In vivo protein trapping produces a functional expression codex of the vertebrate proteome

Karl J Clark[1], Darius Balciunas[2,3], Hans-Martin Pogoda[4], Yonghe Ding[1], Stephanie E Westcot[1,2], Victoria M Bedell[1], Tammy M Greenwood[1], Mark D Urban[1], Kimberly J Skuster[1], Andrew M Petzold[1,2], Jun Ni[1], Aubrey L Nielsen[2], Ashok Patowary[5], Vinod Scaria[5], Sridhar Sivasubbu[2,5], Xiaolei Xu[1], Matthias Hammerschmidt[4] & Stephen C Ekker[1,2]

We describe a conditional *in vivo* protein-trap mutagenesis system that reveals spatiotemporal protein expression dynamics and can be used to assess gene function in the vertebrate *Danio rerio*. Integration of pGBT-RP2.1 (RP2), a gene-breaking transposon containing a protein trap, efficiently disrupts gene expression with >97% knockdown of normal transcript amounts and simultaneously reports protein expression for each locus. The mutant alleles are revertible in somatic tissues via Cre recombinase or splice-site-blocking morpholinos and are thus to our knowledge the first systematic conditional mutant alleles outside the mouse model. We report a collection of 350 zebrafish lines that include diverse molecular loci. RP2 integrations reveal the complexity of genomic architecture and gene function in a living organism and can provide information on protein subcellular localization. The RP2 mutagenesis system is a step toward a unified 'codex' of protein expression and direct functional annotation of the vertebrate genome.

Innovation and evolution have led to diverse biological systems and biochemical pathways in vertebrates[1–3]. Understanding which genes are necessary, important or advantageous for survival in complex, multicellular organisms requires an examination of gene expression and function. The cumulative information on a single gene can be thought of as a 'codex', which contains multiple types of data including sequence, expression domains and function, with each data type containing multiple observations (for example, sequence variations and dynamic expression over time or in response to genetic or environmental cues). Here we present an *in vivo* mutagenesis tool that bridges the gap between sequence data and gene function. The pGBT-RP2.1 (RP2) protein-trap system recapitulates endogenous gene expression, disrupts gene splicing with nearly all tested lines displaying >99% knockdown of the native transcript and allowed us to generate a systematic collection of conditional mutants in the zebrafish. These gene-break transposons (GBTs) are to our knowledge the first *in vivo* protein-trap alleles in a vertebrate with reporter production and line selection occurring in the fish rather than using *in vitro* cells for production or molecular characterization.

RESULTS

Genomic assessment tool RP2

We developed a GBT mutagenesis system for the vertebrate model system *Danio rerio* to facilitate genome annotation and understanding of gene function. The pGBT-RP2.1 (RP2) vector has several features that efficiently cooperate to report gene sequence, expression and function (**Fig. 1**). The mutagenic core of RP2 contains two 'trap' domains that are used to 'capture' genomic information by affecting transcription, resulting in efficient knockdown of endogenous loci by a trapping vector in zebrafish (**Table 1**). These two key components, a protein trap and a 3′ exon trap, are placed in inverted terminal repeats of the *Tol2* transposon to permit efficient genome-wide delivery[4,5].

The protein-trap domain in RP2 generates the expression profile, including subsequent protein localization and accumulation and mutates the gene. In cases where RP2 integrates in the sense orientation of a transcription unit, the protein trap's splice acceptor over-rides normal splicing of the transcription unit, creating a fusion between endogenous upstream exons and the monomeric RFP (*mRFP*) reporter sequences. An in-frame fusion between the reporter and the tagged protein is required to produce red fluorescence owing to removal of the start codon from *mRFP*. As the trapped gene's promoter produces the *mRFP* fusion transcript, it is made when and where the endogenous gene's mRNA is transcribed (**Supplementary Fig. 1**). We note that the localization of the mRFP fusion protein can depend on protein trafficking signals in the trapped protein fragment. The efficient mutagenesis observed in RP2-disrupted loci occurs by termination of the fusion transcript through signals derived from the ocean pout antifreeze gene, consisting of a strong polyadenylation signal (poly(A)), transcriptional terminator and putative border element[6].

The second key domain of RP2 is the previously described 3′ exon trap or poly(A) trap used in zebrafish forward genetic applications[6,7]. GFP activation indicates a high likelihood of transposon insertion in the sense orientation of a transcription unit and does not require concomitant endogenous gene expression. Indeed, all identified mRFP-expressing loci also exhibit

[1]Mayo Clinic, Rochester, Minnesota, USA. [2]University of Minnesota, Minneapolis, Minnesota, USA. [3]Temple University, Philadelphia, Pennsylvania, USA. [4]University of Cologne, Cologne, Germany. [5]Institute of Genomics and Integrative Biology, Council of Scientific and Industrial Research, Delhi, India. Correspondence should be addressed to S.C.E. (ekker.stephen@mayo.edu).

RECEIVED 14 FEBRUARY; ACCEPTED 8 APRIL; PUBLISHED ONLINE 8 MAY 2011; DOI:10.1038/NMETH.1606

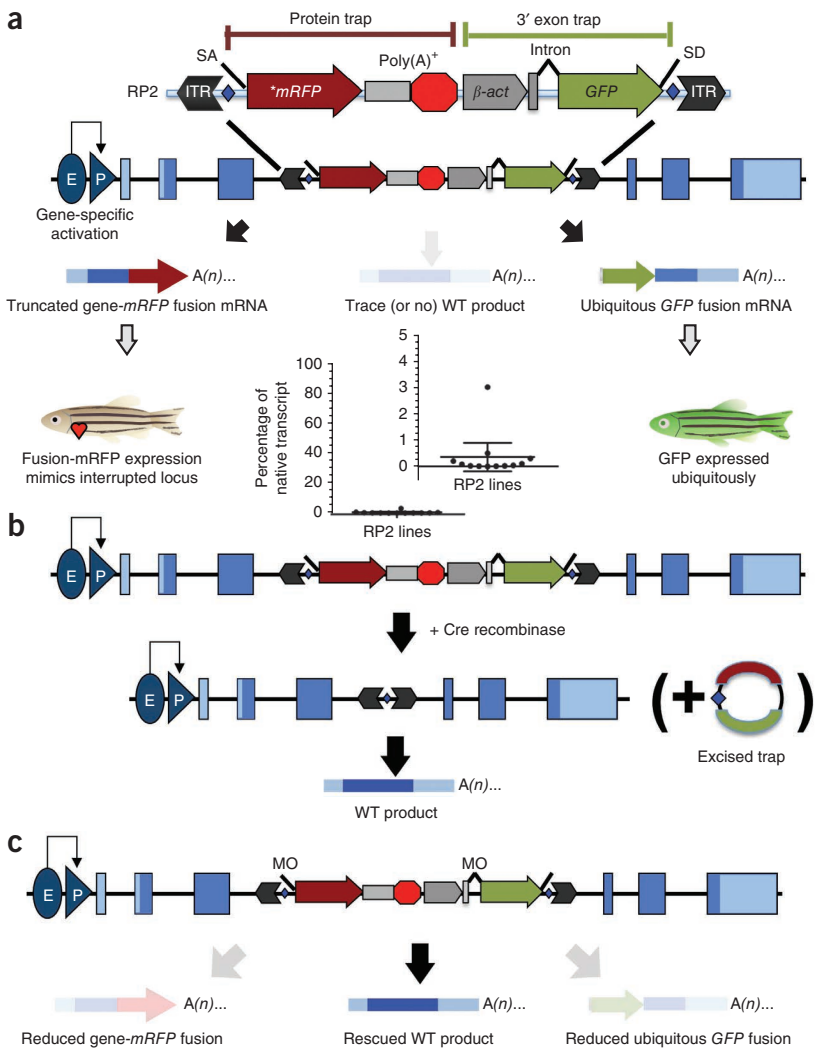

Figure 1 | RP2 gene-break transposon and reversion systems. (**a**) Schematic of the RP2 gene-break transposon system composed of a protein-trap cassette with a transcriptional stop and 3′ exon trap cassette. The interrupted wild-type mRNA is either not produced or is produced in trace amounts. ITR, inverted terminal repeat; SA, splice acceptor; **mRFP*, AUG-less *mRFP* sequence; poly(A)$^+$, polyadenylation signal with extra transcriptional terminator and putative border element; *β-act*, carp beta-actin enhancer, promoter, noncoding exon and intron 1 sequences; SD, splice donor; E, enhancer; P, promoter; and WT, wild type. Scatter plot shows remaining native transcript amounts in RP2 homozygous larvae with the mean and 95% confidence interval indicated by lines (n = 12). (**b**) Schematic of RP2 reversion with Cre recombinase. Blue diamonds show *loxP* sites flanking the mutagenic cassettes. By supplying Cre recombinase, the protein and 3′ exon traps are excised, resulting in normal levels of the nascent mRNA. (**c**) Schematic of RP2 reversion with splice acceptor–masking morpholinos. Both splice acceptors are derived from carp beta actin intron 1. MO, splice acceptor masking morpholino.

linked GFP expression (**Fig. 2a** and Online Methods). Variability of GFP expression is generally limited to ubiquitous changes in GFP signal intensity. Extensive sequencing of molecular tags created by 3′ exon trap activation in zebrafish somatic tissue demonstrated the ability of this vector to identify native transcripts, including new RNA species (**Supplementary Note** and **Supplementary Fig. 2**).

An RP2 protein-trap–mutagenized gene produces mRFP expression that mimics endogenous expression of the gene and quantitatively removes the normal full-length protein. To permit selective correction of the mutagenized locus, we flanked RP2 with Cre recombinase recognition sites (*loxP*) to excise the mutagenic core, leaving behind only the *Tol2* inverted terminal repeats[8] (**Fig. 1b**). In contrast to common mouse conditional mutagenesis methods, RP2 mutagenic core excision results in conditional rescue when the integration occurs in an intron. To revert the GBT mutation in the entire embryo, *Cre* mRNA is delivered at the single-cell stage. Alternatively, inducible or tissue-specific expression of Cre recombinase can result in conditional rescue, permitting investigation of the spatiotemporal requirements for the mutagenized locus.

Cre recombinase–mediated excision results in genetically irreversible deletion of the mutagenic core. However, many genes have additional functions later in development or have adult functions that differ from their early embryonic roles. In these cases, a transient rescue strategy that bypasses the early embryonic roles or functions may be preferable. GBT splice acceptor–targeted antisense morpholino oligonucleotides can be used to transiently rescue a GBT mutation (**Fig. 1c**) for 3 or 4 d after fertilization, permitting the study of a gene's function later in development. Both the primary and secondary splice acceptors, in front of the protein trap and 3′ exon trap *GFP*, respectively, are derived from the same source. Therefore, antisense morpholinos targeted to the common splice acceptor sequence effectively mask both splice acceptors from the spliceosome (**Fig. 1c**), resulting in a transient boost of the wild-type product from the GBT-mutated locus. This strategy can provide rapid proof that rescued phenotypes are caused by the protein trap affecting splicing of the mutant locus.

Expression profiling with RP2

Unlike prior zebrafish vectors (enhancer traps)[9–11], the requirement in RP2 of an in-frame translational fusion means that the expression pattern obtained from a GBT protein trap is directly associated with the tagged gene product (**Fig. 1a**). We prioritized and cataloged lines with robust expression at 2 d and 4 d after fertilization. We are continuing to generate GBT strains and will provide updated results at zfishbook (http://zfishbook.org/). All vectors and zebrafish lines are freely available now through zfishbook, and, once the collection is synchronized, it will be accessible from the Zebrafish International Resource Center.

In our initial collection of 350 protein-trap lines with mRFP expression, about one-third of the lines have neuronal expression, with instances of both diverse and redundant expression domains (**Fig. 2a**). For instance, multiple GBT lines express mRFP in sensory ganglia (for example, GBT0001 and GBT0019)

Table 1 | Gene disruptions of GBTs

Vector	Line	Linkage group	Gene	Ortholog	Zebrafish co-ortholog	mRNA remaining[a] (%)	Phenotype[b]	Fusion tag[c]	Position[d]
R14	GBT0019	17	*k2p10.1*	*KCNK10*	*LOC563228*	20.7	None (HV)	QVNWDPE	17
R14	GBT0031	23	*tnnt2a*	*TNNT2*		5.6	Silent heart	EETTQEH	50
R14	GBT0046	2	*epha4b*	*EPHA5*	*epha4a, ek1*	ND	None (HV)	RNYPENE	42
R15	GBT0001	23	*casz1*	*CASZ1*		13	None (5 d.p.f.)	M	1
R15	GBT0002	14	*si:ch211-51g4.4*	*SORBS2*	*sorbs2a*	24.3	None (5 d.p.f.)	MPSFK	5 (i)
R15	GBT0005	2	*itgb1b*	*ITGB1*	*itgb1a*	33.3	None (HV)	GLSRAQQ	19
R15	GBT0010	7	*cdh11*	*CDH11*		ND	None (5 d.p.f.)	FSLKDNR	551
R15	GBT0039	19	*gabbr1.2*	*GABBR1*	*gabbr1a*	<5	Nicotine response (HV)	HYDRHYT	161
R15	GBT0043	7	*cd99l2*	*CD99L2*		14.4	None (HV)	GKDSGKG	108
R16	GBT0007	17	LOC794348	*C14ORF102*		0.1	None (HV)	TALQKVK	~20
R16	GBT0016	6	*pbx1a*	*PBX1*	*zgc:15882*	ND	None (HV)	CEIKEKT	87
R16	GBT0021	22	*si:ch73-150k18.1*	*CNTNAP5*	*LOC569185*	exon	None (HV)	VHGEGQR	285
RP2	GBT0033	7	*si:dkeyp-73c8.2*	*LRCH4*		ND	None (5 d.p.f.)	LSDITHA	57
RP2	GBT0035	13	LOC559134	PARG	si:dkey-259k14.2	ND	None (5 d.p.f.)	NECLIIT	578
RP2	GBT0040	19	*hoxa5a–hoxa3a*	*HOXA* cluster	*hoxab* cluster	0.1	None (HV)	RGPALVQ	20 (ii)
RP2	GBT0060	9	*crfb4*	*IL10RB*		ND	None (5 d.p.f.)	NIIGVIT	19
RP2	GBT0067	16	*myom3*	*MYOM3*		0.2	None (HV)	LQLSKQC	~263
RP2	GBT0078	4	*grip1*	*GRIP1*		ND	Reduced adult viability	SLLKNVG	151
RP2	GBT0091	9	*enox1*	*ENOX1*		ND	ND	TGQQLVS	49
RP2	GBT0101	11	*dido1*	*DIDO1*		Exon	Lethal 10–16 d.p.f.	LLHHIHS	38 (iii)
RP2	GBT0125	9	LOC100333685	*CD302*		0.02	None (HV)	ARGEDDN	36
RP2	GBT0126	9	*nrp2b*	*NRP2*	*nrp2a*	0.04	None (HV)	LYGCQIT	427
RP2	GBT0133	9	*zic2a*	*ZIC2*	*zic2b*	Exon	None (5 d.p.f.)	DSAHMGA	47
RP2	GBT0137	1	*eef1a1*	*eef1a1*	*e1a, zgc:73138*	0.5	None (HV)	QDVYKIG	257
RP2	GBT0141	1	*gpm6ba*	*GPM6B*	*gpm6bb*	0.02	None (HV)	DERDESK	21
RP2	GBT0145	3	LOC560542	*EPN2*		ND	None (5 d.p.f.)	SPASYHG	198
RP2	GBT0154	8	*si:ch211-163121.8*	*KIAA1324*		ND	None (5 d.p.f.)	GFYSNGT	351
RP2	GBT0156	5	*fras1*	*FRAS1*		ND	Fin malformation	SRAGHCH	652
RP2	GBT0166	23	*atp1b2a*	*ATP1B2*	*atp1b2b*	ND	None (5 d.p.f.)	GRTASSW	33
RP2	GBT0168	14	*fgf13a (1y+1v)*	*FGF13*	*fgf13b*	ND	None (5 d.p.f.)	KENSEPE	72
RP2	GBT0175	6	*arhgef25b*	*ARHGEF25*	LOC100331456	ND	None (5 d.p.f.)	VCCFNQK	89
RP2	GBT0231	7	*neo1*	*NEO1*		ND	None (5 d.p.f.)	AITRPQS	810
RP2	GBT0242	7	*zgc:110022*	*TEX261*		ND	ND	CFVTLAI	23
RP2	GBT0283	24	LOC793623	*SH3KBP1*		ND	ND	MDNEAEK	209
RP2	GBT0325	11	LOC557764	*MEGF6*	*megf6a*	ND	ND	TGVVCNE	446
RP2	GBT0340	7	*nfatc3*	NFATC3	LOC566869	ND	ND	QPPLGPA	30
RP2	GBT0348	18	*ryr1b*	*RYR1*	*ryr1a*	3	Reduced swimming	QMISACK	1,444
RP2	GBT0363	8	*xu0363Gt*			0	Lethal		
RP2	GBT0364	10	*xu0364Gt*			0.3	Lethal		
RP2	GBT0365	12	*xu0365Gt*			0.1	None (5 d.p.f.)		

Putative human ortholog (GBT0137 lists the putative mouse ortholog) and putative zebrafish (zf) co-ortholog are listed. ND, no data.
[a]mRNA remaining in homozygous GBT larvae relative to wild-type transcript levels; in some cases, the transposon inserted into an exon directly disrupting the transcript. [b]HV, homozygous viable as adults. [c]Seven proximal amino acids at the fusion site with mRFP. [d]Truncation position of the predicted wild-type protein; fusions i, ii and iii are described in Online Methods.

or neuromasts (for example, GBT0002 and GBT0157). Although these lines have similar expression domains in sensory ganglia or neuromasts, mRFP expression in these lines differs in other tissues. In many cases, fluorescent puncta appear in a general neural background of mRFP, with distinct differences in the number, location or coverage of mRFP fluorescence spots (for example, GBT0016 and GBT0141). In some cases, a GBT insertion only affects a single protein isoform. For example, owing to the integration site in *fgf13a*, GBT0168 only traps the 1y+1v isoform of *fgf13a* (ref. 12). As the complexity of the protein-trap expression library increases, each line will contribute to its own codex as well as enable morphological, developmental and molecular annotation of the zebrafish.

Annotation of transcriptional organization with RP2

Hox clusters are highly conserved among animals, but the mechanism underlying this strict conservation remains largely unclear[13]. GBT0040 is an insertion in the zebrafish *hoxaa* cluster, between the annotated exons of the *hoxa5a* and *hoxa4a* genes resulting in an mRFP pattern that resembles *hoxa4a* expression (**Fig. 2b**). Upstream exons fused to the mRFP transcript in GBT0040 contained two exons, one noncoding exon 5′ of *hoxa5a* and one protein-coding exon between *hoxa5a* and *hoxa4a* (**Fig. 2b**). Transcriptional assessment demonstrated that the protein-coding exon splices to exons downstream of both *hoxa4a* and *hoxa3a* to produce protein-coding sequences with new N-terminal sequences (**Fig. 2b**). Notably, this single insertion results in the effective loss (>99% knockdown) of both of these alternate mRNA transcripts (**Fig. 2b**). Exon sharing and alternate transcripts in *hox* clusters has been previously observed in the transcriptome[14,15]. The exon sharing and alternate transcripts apparent in the GBT0040 allele suggest a mechanism underlying the conserved retention of *hox* genes in a single genomic cluster.

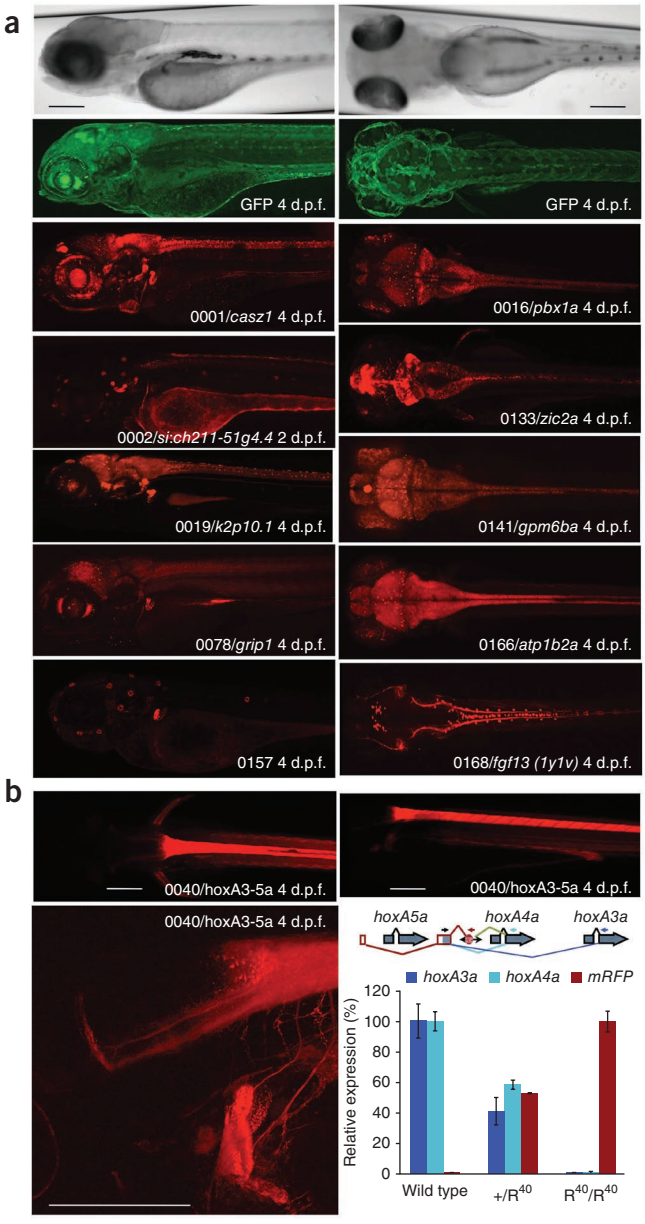

Figure 2 | Protein expression codex: examples of protein-trap expression patterns. (**a**) Maximal image projections of *z*-dimension stacks in sagittal (left) and coronal (right) planes. Brightfield (top) and corresponding green fluorescence images of 4 d after fertilization (d.p.f.) larvae are shown. For each mRFP image, GBT identifier number, gene name and age of the larval fish are indicated. Scale bars, 200 µm. (**b**) mRFP expression pattern in GBT0040, an integration within the *hoxaa* cluster between *hoxa5a* and *hoxa4a*. Scale bars, 100 µm. The schematic demonstrates the annotated genes of this region of the *hoxaa* cluster. The red framed exons and red splicing lines show exons spliced to RP2. The green splice lines show the primary splice event from the 3′ exon trap and the dark and light blue splicing lines show alternative splicing identified by reverse transcription–PCR. The graph shows relative abundance of transcripts containing both the shared exon and indicated downstream cassette (*hoxa3a*, *hoxa4a* or *mRFP*) within the given genotypes (95% confidence interval, $n = 4$). Genotypes are wild type, and heterozygous (+/R^{40}) and homozygous (R^{40}/R^{40}) *mn0040Gt* mutants of the indicated alleles.

Annotating gene function with RP2

The GBT protein-trap system provides *in vivo* functional annotation beyond reporting the dynamic expression patterns of vertebrate genes and promoter or splicing variations at interrupted loci. We illustrate three examples of genes expressed in distinct domains of the zebrafish musculature: *tnnt2a* (GBT0031; **Fig. 3a**), *ryr1b* (GBT0348; **Fig. 3b**) and *myom3* (GBT0067; **Fig. 3c**).

GBT0031 (**Fig. 3a**) is a first-generation protein-trap insertion (R14; **Supplementary Fig. 3**) into troponinT2a *(tnnt2a)* with strong cardiac muscle–specific mRFP expression. Protein-trap disruption of *tnnt2a* results in recessive loss of heartbeat that phenocopies a previously documented *tnnt2a* mutation, silent heart (*sih*)[16,17]. To test the somatic reversibility of this protein-trap system, we injected *Cre* mRNA or splice acceptor masking morpholino into embryos from a GBT0031 incross. Whereas 27.8% of uninjected embryos (70 of 252 embryos) developed the silent heart phenotype, this phenotype was reversed in embryos injected with either masking morpholino (0 of 182 embryos) or Cre recombinase (0 of 176 embryos) (**Fig. 3a**). We also measured the knockdown of intact *tnnt2a* mRNA via quantitative PCR across the exons flanking the protein-trap insertion. Homozygous GBT0031 ($tnnt2a^{mn31gt-/mn31gt-}$) embryos expressed 6% of the intact mRNA as compared to wild-type siblings. Injection of either the gene-break-masking morpholino or *Cre* restored the level of intact mRNA to near heterozygotic ($tnnt2a^{+/mn31gt-}$) or wild-type ($tnnt2a^{+/+}$) amounts, respectively (**Fig. 3a**). Note that we made the GBT0031/*tnnt2a* locus from an early generation protein-trap vector, pGBT-R14 (R14) (**Supplementary Fig. 3**). Although these first successful protein-trap vectors including R14, pGBT-R15 (R15)[7] and pGBT-R16 (R16) often lead to stronger knockdown than earlier published gene trap vectors in zebrafish[11], the overall knockdown can be incomplete (**Table 1** and **Supplementary Fig. 3**). Addition of the strong transcriptional terminator or stop cassette and a second splice acceptor in RP2 results in a notably stronger knockdown effect (**Fig. 1**, **Table 1** and **Supplementary Fig. 3**).

An RP2 insertion into the ryanodine receptor 1b gene (GT0348/*ryr1b*) resulted in only 3% of its mRNA intact in homozygous ($ryr1b^{mn348gt-/mn348gt-}$) offspring. This is the weakest knockdown we have measured in an RP2 line to date (**Table 1**) but is a sharp improvement over the early protein-trap vectors R14, R15 and R16. GBT0348/*ryr1b* produces mRFP expression in the fast-twitch muscle of the zebrafish (**Fig. 3b**). Disruption of *ryr1b* expression in homozygous larvae resulted in a slow swimming phenotype that led to growth impairment and lethality by 14 d after fertilization (data not shown). The GBT0348 phenotype is similar to a previously reported allele of *ryr1b* that impacts swimming called relatively relaxed[18] (J. Dowling, personal communication).

The third muscle-specific gene insertion is in myomesin 3 (GBT0067/*myom3*; **Fig. 3c**), which encodes a structural component of the M-band of intermediate fibers of skeletal muscle[19]. *Myom3* is expressed in the slow or intermediate muscle fibers of the larval zebrafish. Homozygous fish can survive to fertility in a laboratory environment despite having less than 1% of intact *myom3* mRNA.

RP2 annotation of protein localization and trafficking

Protein-trap fusions can reveal protein subcellular localization; for example, the myom3-mRFP fusion is found in the M-band of the sarcomere (**Fig. 3c**). They can also be used to study subsets of proteins with common trafficking properties. The signal most

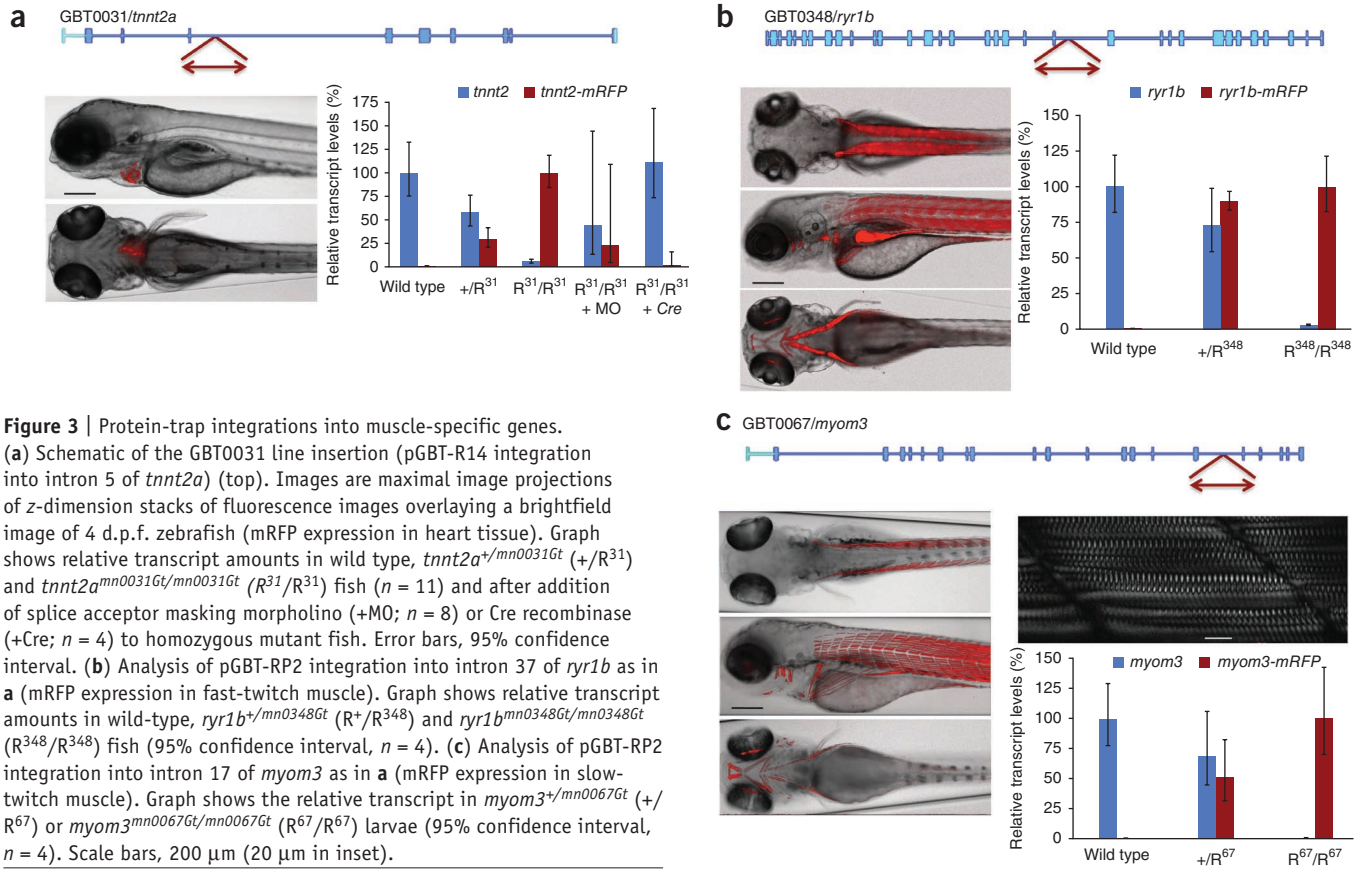

Figure 3 | Protein-trap integrations into muscle-specific genes. (**a**) Schematic of the GBT0031 line insertion (pGBT-R14 integration into intron 5 of *tnnt2a*) (top). Images are maximal image projections of *z*-dimension stacks of fluorescence images overlaying a brightfield image of 4 d.p.f. zebrafish (mRFP expression in heart tissue). Graph shows relative transcript amounts in wild type, $tnnt2a^{+/mn0031Gt}$ (+/R^{31}) and $tnnt2a^{mn0031Gt/mn0031Gt}$ (R^{31}/R^{31}) fish (n = 11) and after addition of splice acceptor masking morpholino (+MO; n = 8) or Cre recombinase (+Cre; n = 4) to homozygous mutant fish. Error bars, 95% confidence interval. (**b**) Analysis of pGBT-RP2 integration into intron 37 of *ryr1b* as in **a** (mRFP expression in fast-twitch muscle). Graph shows relative transcript amounts in wild-type, $ryr1b^{+/mn0348Gt}$ (R^{+}/R^{348}) and $ryr1b^{mn0348Gt/mn0348Gt}$ (R^{348}/R^{348}) fish (95% confidence interval, n = 4). (**c**) Analysis of pGBT-RP2 integration into intron 17 of *myom3* as in **a** (mRFP expression in slow-twitch muscle). Graph shows the relative transcript in $myom3^{+/mn0067Gt}$ (+/R^{67}) or $myom3^{mn0067Gt/mn0067Gt}$ (R^{67}/R^{67}) larvae (95% confidence interval, n = 4). Scale bars, 200 µm (20 µm in inset).

frequently observed in this collection of protein-trap lines was the N-terminal signal sequence, found in members of the secretome[20]. Understanding the function of the secretome is particularly crucial because of the extensive roles that cellular context and cell-cell signaling have in vertebrate biology and physiology. Key subsets of the secretome such as G protein–coupled receptors are critical drug targets.

About 25% of the 350 lines we generated have some mRFP accumulation in the kidney tubules, white blood cells or developing bone (**Fig. 4a**). Based on the presence of signal sequences at the N terminus of proteins identified from fish with these mRFP patterns, we hypothesized that the kidney and white blood cells were filtering or in other ways accumulating the mRFP fusion proteins. To test the ability of these cell types to remove fluorescent particles from the blood, we injected Alexa Fluor 488–conjugated dextran into the bloodstream of GBT0046 embryos 3 d after fertilization. Within 8 h the dextran particles had accumulated in both the kidney ducts and white blood cells where the mRFP fusion protein was also localized (**Fig. 4b**). The presumed trafficking of secreted mRFP-protein fusions resulted in diverse expression outcomes. In some lines, mRFP fusion proteins demonstrated some local retention indicating where the mRFP fusion protein originated[7]. Other lines showed no mRFP fusion protein locally, an effect that is readily documented using an antisense mRFP probe for *in situ* expression data (**Fig. 4**).

Strong mRFP accumulation in developing bone occurred in a few protein-trap lines and depended on the fusion protein (**Fig. 4a**), suggesting an additional protein-trafficking mechanism. For instance, we observed strong mRFP accumulation in developing bone upon RP2 integration into cd99 antigen like-2 (GBT0043/*cd99l2*); the cd99l2-mRFP fusion protein accumulated in or near developing bone (**Fig. 4a**), even though *cd99l2* mRNA was not expressed at or near this tissue (data not shown). We hypothesized that the amino acids present in the cd99l2 protein fusion of GBT0043 directed secretion and subsequent accumulation of mRFP in the bone matrix. We tested this by injecting DNA that would produce mRFP with and without the cd99l2 fusion sequence in a ubiquitous, albeit mosaic, manner in the zebrafish. When we expressed mRFP alone from a ubiquitous promoter, we saw generalized, mosaic expression of mRFP. But when we fused the cd99l2 sequences to mRFP, fluorescence localized to developing bone (**Fig. 4c**).

Protein-trap integrations into secretome genes are efficient mutagens as illustrated by integration into the Fraser syndrome 1 gene (GBT0156/*fras1*) (**Fig. 4d**). The RP2 insertion is in intron 15 of *fras1*, creating a fusion protein with the extracellular domain of *fras1*. The secreted *fras1*-mRFP fusion protein in GBT0156 accumulated in both the kidney and white blood cells. However, the mRFP-fusion mRNA was located in the brain, lens, muscle and developing fins and skin, similar to endogenous expression patterns for *fras1* (refs. 21,22). GBT0156 fish homozygous for the insertion ($fras1^{mn0156gt-/mn0156gt-}$) expressed only 0.01% of intact *fras1* transcript and displayed a phenotype mimicking *pinfin*, a previously described mutation in *fras1*[21,23]. Microinjecting *Cre* mRNA reverted the *pinfin* phenotype in GBT0156 embryos. If additional prioritization of secreted lines based on mRNA expression is desired, *in situ* hybridization using a single antisense mRFP probe documents the expression pattern of tagged lines.

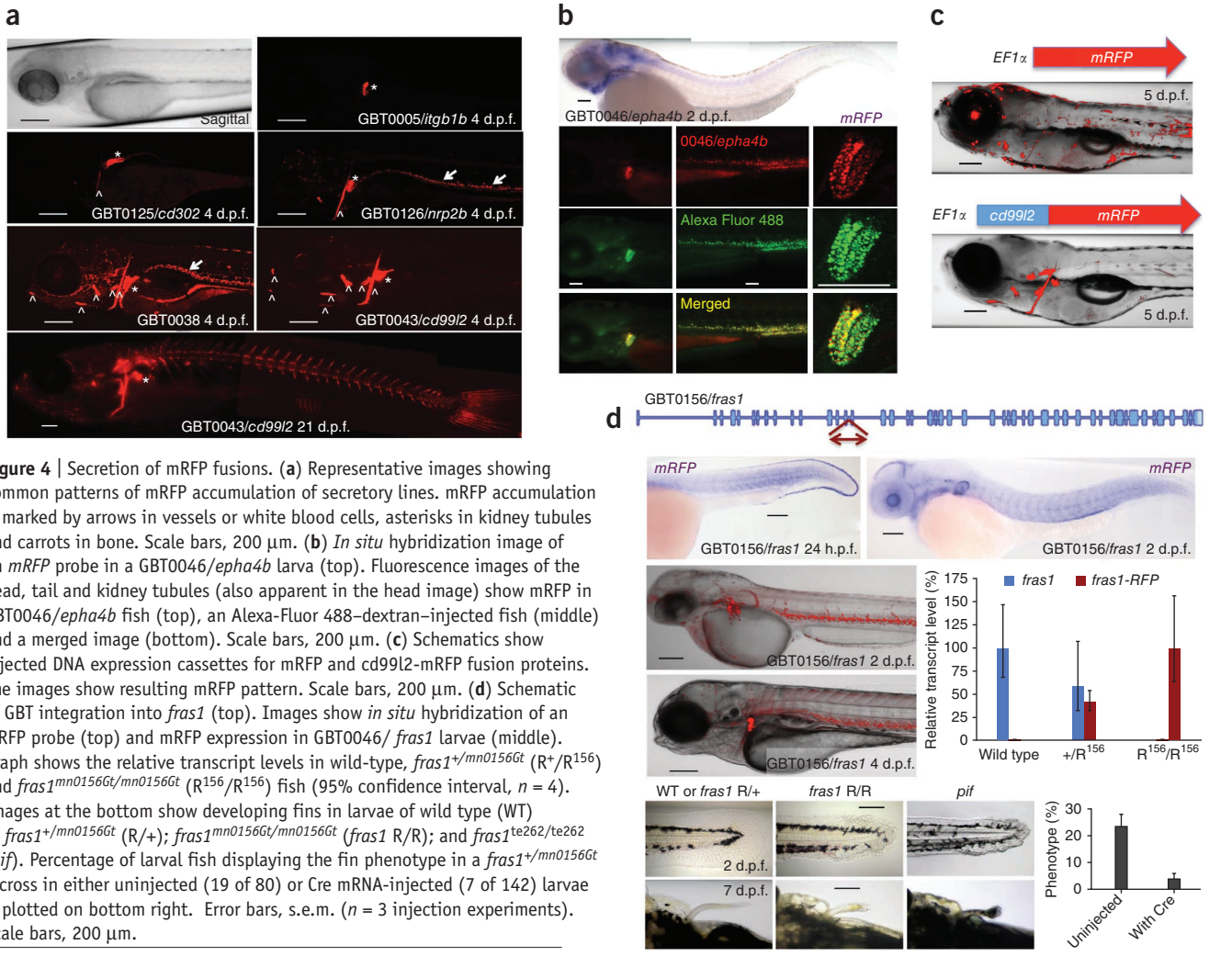

Figure 4 | Secretion of mRFP fusions. (**a**) Representative images showing common patterns of mRFP accumulation of secretory lines. mRFP accumulation is marked by arrows in vessels or white blood cells, asterisks in kidney tubules and carrots in bone. Scale bars, 200 µm. (**b**) *In situ* hybridization image of an *mRFP* probe in a GBT0046/*epha4b* larva (top). Fluorescence images of the head, tail and kidney tubules (also apparent in the head image) show mRFP in GBT0046/*epha4b* fish (top), an Alexa-Fluor 488–dextran–injected fish (middle) and a merged image (bottom). Scale bars, 200 µm. (**c**) Schematics show injected DNA expression cassettes for mRFP and cd99l2-mRFP fusion proteins. The images show resulting mRFP pattern. Scale bars, 200 µm. (**d**) Schematic of GBT integration into *fras1* (top). Images show *in situ* hybridization of an mRFP probe (top) and mRFP expression in GBT0046/ *fras1* larvae (middle). Graph shows the relative transcript levels in wild-type, *fras1*$^{+/mn0156Gt}$ (R^{+}/R^{156}) and *fras1*$^{mn0156Gt/mn0156Gt}$ (R^{156}/R^{156}) fish (95% confidence interval, $n = 4$). Images at the bottom show developing fins in larvae of wild type (WT) or *fras1*$^{+/mn0156Gt}$ (R/+); *fras1*$^{mn0156Gt/mn0156Gt}$ (*fras1* R/R); and *fras1*$^{te262/te262}$ (*pif*). Percentage of larval fish displaying the fin phenotype in a *fras1*$^{+/mn0156Gt}$ incross in either uninjected (19 of 80) or Cre mRNA-injected (7 of 142) larvae is plotted on bottom right. Error bars, s.e.m. ($n = 3$ injection experiments). Scale bars, 200 µm.

DISCUSSION

The approach used here is based in part on the extensive gene-trapping work in mouse embryonic stem cells, including vectors that produce protein fusions, one form of protein traps[24]. It also complements the rich history of *in vivo* protein trapping in *Drosophila melanogaster*, where the protein traps are not designed to terminate the interrupted protein and therefore are less mutagenic[25–27]. The zebrafish model is well-suited for high-throughput genome mutagenesis and real-time expression analysis using protein-trap technology. Moreover, the RP2 transposon system is not restricted to use in zebrafish and can be applied to other model organisms.

Near-random integration of transposons into the vertebrate genome permits unbiased identification of genes regardless of whether they are completely novel, predicted or well characterized in other species. In addition, the visible expression pattern of the trapped locus can be used to intentionally bias a genetic screen by preselecting mutations in genes expressed in a tissue of interest. Furthermore, the dominant reporter in GBT mutants opens the door to an array of genetic screens difficult or impossible to perform with chemical mutagenesis, including behavioral genetics. The ability to sort a population based on the presence of a single gene mutation allows quantitative assessment of the role of a single gene in modifying complex, multifactorial phenotypes. Localized reversion of the mutation by the use of tissue-specific Cre recombinase will permit mapping of tissues or neural networks that are required for proper function.

The combination of conditional mutants in a vertebrate with *in vivo* protein-trap technology enables a direct link between DNA sequence, expression and function for each genetic locus, a concept we term a gene codex. When combined with high-resolution analyses of sequence variation, the zebrafish is a key model for assembling a full codex for the genetic complement of the vertebrate genome and can provide new insights into genomic complexity that have been difficult to directly study and functionally annotate.

The RP2 mutagenesis system presented here is a first step to creating a gene codex. To complete it, multiple alleles for each protein-coding gene will be needed to understand expression, trafficking and function depending on insert location. For example, integration at N-terminal locations is more likely to produce functional null alleles because large portions of the protein are truncated. In contrast, integration into more C-terminal locations is more likely to produce proteins with at least some function but that are informative about subcellular protein localization and trafficking. Comprehensive mutagenesis of a vertebrate genome with insertional mutagens, such as RP2, could require millions of insertional events[28]. Therefore, to maximize genome coverage,

it is recommended to use complementary vector systems to take advantage of differences in integration preferences[28,29]. A first step will be to produce all three reading frames of the RP2 vector. Subsequent complementary transposon systems and yet-to-be established reverse genetic approaches will increase genome coverage to achieve the goal of a comprehensive codex of the vertebrate genome.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.

Accession codes. GenBank: HQ335167 (pGBT-R14), HQ335168 (pGBT-R15), HQ335169 (pGBT-R16), HQ335170 (pGBT-RP2.1), HQ335171 (pT3TS-Tol2), HQ335172 (pT3TS-Cre) and HQ335166 (pGBT-PX).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

The US National Institute on Drug Abuse (DA14546), US National Institute of General Medical Sciences (GM63904), National Institute of Diabetes and Digestive and Kidney Diseases (F30DK083219 and P30DK084567) and the Mayo Foundation provided funding for this research. S.S. and V.S. acknowledge funding support from the Council of Scientific and Industrial Research (grant FAC002), India. We thank members of the Center for Genome Engineering at the University of Minnesota for providing collaborative discussion and resources, the staff of the Zebrafish Core Facility at the Mayo Clinic for providing zebrafish care, D. Argue for programming the zfishbook website, E. Klee for bioinformatic analyses of the zebrafish transcriptome used in theoretical design and testing of the RP2 system, A. Person for assistance in injection of fluorescently labeled dextrans, and InSciEd Out participants and Summer Undergraduate Research Fellow, N. Boczek, for imaging RFP-expressing fish.

AUTHOR CONTRIBUTIONS

K.J.C., D.B., S.S. and S.C.E. designed the experiments. D.B. built *R14, R15, R16, RP2* and pEF1a-cd99l2-RFP vectors. S.S. built pX. J.N. made the pEF1a-RFP vector. D.B. piloted the production of R14, R15 and R16 lines. K.J.C. trained and managed the team that produced >300 RP2 lines and molecularly characterized lines other than GBT0031, GBT0039 and GBT0043 (D.B.). M.D.U., T.M.G., A.L.N., K.J.S. and K.J.C. identified and propagated mRFP-expressing lines. A.M.P., M.D.U. and K.J.C. photographed mRFP expression patterns. K.J.S. and K.J.C. molecularly characterized lines by 5′ rapid amplification of cDNA ends, inverse PCR and quantitative PCR. T.M.G., K.J.C. and D.B. characterized the reversion of GBT0031 by *Cre* mRNA and morpholino injection. D.B. injected fluorescently conjugated dextran into GBT0046 and imaged its uptake. J.N. imaged GBT0043 fish and together with K.J.C. injected fish with EF1a-RFP and EF1a-cd99l2-RFP vectors. K.J.C. imaged the injected fish. S.E.W. analyzed the GBT0156/frasI phenotype and reversion. V.M.B. conducted *in situ* hybridizations. Y.D. identified GBT0348 in a screen performed in X.X.'s laboratory and contributed quantitative PCR data for three more RP2 lines. H.-M.P. tested the GBT protocol in M.H.'s lab and produced several independent lines, including the initial assessment and imaging of GBT0040. A.P., V.S. and S.S. produced and analyzed 3′ exon trap data for **Supplementary Figure 2**. K.J.C. and S.C.E. were primary authors of the text.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

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 (2007).
3. Sauka-Spengler, T. & Bronner-Fraser, M. Evolution of the neural crest viewed from a gene regulatory perspective. *Genesis* **46**, 673–682 (2008).
4. Balciunas, D. *et al.* Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet.* **2**, e169 (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 (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 (2006).
7. Petzold, A.M. *et al.* Nicotine response genetics in the zebrafish. *Proc. Natl. Acad. Sci. USA* **106**, 18662–18667 (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).
9. Parinov, S., Kondrichin, I., Korzh, V. & Emelyanov, A. Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. *Dev. Dyn.* **231**, 449–459 (2004).
10. Balciunas, D. *et al.* Enhancer trapping in zebrafish using the Sleeping Beauty transposon. *BMC Genomics* **5**, 62 (2004).
11. Kawakami, K. *et al.* A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell* **7**, 133–144 (2004).
12. Munoz-Sanjuan, I., Simandl, B.K., Fallon, J.F. & Nathans, J. Expression of chicken fibroblast growth factor homologous factor (FHF)-1 and of differentially spliced isoforms of FHF-2 during development and involvement of FHF-2 in chicken limb development. *Development* **126**, 409–421 (1999).
13. Holland, P.W. Beyond the Hox: how widespread is homeobox gene clustering? *J. Anat.* **199**, 13–23 (2001).
14. Coulombe, Y. *et al.* Multiple promoters and alternative splicing: Hoxa5 transcriptional complexity in the mouse embryo. *PLoS ONE* **5**, e10600 (2010).
15. Hadrys, T., Prince, V., Hunter, M., Baker, R. & Rinkwitz, S. Comparative genomic analysis of vertebrate Hox3 and Hox4 genes. *J. Exp. Zoolog. B Mol. Dev. Evol.* **302**, 147–164 (2004).
16. Chen, J.N. *et al.* Mutations affecting the cardiovascular system and other internal organs in zebrafish. *Development* **123**, 293–302 (1996).
17. Sehnert, A.J. *et al.* Cardiac troponin T is essential in sarcomere assembly and cardiac contractility. *Nat. Genet.* **31**, 106–110 (2002).
18. Hirata, H. *et al.* Zebrafish relatively relaxed mutants have a ryanodine receptor defect, show slow swimming and provide a model of multi-minicore disease. *Development* **134**, 2771–2781 (2007).
19. Schoenauer, R. *et al.* Myomesin 3, a novel structural component of the M-band in striated muscle. *J. Mol. Biol.* **376**, 338–351 (2008).
20. Klee, E.W. The zebrafish secretome. *Zebrafish* **5**, 131–138 (2008).
21. Carney, T.J. *et al.* Genetic analysis of fin development in zebrafish identifies furin and hemicentin1 as potential novel fraser syndrome disease genes. *PLoS Genet.* **6**, e1000907 (2010).
22. Gautier, P., Naranjo-Golborne, C., Taylor, M.S., Jackson, I.J. & Smyth, I. Expression of the fras1/frem gene family during zebrafish development and fin morphogenesis. *Dev. Dyn.* **237**, 3295–3304 (2008).
23. van Eeden, F.J. *et al.* Genetic analysis of fin formation in the zebrafish, Danio rerio. *Development* **123**, 255–262 (1996).
24. Friedel, R.H. & Soriano, P. in *Methods in Enzymology* Vol. 477 (eds., Wassarman, P.M. & Soriano, P.M.) 243–269 (Academic Press, 2010).
25. Aleksic, J., Lazic, R., Müller, I., Russell, S.R. & Adryan, B. Biases in *Drosophila melanogaster* protein trap screens. *BMC Genomics* **10**, 249 (2009).
26. Buszczak, M. *et al.* The carnegie protein trap library: a versatile tool for *Drosophila* developmental studies. *Genetics* **175**, 1505–1531 (2007).
27. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**, 15050–15055 (2001).
28. Nord, A.S. *et al.* Modeling insertional mutagenesis using gene length and expression in murine embryonic stem cells. *PLoS ONE* **2**, e617 (2007).
29. Skarnes, W.C. *et al.* A public gene trap resource for mouse functional genomics. *Nat. Genet.* **36**, 543–544 (2004).

ONLINE METHODS

Zebrafish. All zebrafish work was conducted under Institutional Animal Care and Use Committee approved protocols. All lines are freely available now through http://zfishbook.org/, and, once the collection is synchronized, will be accessible from the Zebrafish International Resource Center.

Vector construction. Plasmids used in this study are freely available on request. The following annotated plasmid sequences have been submitted to GenBank: pGBT-R14 (accession code 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 (ref. 30). The ocean pout antifreeze polyadenylation sequence, transcriptional terminator and putative border element used in pGBT-PX and pGBT-RP2 were cloned from pFRM2 (ref. 31).

pEF1a-RFP was made by ligating a 685 bp BamHI–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 R. Tsien's laboratory 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, encoding the cd99l2 N-terminal fusion to mRFP, was obtained from cDNA of GBT0043 fish using cd99l2-F1 and mRFP-R1 primers. The second PCR product 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 (also known as 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 μl^{-1} each. The injected embryos were raised at ~29 °C. For RP2 injections, the injected embryos were sorted using the GFP fluorescence at 3–4 d after 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 (F_0 generation). The number of class III larvae ranged from 5–50% of injected embryos with an average of about 20%.

F_1 embryos were obtained by crossing the F_0 generation to nontransgenic brood stock. The embryos were scored for mRFP expression. In a recent assessment, 32% of class III fish (173 fish) produced some F1 embryos with mRFP expression compared to 20% (127 fish) and 7% (61 fish) of class II and class I fish, respectively. mRFP-expressing embryos were sorted by pattern if possible, assigned a GBT identifier and raised to adulthood. Imaging and molecular work were done on F_2 or subsequent-generation embryos.

Standard imaging of GBT lines. The mRFP expression pattern of each GBT line was recorded at both 2 and 4 d after fertilization if mRFP was expressed. Coronal-, sagittal- and ventral- oriented *z*-stacks were obtained at 50× magnification as described previously[33].

Identification of tagged genes. To determine the identity of the gene with an activated protein trap, 5′ rapid amplification of cDNA ends (RACE) was performed as previously described[34], with minor modifications to primer sequences. Briefly, total RNA was isolated from 20 mRFP-expressing embryos. cDNA was produced from 250 ng of total RNA using a gene-specific primer 5R-mRFP-P0 for the reverse transcriptase reaction. PCR was 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 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 100 µl reaction. The ligation reactions were diluted tenfold 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. After an initial denaturation for 2 min at 95°, the primary and secondary PCRs were cycled 30 times with 30 s of denaturation, 30 s of annealing at 55 °C and 6 min of extension. The primary PCRs were diluted 50-fold before preparing the nested PCRs. 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 noncarrier siblings.

Identification of orthologs and fusions. Orthologs listed in **Table 1** were determined using the National Center for Biotechnology Information Homologene database. Orthologs for GBT0007, GBT0021, and GBT0067 were identifed as a best match after a BlastX search of the human genome. Fusions mentioned in **Table 1** include (i) an insert into the 5′ untranslated region of *si:ch211-51g4.4* that uses an upstream AUG in the 5′ untranslated region to create this fusion; (ii) an insert in the *hoxa* cluster that

uses a shared exon with unique protein fusion sequences; and (iii) an insert into the first annotated *dido1* exon results in two transcripts. The first transcript uses only the two annotated upstream exons that are noncoding. 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 after 5′ RACE, primers were designed in the flanking exons of the 'tagged' gene priming toward 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 8 kb, primers alternating in orientation were designed every 3 kb 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 reverse transcription–PCR. Individual embryos of a hetereozygous incross breeding were lysed using Trizol reagent (Invitrogen). Total RNA was purified as indicated with the addition of 1 µl of glycoblue (Ambion) as a carrier. Genomic DNA was purified after back-extraction of the organic phase and interphase using 4 M guanidine thiocyanate, 50 mM sodium citrate and 1 M Tris (pH 8.0). The genomic DNA in the new aqueous phase was moved to a new tube, 1 µl 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 10 µl of TE. The genomic DNA was used with genotyping primers to determine which embryos were wild type, heterozygous or homozygous 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 **Supplementary Table 2** for sequences) were used to test for 'wild-type' 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 sample reactions were referenced to the ribosomal protein S6 kinase b, polypeptide 1 transcript using RT-RPS6kb1-F1 and RT-RPS6Kb1-R1. In cases where genotyping of embryos was possible, owing 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 with PmeI. To the purified DNA template, we added 10× Dig RNA labeling mix (Roche) and T7 RNA polymerase (Roche) for 2 h at 37 °C to make the probe. After 2 h, the plasmid was digested using RNase-free DNaseI (Promega) and the probe purified using RNA easy MinElute cleanup kit (Qiagen).

Morpholino and Cre reversion. Embryos from GBT crosses were injected with a mixture of two morpholinos, GBM1 and GBM2 (**Supplementary Table 1**) that target the exogenous splice acceptor in pRP2 derived from carp beta-actin intron 1. Injection of 3 nl of a GBM1 and GBM2 morpholino mixture at concentrations of 100 µM was used to revert the GBT0031 phenotype. *Cre* mRNA was produced using mMessage mMachine (Ambion) from pT3TS-Cre plasmid linearized by SacI digestion. About 25 pg 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 d after fertilization 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]. The 3′ RACE-PCR products were treated with SpeI and SphI restriction enzymes in a final reaction volume of 100 µl and were incubated at 37 °C for 8 h. The SpeI and SphI–digested 3′ RACE-PCR products were purified and ranged from 70 bp to 4 kb. 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. We transformed 2 µl of this reaction mixture 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 medium per well. The inoculated plates were incubated at 37 °C with 250 r.p.m. rotation for 20 h at an incline of 45 °C to increase the surface area. After incubation the plates were centrifuged at 3,000*g* for 5 min 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 ~100 bp were considered as a 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-h.p.f. zebrafish embryos. We incubated 50 µg of freshly isolated RNA with 5 µl of 10 µM adaptor primer in a final volume of 35 µl. This mixture was incubated at 65 °C for 10 min, 25 °C for 5 min, 42 °C for 3 min 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 3,500 millicurie mM^{-1} α-labeled dCTP, 2 µl of M-MLV reverse transcriptase and 5 µl of 10× buffer (Ambion). This mixture is incubated at 42 °C for 1 h and then chilled at 4 °C for 5 min. To this reaction, 5 µl of each 0.5 M EDTA and 1 N NaOH was added and then the mixture was heated to 65 °C for 30 min. This mixture was chilled in ice and then 6.5 µl 1 M Tris-Cl (pH 7.5) is added. Before this, clones containing inserts for which identities were found only in the Ensembl zebrafish genomic DNA database with no representation found

doi:10.1038/nmeth.1606

in the current EST databases were spotted on 0.45 μm nitrocellulose blotting membrane (MDI). This membrane is then UV-light cross-linked using UV cross linker and then hybridized for 1 h at 37 °C using prehybridization buffer containing DIG Easy Hyb buffer (Roche). The radioactive probe prepared above was heated at 95 °C for 2 min and added to the prehybridization buffer directly. The membrane was incubated overnight at 37 °C. After the hybridization, the membrane was washed twice: initially with 1× SSC and 0.1% SDS for 15 min at 37 °C and then with 0.5× SSC and 0.5% SDS for 15 min at 37 °C. As soon as washes were done, the membrane was wrapped with Saran wrap and exposed to a Phosphoimager screen for ~24 h before imaging.

30. Cormack, B.P., Valdivia, R.H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
31. Gibbs, P. & Schmale, M. GFP as a genetic marker scorable throughout the life cycle of transgenic zebra fish. *Mar. Biotechnol.* **2**, 107–125 (2000).
32. Campbell, R.E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. USA* **99**, 7877–7882 (2002).
33. Petzold, A.M. *et al.* SCORE imaging: specimen in a corrected optical rotational enclosure. *Zebrafish* **7**, 149–154 (2010).
34. Clark, K.J., Geurts, A.M., Bell, J.B. & Hackett, P.B. Transposon vectors for gene-trap insertional mutagenesis in vertebrates. *Genesis* **39**, 225–233 (2004).
35. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* **3**, 59–69 (2008).
36. Bill, B.R., Petzold, A.M., Clark, K.J., Schimmenti, L.A. & Ekker, S.C. A primer for morpholino use in zebrafish. *Zebrafish* **6**, 69–77 (2009).