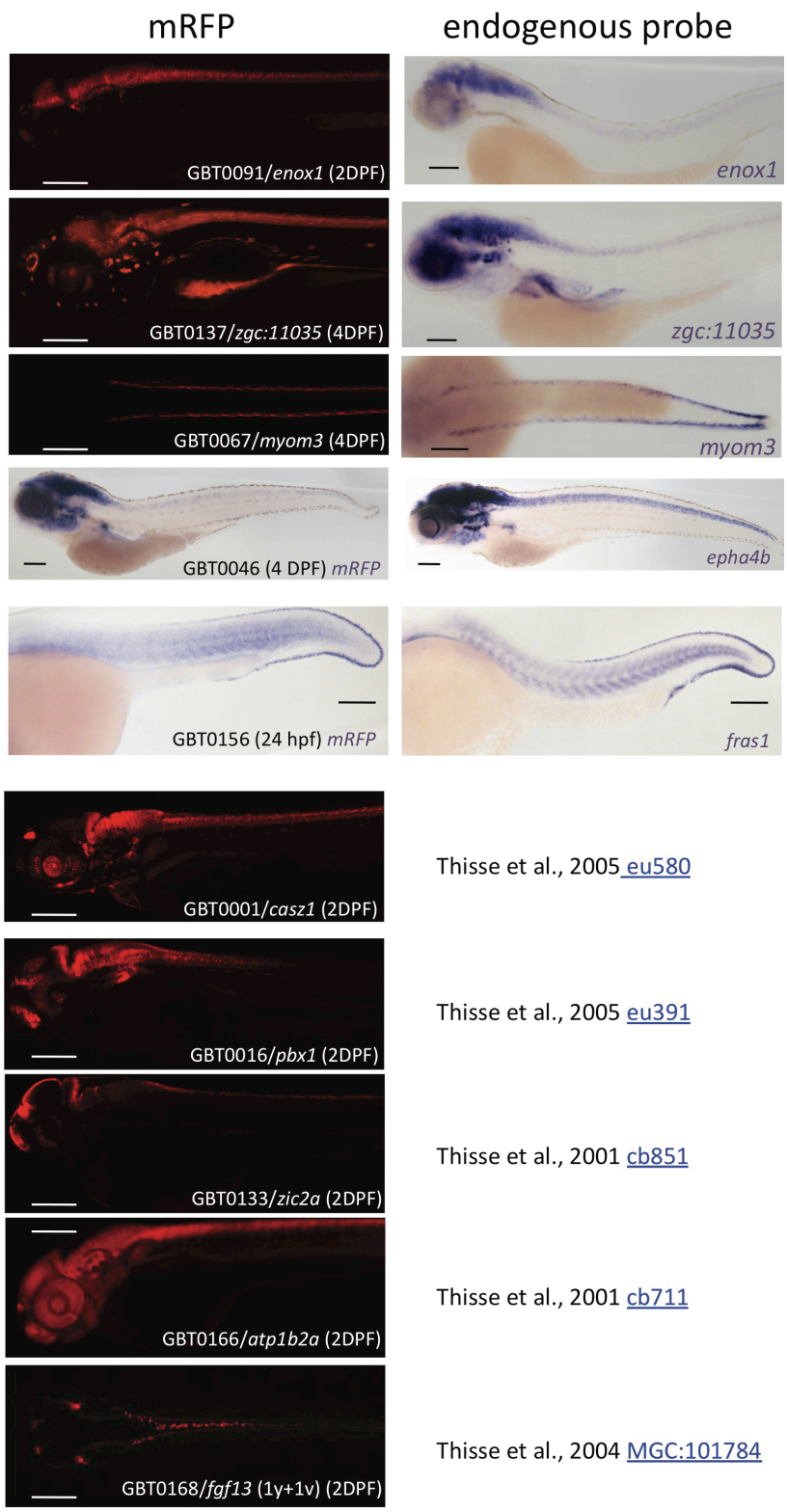
Supplementary Figure 1



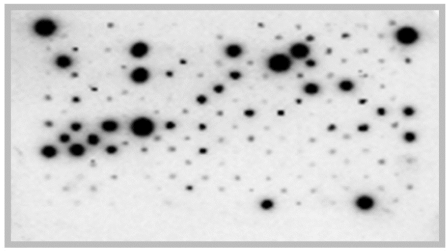
mRFP localization and corresponding endogenous patterns. Shown in the left column are mRFP fluorescence patterns or *in situ* using antisense mRFP probe in GBT lines with secreted fusion proteins. The right column shows the endogenous gene expression pattern as determined by *in situ* hybridization with a gene-specific probe or lists the probe identifier and reference from publicly available images available at http://zfin.org. The images show near matches in expression with subtle differences likely due to several factors. First, the mRFP fusion protein has different kinetics than the mRNA that produces it– as translation occurs after transcription, the protein half-life can be different than the mRNA half-life. Second, the mRFP fusion transcript lacks the nascent 3'UTR that may affect differential spatiotemporal mRNA stability of the endogenous transcript. Third, the mRFP trap may only affect a subset of the transcripts detected by whole-mount *in situ* hybridization. Such is the case for GBT0168/*fgf13* 1y+1v isoform, which is more specific than the *in situ* against all *fgf13* transcripts. Lastly, the quality of imaging from live fish for mRFP fluorescence can exceed the imaging of colorimetric reactions used in a typical whole-mount *in situ* hybridization. [1-3].

Supplementary Figure 2

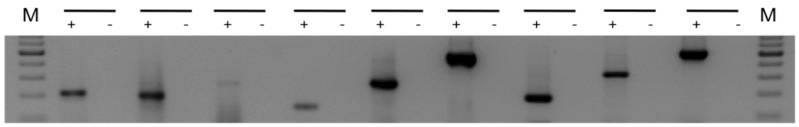
a.

Description of sequences trapped by the 3' exon trap cassette- somatic	
Known genes	125 (10%)
Unannotated transcripts/ESTs	75 (6%)
Putative novel endogenous transcripts	850 (67%)
Low complexity and repeat sequences	223 (17%)
Total trapped sequences analyzed	1273

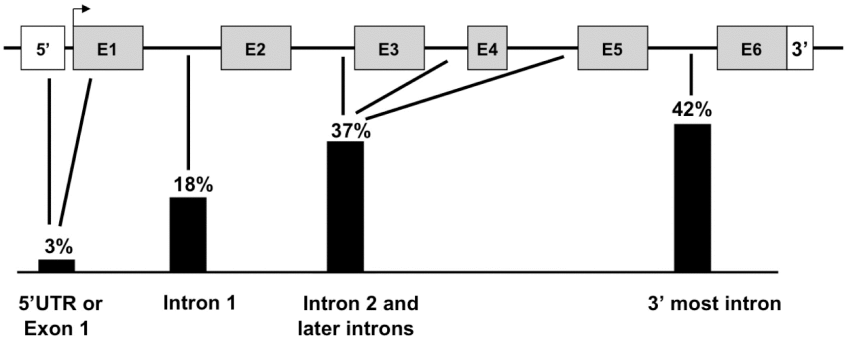
b.



Putative novel endogenous transcript captured by the 3' exon trap validated by Reverse northern assay and RT-PCR assay using RNA from wild type zebrafish embryos as template



c.

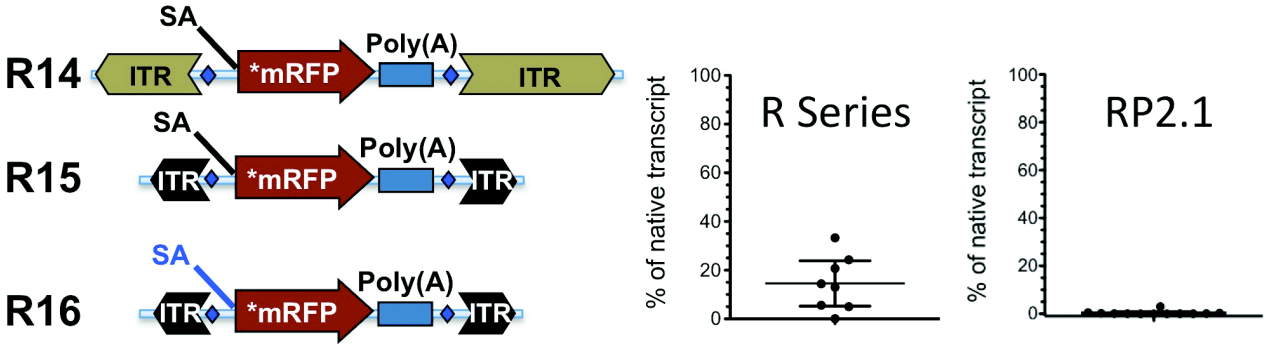


d.

Description of sequences trapped by the 3' exon trap cassette- germline	
Known Genes	9 (32%)
Putative novel endogenous transcripts	13 (46%)
Low complexity and repeat sequences	6 (21%)

Evaluation of 3' exon trap integration sites in zebrafish. (**a**) Description of the sequence captured by the 3' exon trap cassette following somatic integration in the zebrafish genome. (**b**) Reverse northern blot and RT-PCR of RNA from 24-hour post fertilization wild type zebrafish embryos following somatic integration of a 3' exon trap performed on trapped sequences that map to the gene and transcript-free region of the zebrafish genome. The data reveals that a significant subset (75%) of these transcripts is expressed during this single stage in zebrafish development. (**c**) Distribution of the 3' exon trap integration within 100 well-annotated zebrafish genes. (**d**) Description of sequences trapped by the 3' exon trap cassette in the germline of zebrafish.

Supplementary Figure 3



Early generation protein trapping gene-break transposons. A schematic of R14, R15, and R16 are shown. Key differences from RP2 are the lack of the 3'exon trap and the use of a simple polyadenylation signal (poly(A)) from the zebrafish beta-actin gene instead of the poly(A) with a terminator and putative border element from the ocean pout antifreeze gene. R14 used the full-length Tol2 inverted terminal repeats (ITR). R15 uses the mini-Tol2 ITR sequences. R16 has a splice-acceptor from the zebrafish beta-actin gene rather than the carp beta-actin gene.

Supplementary Table 1

Oligo Name	DNA sequence
mRFP/Bam-F1	GGATCCACCATGGCCTCCTCCGAGGA
mRFP/Cla-R1	ATCGATTAGGCGCCGGTGGAGTG
cd99l2-F1	ATGGAGAAGACGCTCTGGACATG
mRFP-R1	CGATCTCGAACTCGTGGCCGTTCA
CDS-*mRFP-F1	AAGAATTCGAAGGTGCCTCCTCCGAGGATGTCATCAAGG
CDS-mRFP-R1	AAACTAGTCTTAGGCTCCGGTGGAGTGGCGG
5R-mRFP-P0	TGGAGCCGTACTGGAACTGA
5R-mRFP-P1	CGCCCTTGGTCACCTTCAGCTT
5R-mRFP-P2	CCTTGAAGCGCATGAACTCCTTGAT
INV-OPT-P1	ATCCCTCACTCCCGTAGCTGTCCA
INV-OPT-P2	GCGAGCCAAACCTGTCTGGTGTA
5R-GFP-P1	TGCATCACCTTCACCCTCTCCACT
5R-GFP-P2	GGAAATTTGTGCCCATTAACATCACC
Tol2-ITR(L)-O1	AATTAAACTGGGCATCAGCGCAATT
Tol2-ITR(L)-O3	CCAAGGGAAAATAGAATGAAGTGATCTCCA
RT-RPS6Kb1-F1	AAATCTCTATGGCGCTCGGACACC
RT-RPS6Kb1-R1	TGGACTCCTTACACGCCCGAAATC
GBM1	ATTCCGGCTGAACTGTAAATGAATG
GBM2	CTGTAAATGAATGAGAAAACCGGTT
Adapter primer	GGCCACGCGTCGACTAGTASTTTTTTTTTTTTTTTT

Oligo names and sequences. Oligos mentioned in the text or methods are listed here with corresponding sequence listed 5' to 3'. Morpholinos are underlined.

Supplementary Table 2

Vector	Line	Gene ID	QRT Forward Primer	QRT Reverse Primer
R15	GBT0001	*casz1*	TGTTTCTCCCAGGCCAGCTTACATC	GCTTGGCACAGATCGCATTCAGTT
R15	GBT0002	*si:ch211-51g4.4*	AACTTCTCCAGTGCCAAGCAAAGC	CCAGGCATTAGGACTGAAAGTAGGCA
R15	GBT0005	*itgb1b*	TGGACGTAAGGCTGCTCCTGATTTC	GCCGAAGCCTTGGTGCATTCAT
R16	GBT0007	*C14ORF102*	CTTCTCTGGACTGTCAGGCAGCAAC	GCTCGCTGGTGAGTCTTCAAAGCA
R14	GBT0019	*k2p10.1*	GGAAACAGGTGAACTGGGATCCTGAAC	GGGTTCTGCATGGCCGAAACA
R14	GBT0031	*tnnt2a*	GTTGCCATGTTTGAAAGCTGTGTGC	TCCATGAAAAAACAGAAGCCGTCAGT
R15	GBT0039	*gabbr1.2*	CGACCACCTAGAGATGGAGGGATCAG	CGCCCTTGGTCACCTTCAGCTT
RP2	GBT0040	*hoxA3a*	AAACGCACCGACCGGATTTACTC	CGCCGTTTCCCAATGTTTATTTCA
RP2	GBT0040	*hoxA4a*	AAACGCACCGACCGGATTTACTC	TTTTCAAGCTCAAGAGCCTGCTGG
R15	GBT0043	*CD9912*	TCGCAGATGCCCTAAACCCTGAC	TGTCGGGTTTGTAAGAGTTATCGTTTCC
RP2	GBT0067	*myom3*	TGACAGCAGAGGAGAAGACGACACTATC	GCAGTAAAGCTTGAAGCCCTGCG
RP2	GBT0078	*grip1*	GAAGAATGTGGGGGAAAGGGTTGT	CACCTCCACGTTTTTGAACATGACC
RP2	GBT0125	*LOC100333685*	TTATTCGGCGAGAGGAGAGGACG	TGAAGTGATAGCAGTTGCGTCCGAAT
RP2	GBT0126	*nrp2b*	TGGAAGAATGGAATCGCTTTACGCT	AATCCGGAAGAAGGCCAGAGAGC
RP2	GBT0137	*zgc:110335*	GGCTGGAAGATCACCCGAAAGG	ATTCCTGTCTCAACACGCCCCA
RP2	GBT0141	*gpm6ba*	AAACCAGAATCGACGGCTGAGGA	CAAAGAGGCATATGGGACACCACC
RP2	GBT0156	*fras1*	CTGCTCAAAGGCTCTGATCTTGAACC	CCACACAGACCCTACAGGACGAATG
RP2	GBT0348	*ryr1b*	GGAGCAGCTGAGATGGTTCTGCAA	TGTTGCCGCCATTAAGGATTGAGAT

Quantitative PCR primer pairs for GBT lines. Listed are the GBT vector, line designator, gene identifier, and the corresponding forward and reverse primer sequences used for QRT.

Supplementary Note.

3' exon-trap cassette efficiency in zebrafish. In Figure S2, we have generated random insertions of the 3' exon trap cassette in the somatic tissue of zebrafish to study the pattern of sequences trapped by the 3' exon trap cassette. A total of 1273 somatic insertion events were generated and analyzed by 3' RACE (Rapid Amplification of cDNA ends) technique as described[4]. An identity search using BLAT function[5] at UCSC Genome Browser (http://genome.ucsc.edu) against the zebrafish reference genome (Zv8 assembly) was used to map the trapped sequences to the genome. Genomic sequences with over 94% identity to the trapped sequence were considered as significant matches. The annotation status, exon prediction and exon–intron boundary of the trapped sequence were also identified where possible. The bioinformatics analysis and mapping of the captured sequences to the zebrafish genome reveal that 10% (n = 125) of the insertions occurred within well-annotated genes in the zebrafish genome. 6% (n = 75) of the insertions trapped un-annotated transcripts or expressed sequence tags in the zebrafish genome. 67% (n = 850) of the insertions mapped to gene and transcript-free regions in the zebrafish genome, and 17% (n = 223) of the trapped sequences were considered as low complexity based on the sequence composition or repetitive sequences for which we could not assign any genomic location.

Since a large proportion (67%) of the trapped sequences mapped to gene and transcript-free regions of the zebrafish genome, we conducted reverse northern blotting and RT-PCR using RNA from 24-hour post fertilization wild type embryos. This demonstrated that a subset (75%) of the trapped sequences that map to the gene and transcript-free region of the zebrafish genome are actually expressed during zebrafish development (Fig. S2B). We also studied the distribution of the 3' exon trap integration within 100 well-annotated zebrafish genes (Fig. S2C). The 3' exon trap cassette is capable of trapping most exons of genes with almost equal efficiency except for the exons at the 5' end of a gene.

In addition to the somatic integration data, we have collated the germline integration data of 3'exon trapping from two published articles[4,6]. The 3' exon trapped sequences were used to conduct an identity search using BLAT function[5] at UCSC Genome Browser (http://genome.ucsc.edu) against the zebrafish reference genome (Zv8 assembly). Genomic sequences with over 94% identity to the trapped sequence were considered as significant matches. The annotation status, exon prediction and exon–intron boundary of the trapped sequence were also identified.

A total of 28 trapped genomic sequences were analyzed from both published articles. The 3'exon trap cassette could capture known and putative novel endogenous transcripts (n = 22, 78%) in the zebrafish genome. 21% (n = 6) of the time the 3' exon trap captured sequences that were considered as low complexity based on sequence composition or repetitive sequences for which we could not assign any genomic location. We also studied the distribution of the 3' exon trap integration within 8 well-annotated zebrafish loci (Fig. S2D). The 3' exon trap cassette is capable of integration within genes of diverse genetic structures in the zebrafish genome. Importantly, the use of the ß-actin promoter may aid in protection against transcript degradation from nonsense-mediated decay as was observed in mammalian systems using the human ß-actin promoter[7].

Supplementary References:

1 Thisse, B. *et al.* Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). ZFIN Direct Data Submission, http://zfin.org (2001).

2 Thisse, B. & Thisse, C. Fast release clones: a high throughput expression analysis. ZFIN Direct Data Submission, http://zfin.org (2004).

3 Thisse, B. & Thisse, C. Hight throughput expression analysis of ZF-models consortium clones. ZFIN Direct Data Submission, http://zfin.org (2005).

4 Sivasubbu, S. *et al.* Gene-breaking transposon mutagenesis reveals an essential role for histone H2afza in zebrafish larval development. *Mech Dev* **123**, 513-529, doi:10.1016/j.mod.2006.06.002 (2006).

5 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664, doi:10.1101/gr.229202. Article published online before March 2002 (2002).

6 Petzold, A. M. *et al.* Nicotine response genetics in the zebrafish. *Proc Natl Acad Sci USA*, doi:10.1073/pnas.0908247106 (2009).

7 Tsakiridis, A. *et al.* Expression-independent gene trap vectors for random and targeted mutagenesis in embryonic stem cells. *Nucleic Acids Res* **37**, e129, doi:10.1093/nar/gkp640 (2009).