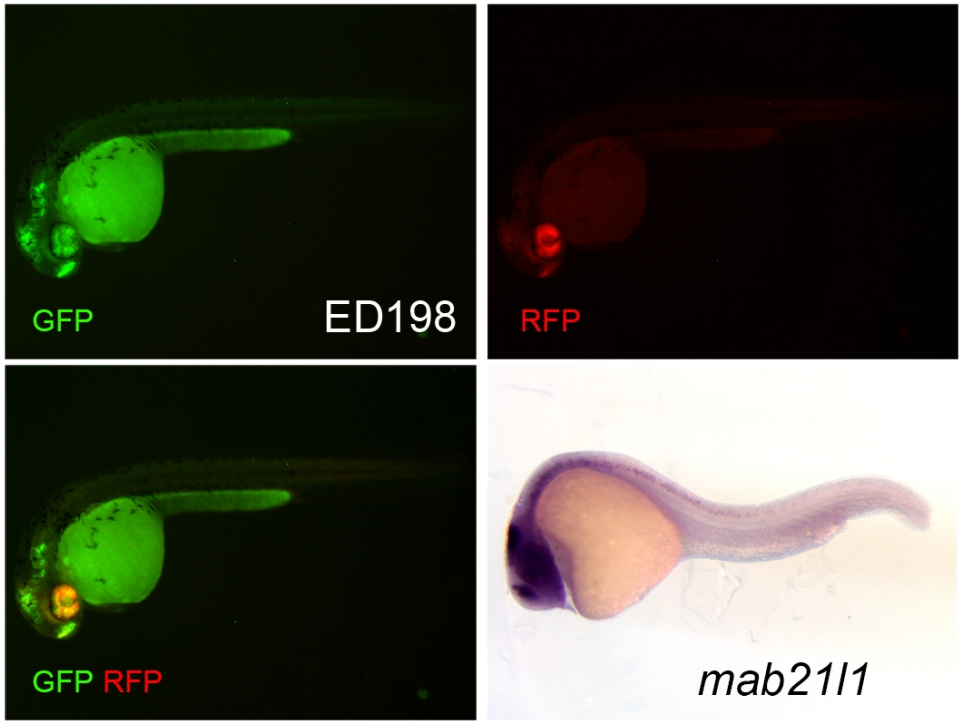
ED198



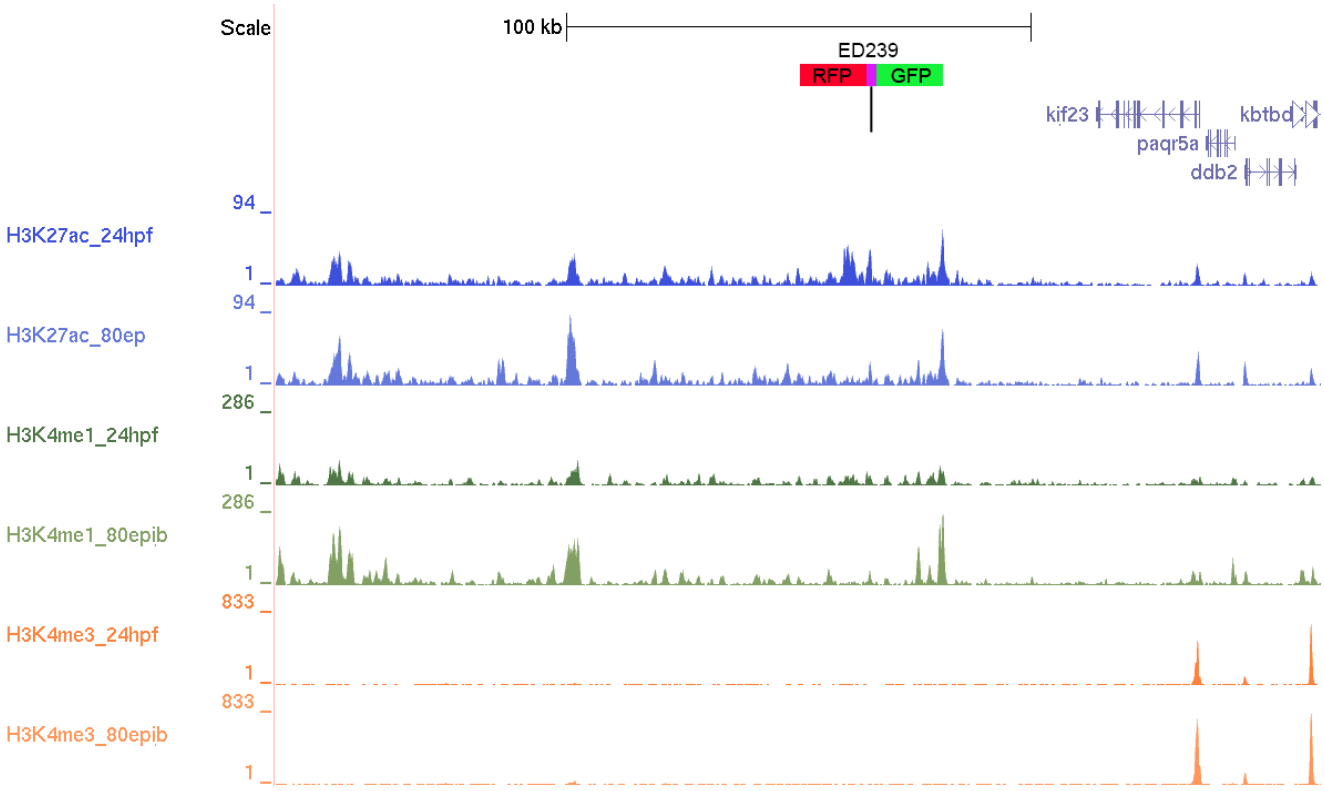
Genomic Landscape:
chr15:32570115-32993341

Size of Landscape
423226 bp

Associated Gene:
mab21l1



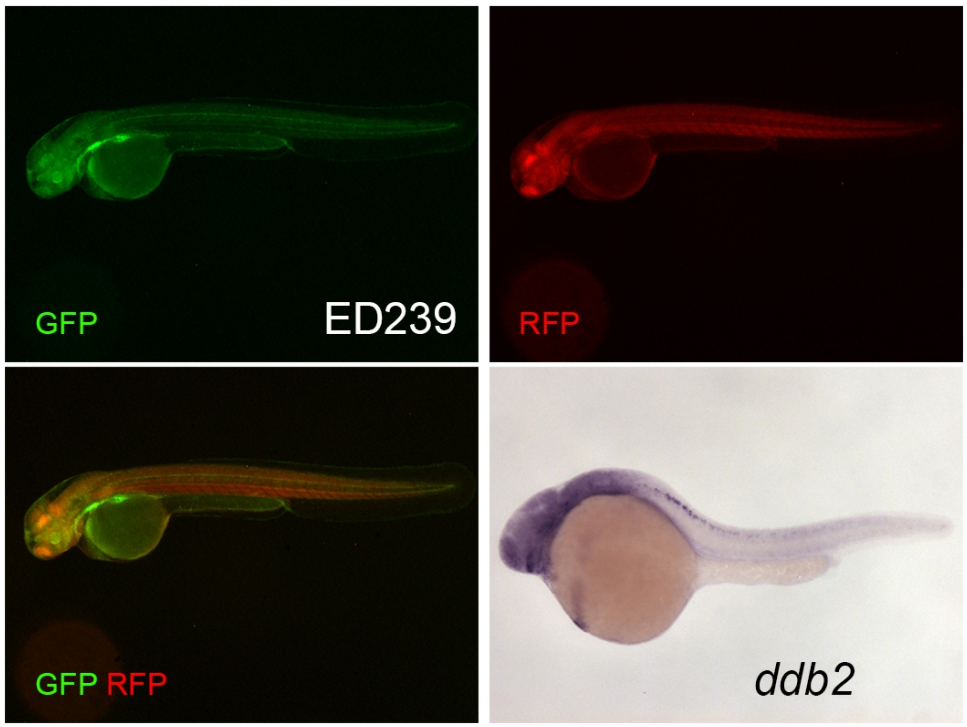
ED239



Genomic Landscape:
chr18:20040522-20265304

Size of Landscape
224782 bp

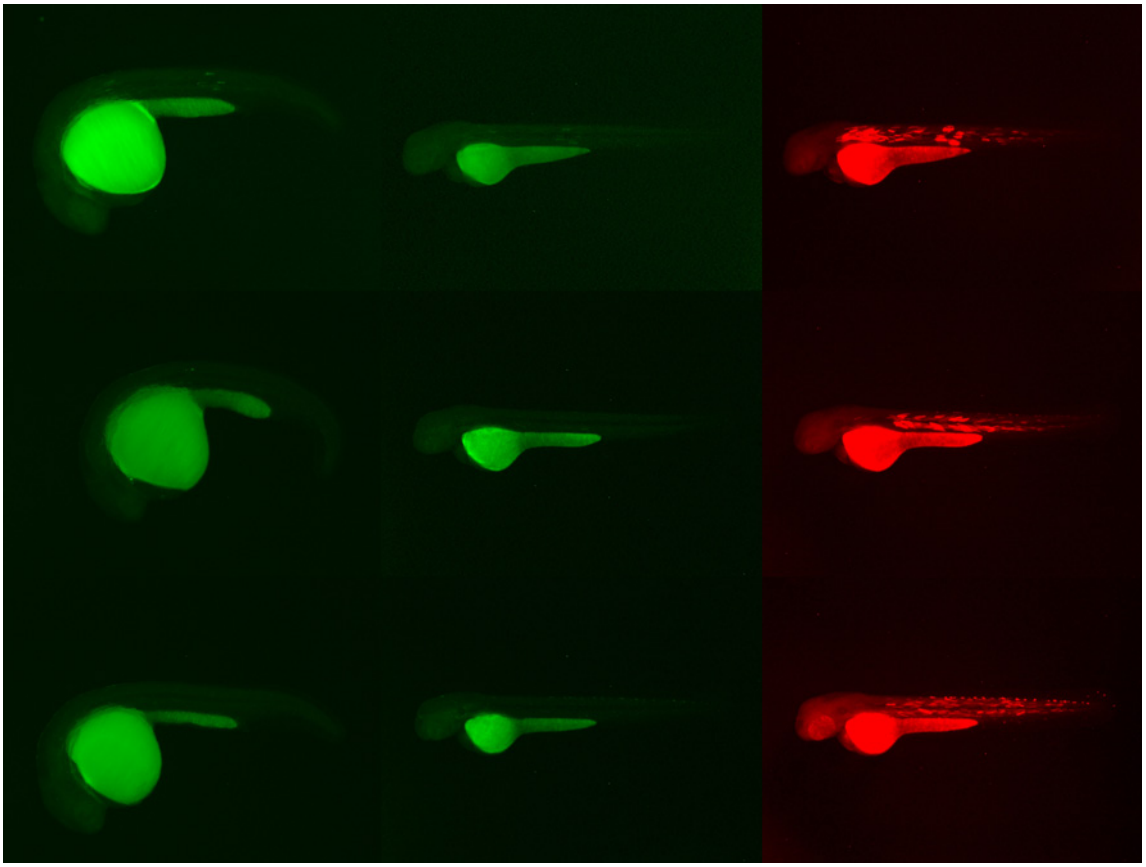
Associated Gene:
ddb2



Supplemental Figure 6 - ED mapped lines associated to a gene.

For each line, a diagram representing the genomic landscape of the ED insertion is available. The first track of this diagram represents the ED insertion point and its orientation (GFP RFP or RFP GFP) in the respective genomic landscape. The second track plots genes present in this genomic landscape (RefSeq genes). Third to eighth tracks represent epigenetic marks obtained by chip-seq (Bogdanovic et al. 2012). Each mark is plotted at two different developmental times, 24hpf (higher track) and 80% epiboly (lower track). Third and fourth tracks plot the H3K27 acetilation mark, the fifth and sixth tracks plot H3K4 mono methylation mark and seventh and eighth tracks plot H3K4 tri methylation mark. Genomic coordinates of the respective landscape are displayed under "Genomic Landscape:" (Zv9/danRer7 zebrafish genome assembly). The size of the landscape is displayed under "Size of Landscape". The gene associated to the ED line is named under "Associated Gene:" and its expression pattern is shown in the lower panels in the image containing the corresponding gene name. Images of the GFP and RFP expression patterns for each ED line are also included in the lower panels. In some cases a probe against *krox20* was used as an internal control for the *in situ* hybridization and its expression pattern is noted with asterisks. *apoeb, atp2b1a, calm1a, ddb2, irf2bp2b, postnb, robo1, traf4a, vsx1,* and *zfhx4 in situ* hybridization images were obtained from The Zebrafish Model Organism Database (Bradford et al. 2011). The expression patterns of *atoh8, mir124*-5, *mir216a-1* and *pld1a* are described by Yao and colleges (Yao et al. 2010), Shkumatava and colleges (Shkumatava et al. 2009) and Zeng and colleges (Zeng et al. 2009).

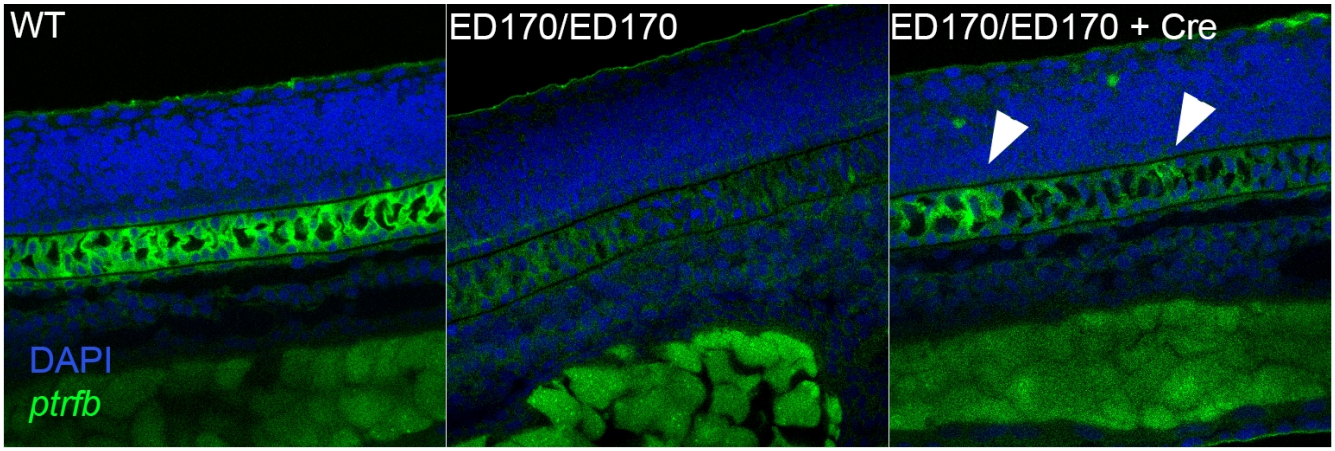
Supplemental Figure 7



Supplemental Figure 7 – Enhancer assay for the 0.9 kb fragment downstream of ED186 insertion.

Three different F0 embryos (rows) injected with a reporter vector (Bessa et al. 2009) to test the enhancer activity of the 0.9 kb fragment downstream of the ED186 insertion. In this assay GFP expression (green) reports enhancer activity, which is not detected at 24 (first column) or 48hpf (second column). RFP expression (Red) is used as an internal control of trangenesis (third column).

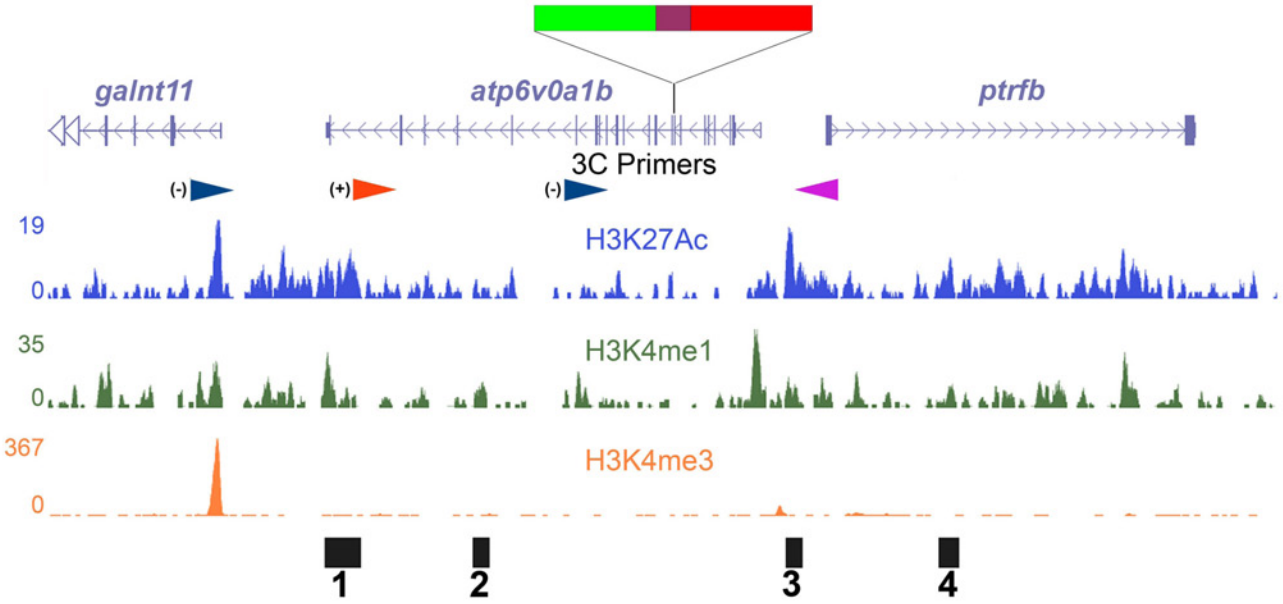
Supplemental Figure 8



Supplemental Figure 8 - Detection of *ptrfb* expression by fluorescent *in situ* hybridization in a wild type and ED170 homozygous mutant background, with and without Cre recombinase.

Wild type embryos present a strong expression of *ptrfb* (green) in the notochord (left image). In ED170 homozygous mutant background *ptrfb* expression levels are strongly reduced (middle image). Cre recombinase injection in ED170 homozygous mutant embryos induces mosaics of cells expressing similar to wild type *ptrfb* levels (arrows; right image). Embryos were stained with the nuclear marker DAPI (blue) and images were acquired with a confocal microscope. Cre injected embryos displaying a high degree of mosaicism were selected to show differences of transcript levels in rescued cells versus non rescued cells within the same sample.

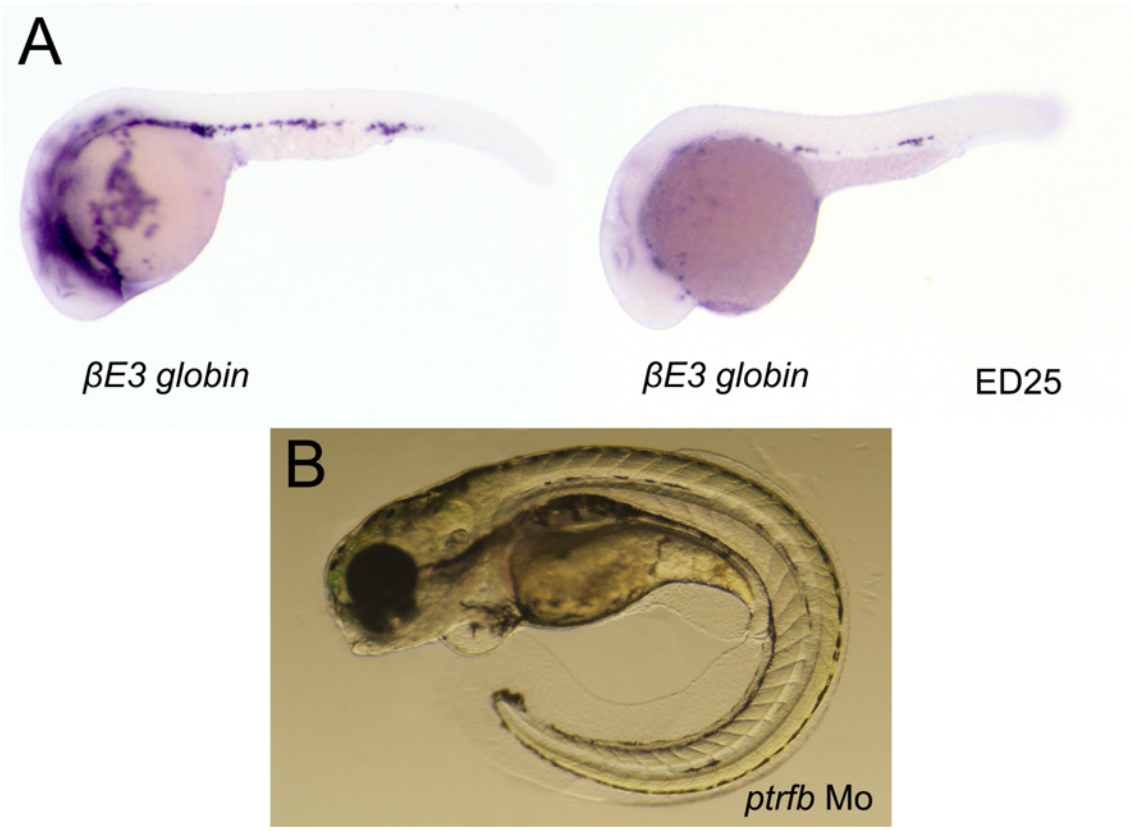
Supplemental Figure 9



Supplemental Figure 9 – Selection of putative enhancers of *ptrfb*.

Based on H3K4me1 (green) or H3K27Ac (blue) marks at dome, 80%, epiboly and 24hpf (shown here) developmental stages, 4 candidate enhancer sequences where selected, 1 to 4 (black boxes). The H3K4me3 promoter associated mark is also displayed (orange).

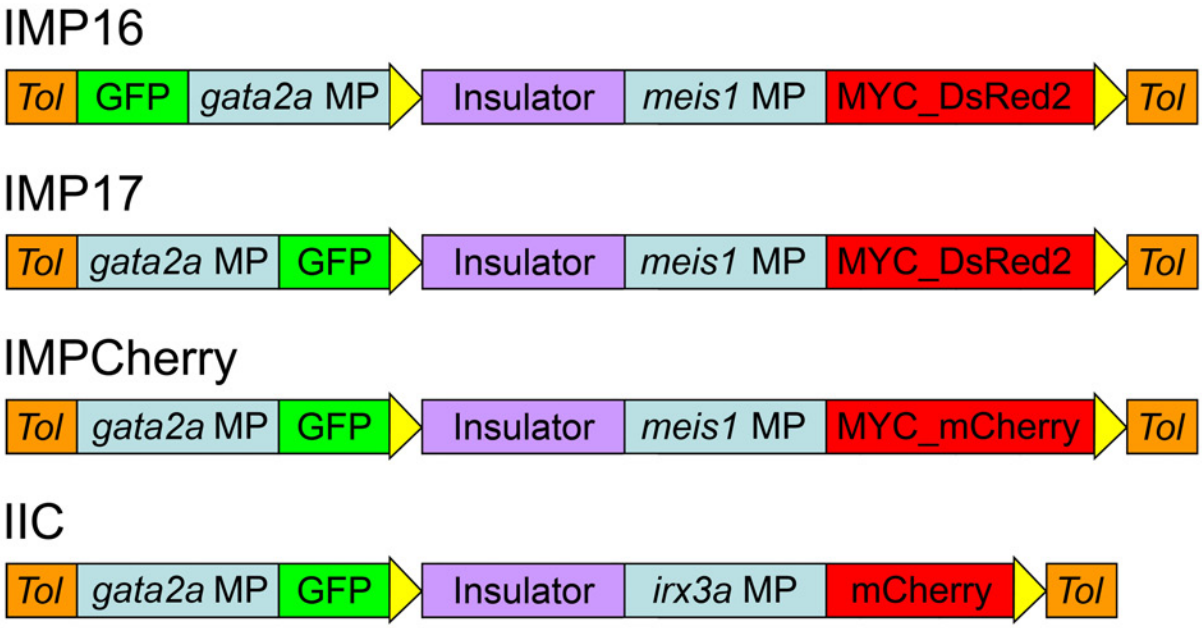
Supplemental Figure 10



Supplemental Figure 10 - Phenotypes associated to the *klf4b*/ED25 mutant and to a strong specific knockdown of *ptrfb* using a morpholino.

(A) *beta E3 globin* is expressed in blood cells of wild type embryos at 24hpf (left). This expression is reduced in the *klf4b*/ED25 mutant (right). (B) The knockdown of *ptrfb* by injecting 10 ng of a morpholino that targets specifically *ptrfb* generates a strong phenotype characterized by a dramatically bent tail.

Supplemental Figure 11



Supplemental Figure 11 – Diagram of the different versions of ED vectors used in the ED screen.

Four different vectors are represented, IMP16, IMP17, IMPCherry and IIC. Each vector is composed by two *Tol2* transposon terminal inverted repeats (orange boxes; *Tol*), two minimal promoters (blue boxes; MP), one insulator (purple, Insulator), two *loxP* sequences (yellow triangles) and two reporter genes, eGFP (green boxes; GFP) and RFP (red boxes). Different versions of RFP were used, DsRed2 fused to a myc tag (myc_DsRed2), mCherry fused to a myc tag (myc_mCherry) and mCherry alone. The vast majority of the ED lines, 223 in total, were generated using the IIC vector.

REFERENCES

Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, Fernández A, Montoliu L, Akalin A, Lenhard B, Casares F, et al. 2009. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn* **238**: 2409–2417.

Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJC, Gómez-Skarmeta JL. 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res* **22**: 2043–2053.

Bradford Y, Conlin T, Dunn N, Fashena D, Frazer K, Howe DG, Knight J, Mani P, Martin R, Moxon SAT, et al. 2011. ZFIN: enhancements and updates to the Zebrafish Model Organism Database. *Nucleic Acids Res* **39**: D822–829.

Shkumatava A, Stark A, Sive H, Bartel DP. 2009. Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes Dev* **23**: 466–481.

Yao J, Zhou J, Liu Q, Lu D, Wang L, Qiao X, Jia W. 2010. Atoh8, a bHLH transcription factor, is required for the development of retina and skeletal muscle in zebrafish. *PLoS ONE* **5**: e10945.

Zeng X-XI, Zheng X, Xiang Y, Cho HP, Jessen JR, Zhong TP, Solnica-Krezel L, Brown HA. 2009. Phospholipase D1 is required for angiogenesis of intersegmental blood vessels in zebrafish. *Dev Biol* **328**: 363–376.