RP2	GBT01565	*fras1*	FRAS1		ND	* fin	SRAGHCH	652
RP2	GBT016623	*atp1b2a*	ATP1B2	atp1b2b	ND	None (5 DPF)	GRTASSW	33
RP2	GBT016814	*fgf13 (1y+1v)*	FGF13	fgf13l	ND	None (5 DPF)	KENSEPE	72
RP2	GBT01756	*arhgef25b*	GEFT		ND	None (5 DPF)	VCCFNQK	89
RP2	GBT02317	*neo1*	NEO1		ND	None (5 DPF)	AITRPQS	810
RP2	GBT02427	*zgc:110022*	TEX261		ND	ND	CFVTLAI	23
RP2	GBT028324	*LOC793623*	SH3KBP1		ND	ND	MDNEAEK	209
RP2	GBT032511	*LOC557764*	MEGF6	megf6a	ND	ND	TGVVCNE	446
RP2	GBT03407	*NFATC3*	NFATC3	LOC566869	ND	ND	QPPLGPA	30
RP2	GBT034818	*ryr1b*	RYR1	ryr1a	3	* swim	QMISACK	1444
RP2	GBT03638	*XU0032*			0	* lethal		
RP2	GBT036410	*XU0135*			0.3	* lethal		
RP2	GBT036512	*XU0231*			0.1	None (5 DPF)		

Table 1: Gene disruptions of GBTs. The column categories present data for each molecularly characterized GBT line. The putative human ortholog (GBT0137 lists the putative mouse ortholog) and putative zebrafish (zf) co-ortholog are listed. % mRNA denotes the relative mRNA remaining in homozygous GBT larvae; in some cases, the transposon inserts into an exon directly disrupting the transcript. Fusion tag lists the seven proximal amino acids at the fusion site with mRFP. Pos indicates the relative position of the truncation in the predicted wild-type protein; for a description of special fusions a, b, and c please refer to the methods. LG, linkage group; DPF, days post-fertilization; HV, homozygous viable as adults; sih, silent heart; nic, nicotine response; swim, reduced swimming; fin, fin malformation; ND, no data.

Online Methods

All zebrafish work was conducted under institutional IACUC-approved protocols.

Vector construction. The following annotated plasmid sequences have been submitted to Genbank: *pGBT-R14* (HQ335167), *pGBT-R15* (HQ335168), *pGBT-R16* (HQ335169), *pGBT-RP2.1* (HQ335170). *pT3TS-Tol2* (HQ335171), *pT3TS-Cre* (HQ335172), and *pGBT-PX* (HQ335166). The fusion of protein trap components from the R-series and poly(A) trap components from the P-series of GBT vectors were used to create the RP series of vectors. The GFP variant used in *pGBT-PX* and *pGBT-RP2.1* is GFPmut2 or GM2[30]. The ocean pout antifreeze polyadenylation sequence, transcriptional terminator, and putative border element used in *pGBT-PX* and *pGBT-RP2* were cloned from *pFRM2*[31].

pEF1a-RFP was made by ligating a 685 bp BamHI to ClaI mRFP fragment into *pDB774* (an *EF1α-GFP* vector) cut with BamHI and ClaI to remove the GFP cassette. The mRFP was originally amplified from source material obtained from the Tsien lab using mRFP/Bam-F1 and mRFP/Cla-R1 primers (all primer sequences are available in Supplementary Table 1)[32].

pEF1a-cd99l2-mRFP was made by combining two PCR products in vitro and subcloning. The first PCR product, the cd99l2 N-terminal fusion to mRFP, was obtained from cDNA of GBT0043 fish using cd99l2-F1 and mRFP-R1 primers. The second was the mRFP product obtained by mRFP/Bam-F1 and mRFP/Cla-R1. The two PCR products were mixed and amplified with cd99l2-F1 and mRFP/Cla-R1 to make the full-length fusion transgene, which was cloned into pJet1 (Fermentas). The cd99l2-mRFP cassette was excised from pJet1 with BglII and ClaI and subcloned into *pDB774*.

pDB774 was produced by cloning the NruI to SphI fragment of *pT2/S2EF1a-GFP* (aka *pDB371*)[10] into *pmini-Tol2* (*pDB739*)[4] digested with MscI.

*pCR4-*mRFP* was produced by subcloning a fragment of mRFP into *pCR4-Topo* (Invitrogen). The mRFP fragment was amplified from *pGBT-R15* using CDS-*mRFP-F1 and CDS-mRFP-R1 primers.

Production of tagged fish lines. Fertilized embryos were obtained from wild-type strains of adult zebrafish. Single cell embryos were injected with 1-2 nL of a combination of protein trap transposon and *Tol2* mRNA at 12.5 ng/uL each. The injected embryos were raised at ~29º C. For RP2 injections, the injected embryos were sorted using the GFP fluorescence at 3-4 days post-fertilization. In summary, a small amount of GFP fluorescence was considered class I, intense GFP covering >40% of the embryo in mosaic patches was considered class II, and GFP (often low intensity) covering >80% of the embryos or having uniform expression in the lenses or brains of the larvae was considered class III. Class III embryos are thought to represent an early transposition event into a transcription unit that results in the widespread expression of GFP in these larvae. Thus, class III represent the best population to produce transgenic offspring and were raised to adulthood (F0 generation). The number of class III larvae ranges from 5-50% of injected embryos with an average of about 20%.

F1 embryos were obtained by crossing the F0 generation to non-transgenic brood stock. The embryos were scored for mRFP expression. In a recent assessment, 32% of class III fish (n = 173) produced some F1 embryos with mRFP expression compared to 20% (n = 127) and 7% (n = 61) of class II and class I fish, respectively. mRFP-expressing embryos are sorted by pattern if possible, assigned a GBT identifier, and raised to adulthood. Imaging and molecular work are done on F2 or subsequent generation embryos.

Standard imaging of GBT lines. The mRFP expression pattern of each GBT line was recorded at both 2 and 4 days post-fertilization if mRFP was expressed. Using the SCORE imaging system, coronal, sagittal, and/or ventral oriented z-stacks were obtained at 50X magnification as described[33].

Identification of tagged genes. To determine the identity of the gene with an activated protein trap, 5'RACE was performed as previously described[34], with minor modifications to primer sequences. Briefly, total RNA was isolated from 20 mRFP positive embryos. cDNA was produced from 250ng of total RNA using a gene specific primer 5R-mRFP-P0 for the reverse transcriptase reaction. PCR is performed with the following gene-specific primers 5R-mRFP-P1 and 5R-mRFP-P2. The resulting products were cloned, sequenced, and an in-frame fusion with the mRFP was verified. The cloned sequences obtained by 5'RACE are available on http://zfishbook.org.

Alternatively, inverse PCR was used to identify the interrupted gene. A restriction enzyme cocktail of AvrII, NheI, SpeI, and XbaI was used to digest about 800 ng of genomic DNA. About 200 ng of this digestion was self-ligated in a 100µL reaction. The ligation reactions were diluted 10-fold and used as template for inverse PCR. Primary and nested PCR primers used for the 5' side (mRFP-side) included 5R-mRFP-P1 and 5R-mRFP-P2 paired with INV-OPT-P1 and INV-OPT-P2, respectively. The primers used for the 3' side (GFP side) were 5R-GFP-P1 and 5R-GFP-P2 with Tol2-ITR(L)-O1 and Tol2-ITR(L)-O3, respectively. Following an initial denaturation of two minutes at 95º, the primary and secondary PCR reactions were cycled 30 times with 30 seconds of denaturation, 30 seconds of annealing at 55º C, and 6 minutes of extension. The primary PCR reactions are diluted 50-fold prior to preparing the nested PCR reactions. The resultant products were gel-isolated, cloned, and sequenced. The cloned

sequences obtained by inverse PCR are available on http://zfishbook.org. Candidates with in-frame fusion with mRFP were verified for linkage to the GBT expression pattern by PCR on genomic DNA or cDNA from mRFP carrier siblings versus non-carrier siblings.

Identification of orthologs and fusions. Orthologs listed in Table 1 were determined using NCBI's Homologene database. Orthologs for GBT0001, GBT0007, GBT0040, GBT0067, GBT0091, and GBT0325 were identifed as a best match following a BLASTX search of the human genome. Special fusions from Table 1 include: a) An insert into the 5'UTR of *si:ch211-51g4.4* that uses an upstream AUG in the 5'UTR to create this fusion. b) An insert in the HoxA cluster that uses a shared exon with unique protein fusion sequences. c) An insert into the first annotated *dido1* exon results in two transcripts. The first transcript uses only the two annotated upstream exons that are non-coding. The second transcript creates a fusion protein that uses an alternate exon between the second and third annotated exon. For each locus, nucleotide sequences encoding the mRFP fusion proteins obtained by 5' RACE are available at http://zfishbook.org.

Identification of transposon integration site. To design primers to genotype offspring of a heterozygous incross, the integration site of a transposon within an intron was required. With the target intron known following 5' RACE, primers were designed in the flanking exons of the "tagged" gene priming towards the intron and the protein trap transposon. The flanking gene-specific primers were paired with either 5R-mRFP-P2 or Tol2-ITR(L)-O1 to amplify the junction fragment of the transposon. In some cases where the introns exceeded 8kb, primers alternating in orientation were designed every 3kb across the intron and were similarly used with the 5R-mRFP-P2 or Tol2-ITR(L)-O1

primers. The junction fragments were cloned, sequenced, and used to identify the integration site of the transposon.

Quantitative RT-PCR. Individual embryos of a hetereozygous incross breeding were lysed using Trizol reagent (Invitrogen). Total RNA was purified as indicated with the addition of 1µL of glycoblue (Ambion) as a carrier. Genomic DNA was purified after back-extraction of the organic phase and interphase using 4M guanidine thiocyanate, 50mM sodium citrate, and 1M Tris pH 8.0. The genomic DNA in the new aqueous phase was moved to a new tube, 1µL of glycoblue (Ambion) was used as a carrier, and the DNA was precipitated by the addition of an equal volume of isopropanol. The pellet was washed with 70% ethanol and resuspended in 10µL of TE. The genomic DNA was used with genotyping primers to determine which embryos were wildtype (+/+), heterozygous (R/+), or homozygous (R/R) for each transposon trap tested. RNA from four embryos of each class was used to produce cDNA using random hexamer primers. Primers in the exons flanking the transposon insertion site (GS-F1/GS-R1- see Table S2 for sequences) were used to test for "wildtype" product that occurs when the two exons are properly spliced together. The GS-F1 primer was used with 5R-mRFP-P2 to examine the amount of RFP fusion transcript. In both cases the products were referenced against the level of ribosomal protein S6 kinase b, polypeptide 1 using RT-RPS6Kb1-F1 and RT-RPS6Kb1-R1. In cases where genotyping of embryos was possible, due to a distinguishable phenotype or availability of viable homozygous adults, genotyped embryos were collected in groups of ten, and the cDNA was tested as above with three technical replicates.

***in situ* hybridization.** *in situ* hybridizations were performed using digoxygenin-labeled probe following previously published protocol[35]. The mRFP probe was made by

linearizing 1 µg of *pCR4-mRFP* plasmid with PmeI. To the purified DNA template, we added 10X Dig RNA labelling mix (Roche) and T7 RNA polymerase (Roche) for 2 hours at 37º C to make the probe. After 2 hours, the plasmid was digested using Rnase-free DnaseI (Promega) and the probe purified using RNA easy MinElute clean up kit (Qiagen).

Morpholino and Cre reversion. Embryos from GBT crosses were injected with a mixture of two morpholinos, GBM1and GBM2 (see Supplementary Table 1) that target the exogenous splice acceptor in pRP2 derived from carp beta-actin intron 1. Injection of 3nL of a GBM1 and GBM2 morpholino mixture at concentrations of 100µM was used to revert the GBT0031 phenotype. Cre mRNA was produced using mMessage machine (Ambion) from pT3TS-Cre plasmid linearized by SacI digestion. About 25pg of *Cre* mRNA was injected into each embryo to revert the GBT0031 and GBT0156 phenotypes. We followed general considerations for morpholino use in zebrafish[36].

Somatic efficiency of 3' exon trap. Total RNA was isolated from batches of 30 GFP expressing 3 DPF zebrafish embryos using Trizol reagent (Invitrogen). First strand cDNA was synthesized by using 5 µg of the RNA and Superscript II Reverse Transcriptase (Invitrogen) with a final reaction mixture volume of 20 µL. Two rounds of 3' RACE-PCR were performed using 2 µL of this first strand cDNA as template as described[6]. 3' RACE-PCR products were treated with SpeI and SphI restriction enzymes in a final reaction volume of 100 µL and were incubated at 37º C for 8 hr. The SpeI and SphI digested 3' RACE-PCR products were purified and ranged from 70bp to 4kb. Purified SpeI and SphI treated 3' RACE-PCR products were cloned using *pCR-4-TOPO* TA Cloning vector (Invitrogen) in a final reaction volume of 6 µL. 2 µL of this reaction mixture was transformed into TOP10 competent cells (Invitrogen). Clones

were grown on carbenicillin resistance plates. Subsets of the clones were inoculated into 96-well deep well plates (Promega) containing 1.3 ml of the culture media per well. The inoculated plates were incubated at 37º C with 250 rpm for 20 hrs at an inclination of 45º C to increase the surface area. After incubation the plates were centrifuged at 3000 x g for 5 minutes to pellet down the bacterial cells. Plasmids were isolated from these pellets and were subjected to capillary sequencing. The clones containing inserts were sequenced and mapped to the zebrafish genome and analyzed as described[6]. Trapped sequences mapping to Expressed Sequence Tags (EST) and Genomic loci with ≥94% identity over a region of ~ 100bp were considered as significant match.

Validation of somatic-trapped sequences by reverse northern dot blot assay.

Validation of trapped sequences for which identities were found only in the Ensembl zebrafish genomic DNA database with no representation found in the current EST databases was done using reverse northern dot blot assay. RNA was isolated from batches of 50 wild type 24-hpf zebrafish embryos. 50 µg of freshly isolated RNA is incubated with 5 µL of 10 µM Adapter primer in a final volume of 35 µL. This mixture is incubated at 65º C for 10 minutes, 25º C for 5 minutes, 42º C for 3 minutes followed by addition of 1 µL of 100 µM dATP, 1 µL of 100 µM dGTP, 1 µL of 100 µM dTTP, 5 µL of 3500 milicurie per mM α-labeled dCTP, 2 µL of M-MLV Reverse Transcriptase and 5 µL of 10X RT Buffer. This mixture is incubated at 42º C for 1 hr and then chilled at 4º C for 5 minutes. To this reaction, 5 µL of each 0.5M EDTA and 1N NaOH is added and then heated to 65º C for 30 minutes. This mixture is chilled in ice and then 6.5 µL 1M Tri-Cl, pH 7.5 is added. Prior to this, clones containing inserts for which identities were found only in the Ensembl zebrafish genomic DNA database with no representation found in the current EST databases were spotted on 0.45 µM

nitrocellulose blotting membrane (MDI). This membrane is then UV cross-linked using UV cross linker and then hybridized for 1 hr at 37º C using pre-hybridization buffer containing DIG Easy Hyb Buffer (Roche). The radioactive probe prepared above is heated at 95º C for 2 minutes and added to the pre-hybridization buffer directly. The membrane is incubated overnight at 37º C. After the hybridization, the membrane is washed twice - initially with 1X SSC and 0.1 % SDS for 15 minutes at 37º C and then with 0.5X SSC and 0.5 % SDS for 15 minutes at 37º C. As soon as washes are done, the membrane is wrapped with saran wrap sheet and exposed to a Phosphoimager screen for ~ 24 hours to obtain an autoradiographic image.

References:

1 Stevens, C. W. The evolution of vertebrate opioid receptors. *Front Biosci* 14, 1247-1269 (2009).

2 Huxley-Jones, J., Robertson, D. L. & Boot-Handford, R. P. On the origins of the extracellular matrix in vertebrates. *Matrix Biol* 26, 2-11, doi:10.1016/j.matbio.2006.09.008 (2007).

3 Sauka-Spengler, T. & Bronner-Fraser, M. Evolution of the neural crest viewed from a gene regulatory perspective. *Genesis* 46, 673-682, doi:10.1002/dvg.20436 (2008).

4 Balciunas, D. *et al.* Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet* 2, e169, doi:10.1371/journal.pgen.0020169 (2006).

5 Urasaki, A., Morvan, G. & Kawakami, K. Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* 174, 639-649, doi:10.1534/genetics.106.060244 (2006).

6 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).

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

8 Branda, C. S. & Dymecki, S. M. Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* 6, 7-28 (2004).