RESOURCE/METHODOLOGY

A versatile gene trap to visualize and interrogate the function of the vertebrate proteome

Le A. Trinh,[1,5] Tatiana Hochgreb,[1] Matthew Graham,[2] David Wu,[1] Frederique Ruf-Zamojski,[1,3] Chathurani S. Jayasena,[1] Ankur Saxena,[1] Rasheeda Hawk,[1] Aidyl Gonzalez-Serricchio,[1] Alana Dixson,[1] Elly Chow,[1] Constanza Gonzales,[1] Ho-Yin Leung,[1] Ilana Solomon,[1] Marianne Bronner-Fraser,[1] Sean G. Megason,[1,4] and Scott E. Fraser[1]

[1]Beckman Institute, Division of Biology, and [2]Center for Advanced Computing Research, California Institute of Technology, Pasadena, California 91125, USA

We report a multifunctional gene-trapping approach, which generates full-length Citrine fusions with endogenous proteins and conditional mutants from a single integration event of the FlipTrap vector. We identified 170 FlipTrap zebrafish lines with diverse tissue-specific expression patterns and distinct subcellular localizations of fusion proteins generated by the integration of an internal citrine exon. Cre-mediated conditional mutagenesis is enabled by heterotypic lox sites that delete Citrine and "flip" in its place mCherry with a polyadenylation signal, resulting in a truncated fusion protein. Inducing recombination with Cerulean-Cre results in fusion proteins that often mislocalize, exhibit mutant phenotypes, and dramatically knock down wild-type transcript levels. FRT sites in the vector enable targeted genetic manipulation of the trapped loci in the presence of Flp recombinase. Thus, the FlipTrap captures the functional proteome, enabling the visualization of full-length fluorescent fusion proteins and interrogation of function by conditional mutagenesis and targeted genetic manipulation.

[*Keywords*: endogenous proteins; proteome; conditional mutagenesis; gene trap; genetic manipulation; gene knockdown]

Supplemental material is available for this article.

Received July 7, 2011; revised version accepted September 16, 2011.

A key to understanding developmental processes is to identify the network of genes expressed in the developing embryo and study their function. Strategies to assess the expression of developmentally regulated genes range from mRNA expression screens to insertion of trapping vectors that monitor transcriptionally active regions of the genome by the random insertion of a report gene for lacZ or a fluorescent protein (Gossler et al. 1989; Friedrich and Soriano 1991). To be most useful, expression studies must move beyond the mRNA level to provide information at the level of protein expression, as it is the protein products that in general carry out the function of genes in cells.

Present addresses: [3]Department of Basic Neurosciences, University of Geneva Medical Center, CH-1211 Geneva 4, Switzerland; [4]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.
Corresponding author.
[5]E-mail ltrinh@caltech.edu.
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.174037.111.

Ample evidence of the power of protein localization studies is provided by the new insights they have presented into gene function (Kerrebrock et al. 1995; Lippincott-Schwartz et al. 1998). Such studies in vertebrates typically focus on single-gene products, using antibodies or fluorescent fusion proteins to assess protein localization. While informative, single-gene approaches can be costly and time-consuming. Large-scale efforts assessing protein localization in the yeast *Saccharomyces cerevisiae* have taken advantage of its short generation time and completely sequenced genome to systematically tag the 3′ end of known genes with a fluorescent protein or a purification tag by homologous recombination (Gavin et al. 2002; Ghaemmaghami et al. 2003; Huh et al. 2003). Translating such systematic approaches to multicellular organisms, in particular to vertebrates, requires the development of new experimental strategies.

Gene trapping offers an alternative approach to the systematic tagging of known genes, relying instead on the

random insertion of a reporter throughout the genome to create fusion proteins with a marker such as lacZ or GFP. Traditional gene-trapping vectors, with a splice acceptor site immediately upstream of a promoterless reporter construct, provide expression data by creating a fusion transcript between a splice donor of the endogenous gene and the splice acceptor of the reporter gene (Gossler et al. 1989; Skarnes et al. 1992; International Gene Trap Consortium 2004). The reporter sequence reveals the endogenous gene expression during development, while the splice fusion of the 5′ end of the truncated endogenous gene creates a defined mutation. Gene trapping has become a powerful approach for studying developmental-regulated genes because it can provide information on both the spatiotemporal expression and the function of the gene. However, the main advantage of traditional gene trapping is also a disadvantage, as capturing the endogenous gene results in the production of a truncated protein, which can alter binding partners and subcellular localization, precluding the assessment of the functional wild-type protein. Additionally, truncated proteins might misfold or otherwise be targeted for degradation, which would make them less informative for expression analysis. In *Drosophila*, a variation of the gene trap vector (sometimes called a protein trap) employs an internal exon incorporating both the splice acceptor and splice donor sequences flanking GFP to generate full-length GFP-tagged proteins. These fusions can be used to assess wild-type protein localization and dynamics (Morin et al. 2001; Clyne et al. 2003; Buszczak et al. 2007) but are unable to create mutant alleles.

Dissecting complex events such as embryogenesis requires a systematic means to combine expression studies with the wealth of information provided by forward genetic screens (Driever et al. 1994; Mullins et al. 1994; Anderson 2000; Hrabé de Angelis et al. 2000). Conditional mutagenesis approaches, creating a large family of mutations that can be deployed in a stage- and tissue-specific fashion, are an essential tool. Building on the availability of homologous recombination in a few model organisms (Huang et al. 1993; Rijli et al. 1993; Rosahl et al. 1993), Cre, Flp, and other site-specific recombinases have been used to create gene deletions, insertions, and inversions in cell culture systems, mice, and *Drosophila* (Senecoff et al. 1985; Golic and Lindquist 1989; Lakso et al. 1992; Orban et al. 1992; Sauer and Henderson 1998; Schnütgen et al. 2003; Branda and Dymecki 2004). Cre conditional alleles allow researchers to circumvent experimental problems caused by genetic redundancy and pleiotropy by mutating a gene in only a specific tissue or at a specific time. Unfortunately, Cre conditional alleles have only been made in a few select organisms in which homologous recombination in embryonic stem cells is possible, such as mice, and each allele represents a significant effort. For model systems that lack the technology for homologous recombination, such as zebrafish, site-directed recombinase techniques have been adapted to control the expression of reporter genes (Hans et al. 2009; Collins et al. 2010). These reporter systems target recombination of transgenes that express fluorescent proteins or exogenous genes in over-expression systems without the ability to target the endogenous loci (Le et al. 2007).

Here, we report an approach that combines the strengths of gene traps, insertional mutagenesis, Cre-mediated mutagenesis, and Flp-mediated targeted genetic manipulation in a vertebrate. The "FlipTrap" generates full-length fluorescent fusion proteins by encoding an internal exon for the fluorescent protein Citrine, integrated by transposable elements (TEs). The vector is engineered with *lox* sites that permit Cre recombinase to remove the citrine and splice donor sequence, resulting in a truncated protein after exposure to Cre recombinase. FRT sites in the vector allow replacement of the FlipTrap cassette with any DNA sequence once an integration event is established. We demonstrate the feasibility of using this approach to visualize the fully functional proteome in the embryo at subcellular resolution, convert the functional proteome into mutant forms, and perform targeted genetic manipulation.

Results

Visualizing the functional proteome by fluorescent tagging of full-length proteins

To generate full-length functional fusion proteins with endogenous genes, we created the FlipTrap, an integration vector containing an internal exon encoding citrine, flanked with TEs to allow integration in the presence of transposase. The Tol2 and maize Ac/Ds transposable systems were chosen to integrate the vector into the genome because of their reported high insertional rate, large cargo capacity, ease of use, and functionality in a wide spectrum of heterologous hosts (Weil and Kunze 2000; Kawakami et al. 2004; Balciunas et al. 2006; Emelyanov et al. 2006). The internal exon is devoid of initiation and stop codons and is flanked by splice acceptor and donor sequences from the zebrafish *ras-associated domain family 8* (*rassf8*) gene, which contains a naturally occurring large internal exon and is constitutively spliced (Fig. 1A). When integrated into an intron of an actively expressed gene, the splice acceptor and donor sequences are recognized by the endogenous splicing machinery, creating a fusion mRNA of citrine flanked by the 5′ and 3′ exons of the endogenous gene (Fig. 1B). Citrine fusion protein expression is detected when the citrine exon is spliced in-frame with the trapped gene. The initial FlipTrap vector has been created with a frame 0 citrine exon. Linker peptides encoded by the splice acceptor and donor sequences provide a flexible linker between the trapped protein domains and Citrine. Finally, the small, neatly folded nature of GFP (Citrine is a 5-amino-acid variant of GFP) (Heikal et al. 2000) and the close proximity of its N and C termini make it a good fusion partner (Ormö et al. 1996).

To visualize endogenous protein expression in the developing embryo using the FlipTrap approach, we performed an F_1 screen in zebrafish. The FlipTrap vector was coinjected with tranposase mRNA into fertilized zebrafish embryos, and the progeny of the F_0 adults were screened for positive Citrine expression (Supplemental Fig. S1). We

tested two variants of the FlipTrap vector: one with Tol2, and a second with Ds elements flanking the citrine and mCherry cassettes. From the injections with the Tol2-containing vector, we identified 122 positive F_0 carriers from 840 F_0 individuals. Of these, 22 individuals generated progeny with multiple expression patterns; these segregated independently in F_1 embryos, suggesting multiple in-frame integration events per F_0 individual. Subsequent cloning of the trapped genes and their integration sites confirmed that the expression patterns with independent segregation in F_1 embryos corresponded to independent insertion events. In total, 154 independent insertions were identified using the Tol2-containing FlipTrap vector, resulting in a trapping rate of 18.3%. We observed a trapping rate of 6.7% with the FlipTrap vector containing the Ds TE from screening 233 injected F_0 individuals (Table 1). A total of 170 FlipTrap lines were identified with the two variants of the FlipTrap vector.

FlipTrap lines exhibit a wide array of distinct expression patterns in the embryo, reflecting both the spatiotemporal differences in the expression of the genes and the subcellular localization of the predicted gene function (Fig. 1C–K). The expression patterns were categorized into three classes (Fig. 2A): tissue-specific (37%), ubiquitous with heightened expression in distinct tissues (14%), and uniform ubiquitous expression (49%). Over a third of the trapped genes exhibit tissue-specific expression at the protein level (Fig. 2A). Figure 1, C–E, shows three FlipTrap lines in which Citrine fusion proteins are expressed exclusively in specific tissues, with distinct subcellular localizations that are reflective of the gene products' cellular function. The trap of a predicted guanylate kinase gene, *Gt(magk-citrine)*^{ct30c}, is expressed only in the cytoplasm of the lateral nerve bundles in the spinal cord starting at 3 d post-fertilization (dpf) (Fig. 1C); two components of the skeletal muscle machinery, *myosin heavy chain 7b β* (*mhc7bb*) [*Gt(mhc7bb-citrine)*^{ct42a}] (Fig. 1D) and dystrophin (*dmd*) [*Gt(dmd-citrine)*^{ct90a}] (Fig. 1E), are expressed as expected in muscle fibers from the start of myogenesis (1 dpf) throughout development. The myosin component Mhc7bb-citrine is expressed in the dorsal muscle fibers in a striated pattern indicative of sarcomeres, and Dmd-citrine localizes to the junction of the sarcolemma as previously reported from antibody staining and as predicted for its role in maintaining muscle integrity (Berger et al. 2010; data not shown).

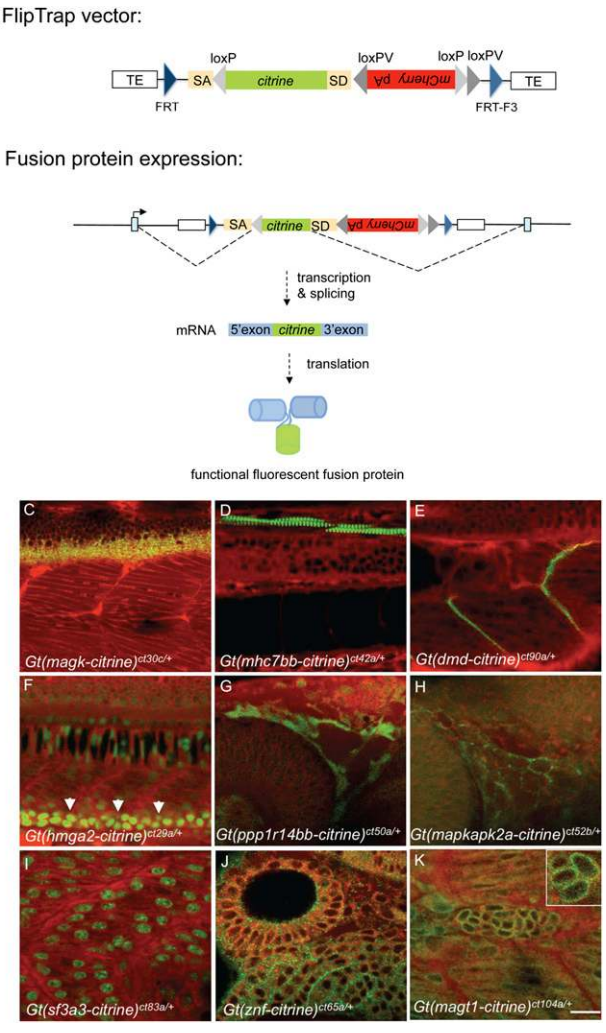

Figure 1. Visualizing the proteome by fluorescent tagging of endogenous full-length proteins. (*A*) The FlipTrap vector consists of a citrine coding sequence (green) flanked by intronic sequences that contain a splice acceptor (SA, beige) and donor (SD, beige). In the reverse orientation are mCherry (red) and polyA (red) sequences. Heterotypic *lox* sites are shown for *loxP* (light gray) and *loxPV* (dark gray). FRT sites (blue) are positioned internal to the TEs (outlined boxes). (*B*) Insertion of the FlipTrap by transposition into the intron of an actively expressed gene leads to splicing of the citrine exon, allowing the formation of an endogenously expressed full-length fluorescent fusion protein when the insertion is in-frame with the trapped gene. (*C–K*) Confocal images of Citrine fusion protein expression (green) in living FlipTrap embryos counterstained with the vital stain bodipy TR methyl ester (red). (*C–E*) Examples of tissue-specific expression. (*C*) *Gt(magk-citrine)*^{ct30c} expression in the nerve bundle of the spinal cord at 82 hpf. (*D*) *Gt(mhc7bb-citrine)*^{ct42a} is expressed in dorsal muscle fibers in a banding pattern of sacromeres at 78 hpf. (*E*) *Gt(dmd-citrine)*^{ct90a} is expressed and localized to the boundary of the myotome at 32 hpf. (*F–H*) Examples of ubiquitously expressed FlipTraps that exhibit heightened expression in specific tissues. (*F*) *Gt(hmga2-citrine)*^{ct29a} is expressed ubiquitously with heightened levels of expression in the pronephric ducts (arrowhead) at 23 hpf. (*G*) *Gt(ppp1r14bb-citrine)*^{ct50a} exhibits heightened expression in the head vasculature at 26 hpf. (*H*) *Gt(mapkapk2a-citrine)*^{ct52b} is ubiquitously expressed but only exhibits localized expression within the head vasculature at 26 hpf. (*I–K*) Examples of ubiquitously expressed FlipTraps with distinct subcellular localization. (*I*) *Gt(sf3a3-citrine)*^{ct83a} is localized to the nucleus of all cells. (*J*) *Gt(znf-citrine)*^{ct65a} is localized to the cell membrane as puncta. (*K*) *Gt(magt1-citrine)*^{ct104a} is expressed in all cells but localizes to the nuclear envelope of the lateral line cells. The *inset* in the *top right* corner of *K* shows magnified image of four lateral line cells. *C*, *G*, and *H* are projections of confocal *Z*-stacks. Lateral view, anterior to left, dorsal top. Bar, 20 μm. According to zebrafish nomenclature guidelines, Gt indicates gene trap. See also Supplemental Figure S2.

Table 1. *Screen summary*

	FlipTrap with Tol2	FlipTrap with Ac/DS
F_0 screened	840	233
Citrine-positive F0	124	15
F_0 with multiple expression patterns	22	1
Total number of traps	154	16
Trapping rate	18.30%	6.70%

The ubiquitously expressed trap lines fall into two classes: those uniform in levels of expression, and those that exhibit heightened expression in a subset of tissues. As shown in Figure 1, F and G, the majority of the 14% with heightened expression display more intense Citrine fluorescence in specific tissues, suggesting differential expression regulation and perhaps increased requirements for the trapped genes in the specific tissue types. In a minority, we observed differential localization of the Citrine fusion in a subset of cell types; for example, the trap of *mitogen-activated protein kinase-activated protein kinase 2a* (*mapkapk2*), which is expressed in all cell types of the embryo but only localizes to the membrane in the vasculature (Fig. 1H) and vacuolating notochord cells (Supplemental Fig. S2C–E). The membrane localization of Mapkapk2-Citrine is spatially and temporally regulated, as notochord cells that have completed vacuolation do not localize Mapkapk2-Citrine (Supplemental Fig. S2F–H). This differential subcellular localization suggests that Citrine fusion proteins can be a direct readout of protein activity.

The FlipTrap lines that show uniform ubiquitous expression (49%) nonetheless display distinct subcellular localization patterns that are indicative of where the protein functions in the cell. Examples include *splicing factor 3a, subunit 3* (*sf3a3*), which localizes to the nucleus (Fig. 1I), and *magnesium transporter 1* (*magt1*), which localizes predominantly to the nuclear envelope in a number of tissues, including the lateral line and neural tube (Fig. 1K, localization shown for lateral line). The patterns of expression of the Citrine fusions colocalize with the patterns of antibody staining for the few cases where antibodies are available for the trapped gene (Supplemental Material; Supplemental Fig. S3).

Imaging of the living embryos revealed the localization of novel Citrine fusion proteins. Figure 1J shows the expression pattern of a novel six-C2H2 Zn finger protein has been trapped. Confocal imaging shows that this novel Zn finger Citrine fusion protein is diffusely localized in the cytoplasm with heightened expression at the cell membrane. While Zn fingers are often thought of as binding to DNA, there are examples of multi-Zn finger proteins that bind intracellular proteins and lipids (Luchi 2001). Visualization of the fluorescent protein in *Gt(znf-citrine)*^{ct65a} embryos suggests that this novel Zn finger Citrine fusion protein interacts with intracellular components and membrane lipids rather than DNA.

Annotating the functional proteome

The presence of the citrine exon sequence in the trapped transcript allows for identification of the trap gene by rapid amplification of cDNA ends (RACE). Using RACE, we identified 158 of the trapped genes. The majority of the lines capture orthologs of known vertebrate genes (66%) (Fig. 2B). Having the full-length gene of these orthologs tagged by Citrine provides opportunities to follow and assess these gene products in their full-length functional

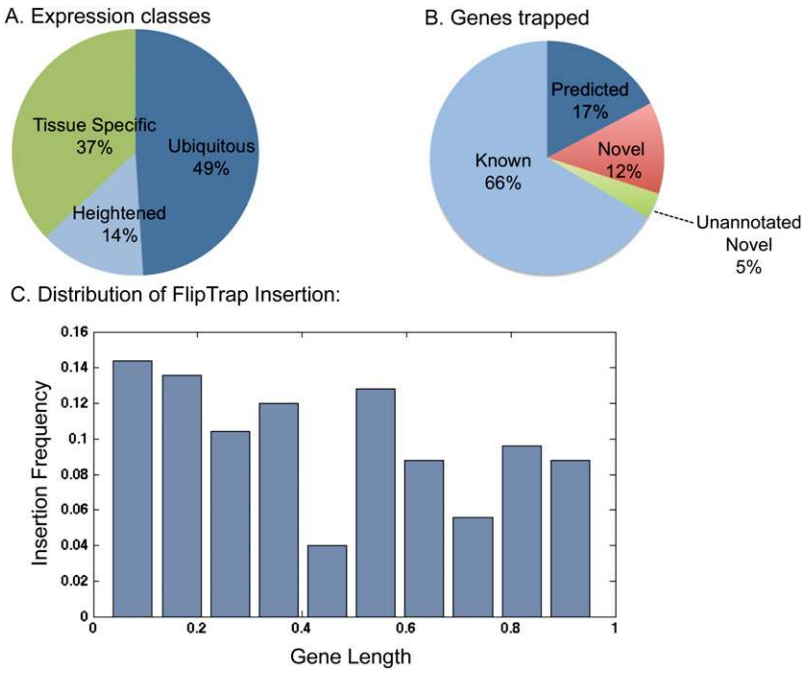

Figure 2. Classification of expression, genes, and distribution of FlipTrap insertion identified by screening. (*A*) Pie chart of the expression classes identified from screening of FlipTrap-injected F_1 embryos (N = 170). (*B*) Distribution of the classes of genes trapped by the FlipTrap vector (N = 158). (*C*) The distribution of the frequency of FlipTrap insertions mapped against gene length. (N = 128). Only lines isolated from the Tol2-containing vector were assessed due to the low number of lines isolated from the Ds-containing vector. Only lines with full in silico annotation by Ensembl were used for this analysis.

state. Twelve percent of the FlipTrap lines are of novel genes previously identified by ESTs, but have no known function attributed to them (Table 2). Seventeen percent are traps of hypothetical proteins that have been predicted to encode gene products based on genomic annotation (Table 3). The Citrine fusion proteins provide experimental data that these genomic sequences encode gene products. Finally, 5% of the trap genes do not correspond to known ESTs or predicted genes previously identified through standard genomic annotation, indicating that these are novel transcripts. Our ability to isolate these sequences by RACE indicates that these sequences correspond to genes that are actively transcribed, providing experimental data to complement in silico annotation efforts. The insertions sites for traps of hypothetical and novel genes were further confirmed by assessing the genomic integration of the FlipTrap vector by splinkerette PCR.

Sequence analysis shows that the FlipTraps generate in-frame fusions of the citrine exon, with the endogenous gene in 98.7% of FlipTrap lines ($N = 158$), indicating that there are few aberrations in splicing in lines expressing visible Citrine. The FlipTrap vector was designed with the citrine exon in frame 0; therefore, only integration between exons whose boundaries are in the same phase should lead to visible Citrine expression. The splicing of the citrine exon into frame with the endogenous trapped gene does not appear to be decreased by the large intron size; cloning of integration sites shows that the FlipTrap vector has made functional fusions after integration into introns ranging from 211 base pairs (bp) to 88 kb in length (average intron size 11.9 kb ± 23.5 kb). This indicates that the splice sites in the FlipTrap vector can function over a large genomic space.

As TEs have been reported to have integration site bias toward transcriptional start sites (Pan et al. 2005; Yant et al. 2005), we investigated the distribution of the FlipTrap integration by mapping the frequency of insertion sites against the length of the trapped genes (for the 128 trapped genes that have been annotated in silico; only Tol2-based FlipTraps were used for this analysis). The integration sites were mapped relative to the fractional gene length (Fig. 2C). The gene length was defined as the number of exons per gene. The distribution of the insertion sites against gene length appears relatively uniform, with a slight slope. To determine whether this reflected a significant bias, we performed χ^2 and Kolmogorov-Smirnov tests, comparing the observed frequency with that from 10,000 computer-simulated random integrations. The results suggest that we cannot reject the null hypothesis ($P > 0.97$) and that the insertions from our experiments are also random.

Functionality of Citrine fusion proteins

The majority of the FlipTrap lines (89.8%) have the citrine exon inserted between exons that together do not encode for single-protein domains (115 of 128 annotated), making it less likely that the Citrine fusion proteins disrupt function or interactions. We tested the functionality of the Citrine fusion proteins in the developing embryo by inbreeding the trap lines and assessing for morphological phenotypes in the homozygous individuals. We found that the majority of the FlipTrap embryos do not exhibit any phenotypes and are viable as homozygous (94.2%, $N = 122$) (see the Supplemental Material).

Of the seven Citrine fusion lines that exhibited phenotypes as homozygotes, one created a fusion between two adjacent genes [*forkhead box D3* (*foxd3*) and *asparagine-linked glycosylation 6 homolog* (*alg6*); *Gt(foxd3-citrine-alg6)*^{ct110a}]; in contrast, one created a clear phenotype without creating detectable defects in splicing or interruption of known protein domains [*Gt(rfx2-citrine)*^{ct81a}] (Supplemental Fig. S5C). In the remaining five mutagenic lines, the FlipTrap vector inserted between exons encoding for a single defined protein domain; thus, in these cases, the Citrine fusion would be expected to disrupt protein motifs that are essential for the function of the trapped gene. For example, in the FlipTrap of *5′-nucleotidase, cytosolic IIa* (*nt5c2*), *Gt(nt5c2-citrine)*^{ct103b}, the insertion site disrupts the haloacid dehydrogenase (HAD) domain. HAD domains form hydrolase folds that are important for substrate binding (Cronin et al. 2002; Kim et al. 2004). The homozygous *Gt(nt5c2-citrine)*^{ct103b} embryos display defects in posterior

Table 2. *Novel gene*

Alleles	Expression	Gene	NCBI accession number
ct30a	Neuronal, MHB, branchial arches, heart	RNA-binding motif, single-stranded interacting protein	NM_001076716
ct33a	Ubiquitous	Tetratricopeptide domain	NM_173285
ct47a	Ubiquitous, prominent in neural tube	Protein phosphatase 2, regulatory subunit B, δ	NM_213318
ct49a	Ubiquitous	Erythrocyte membrane protein band 4.1-like 3b	NM_214812
ct56a	Ubiquitous	LOC327040	NM_212698
ct58a	Ubiquitous	Dipeptidyl-peptidase 9	NM_001077313
ct65a	Ubiquitous	LOC792098	NM_001123305
ct71a	Tip of tail, notochord, MHB	Family with sequence similarity 101, member B	XM_002664625
ct82a	Ubiquitous	Transmembrane protein 44	NM_001045355
ct119a	Ubiquitous	LOC55377,extracellular domain	NM_001020737
ct122b	Ubiquitous	Ribosomal L22 domain	NM_212760
ct126b	Ubiquitous	Protein kinase C-binding protein NELL2	NM_001080075
ct140a	Ubiquitous	LOC431762	NM_001002215
ct149a	Ubiquitous	Coiled-coil domain-containing protein 109A	NM_001077325

Table 3. *Predicted genes*

Alleles	Expression	Gene	NCBI accession number
ct30b	Ubiquitous, heightened in tip of tail	Pappalysin-1 precursor	NM_001123252
ct30c	Neural tube	Guanylate kinase	XM_002663262
ct39a	Jaw	Zinc finger protein 385C	XM_001921873
ct42a	Skeletal muscle	Myosin, heavy chain 7b, β	XM_003201195
ct51a	Ubiquitous	Zn finger protein, transcriptional repressor	NM_001123276
ct69a	Ubiquitous, heightened in jaw	Zinc finger protein 385C	XM_688397
ct78b	Ganglia, spinal cord	RGD motif, leucine-rich repeats, tropomodulin domain and proline-rich	XM_691060
ct96a	Midbrain	Spermatogenesis-associated protein 5-like	XM_002663196
ct98a	Tail fin and pectoral fin	Collagen α-1(IX) chain-like	XM_003200525
ct99a	Ubiquitous, heightened in telencephalon	Vacuolar protein sorting 41 homology	XM_686579
ct110b	Trunk somite, pharyngeal arch, fins	Plectin 1/zinc finger, RAN-binding domain 1	XM_003200128
ct123a	Tail vein, proneprhric, and gut	IMAGE:7139509, VAMP	XM_003199278
ct127a	Ubiquitous	Exocyst complex4	NM_001030134
ct133a	Neural tube	Si:ch211-245g11.2, RNA poly II domain	NM_680633
ct146a	Muscle	Transmembrane protein C14orf109 homolog	NM_001033746
ct151a	Ubiquitous, heightened in spinal cord, eye, ear	Replication factor C (activator 1) 1	NM_001089491
ct158a	Ubiquitous	Pyridoxial kinase-like	XM_002660547
ct162a	Telencephalon, hindbrain	RAB3A, member RAS oncogene family, b	NM_001017761
ct168a	Telencephalon, midbrain, hindbrain	Sine oculis-binding protein homolog a	NM_001098618
ct170a	Neural tube	Phosphoinositide-specific phospholipase C β	

tail formation in the homozygous embryos (Supplemental Fig. S5D). Thus, the majority of the Citrine fusions that disrupt the function of the trapped gene reflect disruption of recognized protein domains.

Database resource

We developed two Web-based MySQL interfaces to store and report the results of FlipTrap expression and molecular data. The data entry interface allows for uploading of images, metadata, sequence data, and annotations, as well as curation of the data. The querying interface provides FlipTrap expression, integration site, and gene sequence data to other researchers as a general resource (http://www.FlipTrap.org). As additional FlipTrap lines are generated and the trapped genes are identified, the data will be made public through the Web site.

Both interfaces are semantically enabled, being underpinned by the zebrafish anatomical ontology and associated data products defined by Zfin. For data entry, this allows only metadata specific to a particular developmental stage to be entered; for example, selecting a particular stage of development limits a selection of anatomical features to only those that are present during the development stage selected.

For the purpose of data retrieval, the FlipTrap expression and molecular data can be queried by four parameters; gene name, anatomy, developmental stage, and expression pattern. The semantic infrastructure allows these queries to be contextually aware; for example, a request for data relating to a particular anatomical feature that is present at a specific stage of development will retrieve data relating to relevant developmental features in earlier stages. Additionally, smart text is used for anatomical query that is underpinned by the zebrafish anatomical ontology. Terms that match the zebrafish anatomical ontology are suggested as the user types in the query. Retrieved data are dynamically augmented with associated data from the Ensembl, NCBI, and University of California at Santa Cruz Genomic Browser Web sites in real time. Changes to the ontologies—new terms and definitions, restructurings, etc.—are propagated through the system so that data retrieval always employs the most up-to-date scientific understanding. Figure 3 is an example of the expression report provided for *Gt(rbms3-citrine)*^{ct30a}.

Mutating the proteome by Cre-lox-mediated recombination in FlipTraps

In addition to generating full-length fusion proteins, the FlipTrap vector contains elements to enable the creation of Cre conditional mutant alleles. In reverse orientation to the citrine exon are the coding sequences of mCherry (a variant of red fluorescent protein), a stop codon, and a polyA signal (Figs. 1A, 4A). Heterotypic *lox* sites flank the citrine/splice donor sequence, as well as the mCherry/polyA sequence. In the presence of Cre, the *lox* sites undergo reversible flipping, followed by irreversible deletion. This excises the citrine/splice donor sequence and "flips" the mCherry/polyA sequence into the forward orientation immediately after the splice acceptor (Fig. 4A); thus, a traditional gene trap allele is generated with a splice acceptor site immediately upstream of the reporter (mCherry) sequence. Transcription and splicing of the mutant allele results in the generation of a fusion mRNA that is devoid of the 3′ exon(s) of the trapped gene downstream from the insertion site. Translation of the mRNA from the flipped

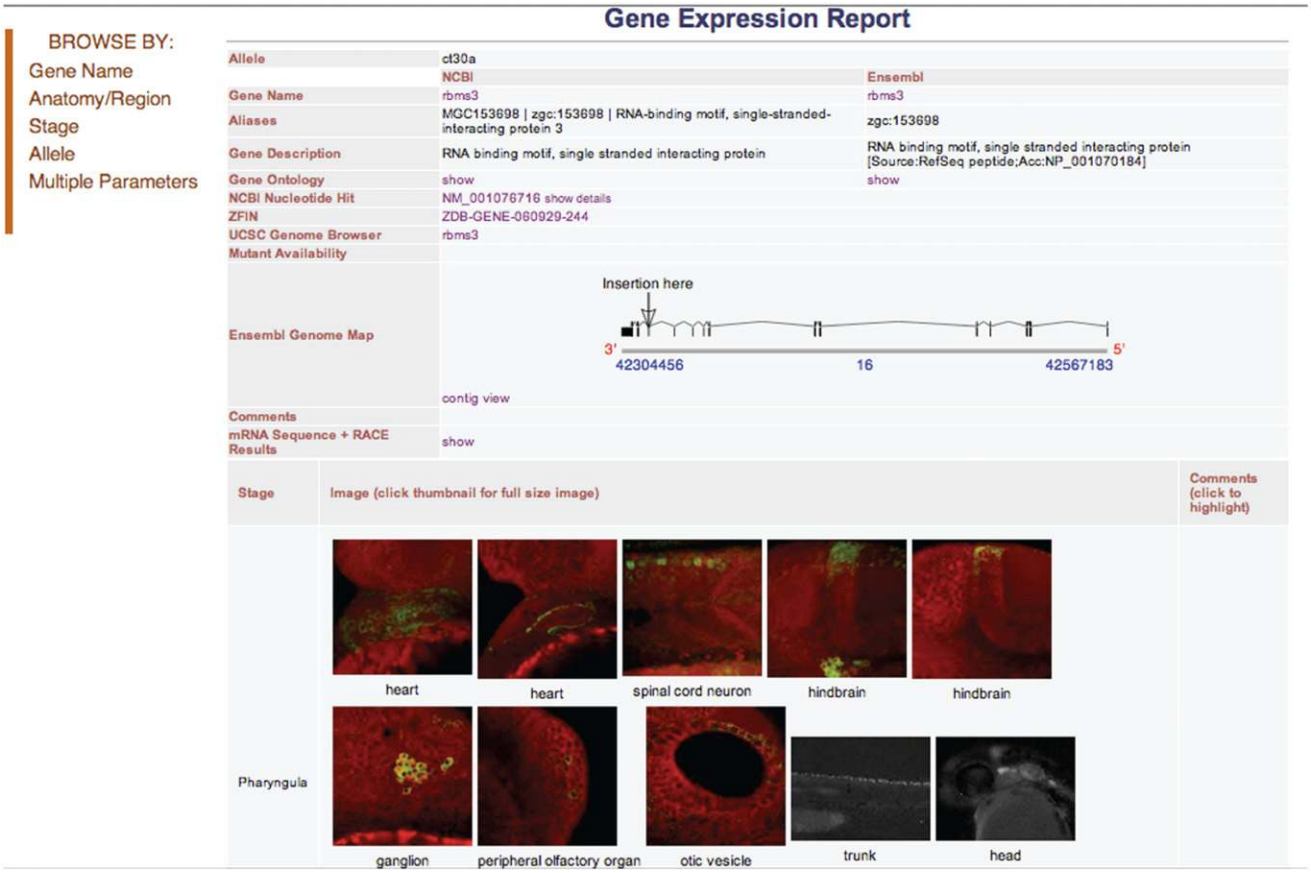

Figure 3. Web-based interface for storage and retrieval of FlipTrap expression and molecular data. The Web-based interface database stores the FlipTrap expression and molecular data, which can be queried by four parameters: gene name, anatomy, developmental stage, and expression pattern. An example of a gene expression report for *Gt(rbms3-citrine)*^{ct30a} displays associated molecular data consisting of allele designation, gene name and alias, and links to NCBI, Zfin, Ensembl, and the University of California at Santa Cruz Genome Browser. The FlipTrap integration site is mapped to genomic annotation of the trapped locus based on Ensembl data. Wide-field and confocal fluorescent images of a protein expression pattern are displayed based on developmental stage and are anatomically annotated.

trap therefore produces a truncated protein fused to mCherry (Fig. 4A).

We tested recombination of the lox sites in the FlipTrap lines by two methods: injection of *cre* mRNA into FlipTrap embryos (Fig. 4B–E), and breeding with a transgenic line expressing a Cerulean-Cre fusion under the control of the *β-actin2* promoter *Tg(bactin2:cerulean-cre)*^{ct5000} (Fig. 4F–I). The Cerulean-Cre fusion was created to complement the FlipTrap technology, as it permits the tracking of Cre recombinase expression in the developing embryo (details in the Supplemental Material). Both injections of *cre* mRNA into FlipTrap lines and breeding with the *Tg(bactin2:cerulean-cre)*^{ct5000} fish lead to efficient recombination of the lox sites as observed by conversion of Citrine to mCherry fusion proteins. In the few cells where recombination has not occurred, perhaps due to mosaic expression of Cre recombinase in the F_0 embryos, Citrine expression remains (Fig. 4B–I, arrowhead). Once recombination occurs in the germline, all cells convert from Citrine expression to mCherry, as expected (see Supplemental Fig. S6C–F). Therefore, the conversion or lack of conversion of Citrine to mCherry expression provides a readout for Cre-lox recombination.

To assess the efficiency of mutagenesis by Cre-lox recombination, we performed real-time quantitative PCR (RT-qPCR) to determine the level of wild-type transcripts in Cre-induced homozygous mutant embryos (Fig. 4J). The presence of the polyA signal and absence of a splice donor in the Cre-induced mutant alleles lead to the exclusion of exons 3′ of the insertion sites in mutant mRNAs; therefore, we used primers for RT-qPCR for each trapped gene that span exons flanking the site of the FlipTrap insertion. Using this approach, we assessed the level of wild-type transcripts in six of the Cre-induced mutant alleles (Fig. 4J). We found varying levels of wild-type transcript knockdown for each line, ranging from a 70% [*Gt(hmga2-mCherry)*^{ct29aR}] to a 97% [*Gt(msi2b-mCherry)*^{ct57aR}] knockdown of wild-type transcripts in the homozygous mutant embryos, indicating that Cre-induced mutant alleles can lead to efficient knockdown of the trapped genes. The residual levels of wild-type transcript in the homozygous Cre mutant embryos suggest that the endogenous splicing machinery can skip over the mCherry 3′ exon to generate a low level of wild-type transcript. Importantly, 80% knockdown of the *tpm4* transcript in the *Gt(tpm4-mCherry)*^{ct31aR}

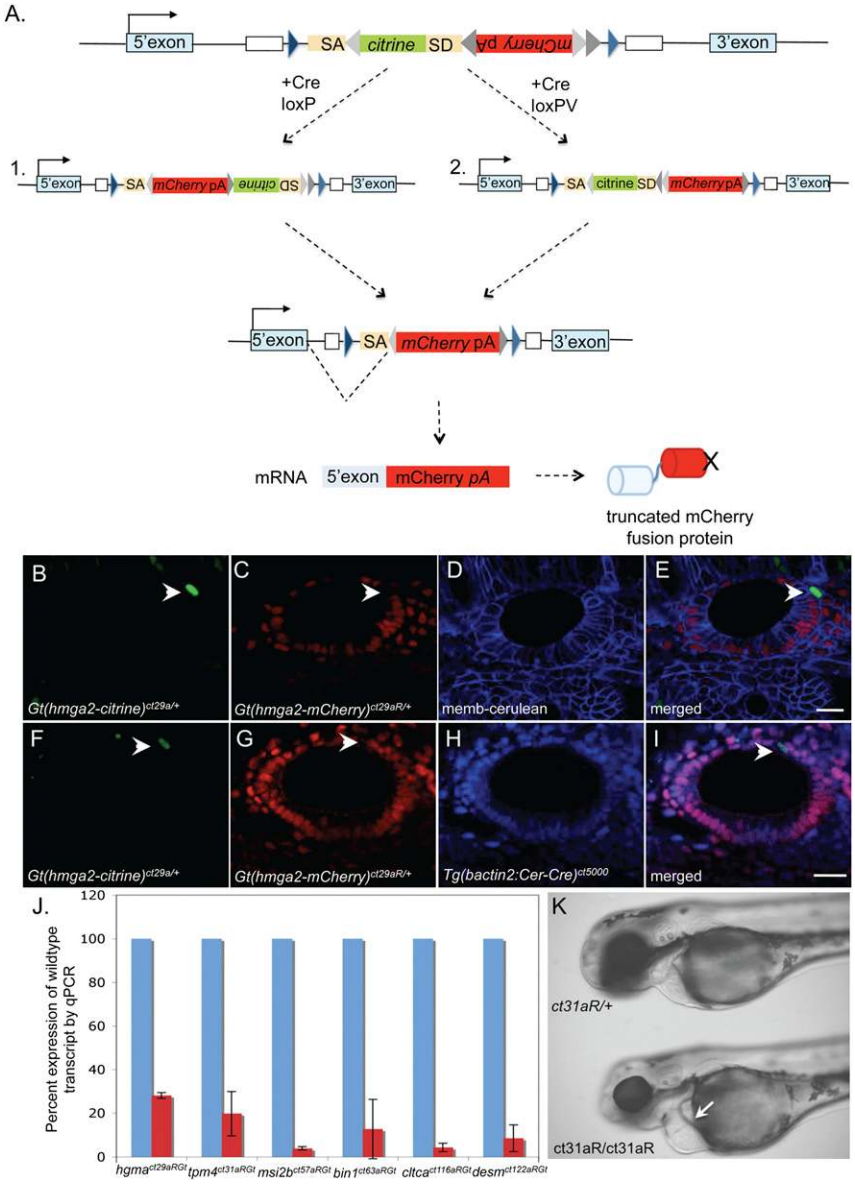

Figure 4. Cre-lox-mediated recombination of FlipTrap lead to knockdown of trap genes. (*A*) Schematic of Cre-mediated recombination. Cre recombination of the *lox* sites lead to two intermediates: (1) Recombination of the *loxP* sites lead to flipping of the mCherry and polyA sequences into the forward orientation and citrine and the splice donor into the reverse orientation. (2) Recombination of the *loxPV* sites lead to flipping of the mCherry and polyA sequences into the forward orientation. Further recombination of either intermediate lead to the excision of the citrine and splice donor, resulting in a mutant gene trap allele that contains a splice acceptor, followed by a 3′ exon encoded by mCherry. Expression of the Cre-induced mutant allele lead to the production of a truncated protein fused to mCherry. (*B–E*) Confocal images of the otic vesicle of a *Gt(hmga2-citrine)*^{ct29a} embryo injected with *cre* and membrane-cerulean mRNA. (*F–I*) Confocal images of the otic vesicle in embryos from a cross between *Gt(hmga2-citrine)*^{ct29a} and *Tg(bactin2:cerulean-cre)*^{ct5000} adults. (*B,D*) Citrine expression after Cre-lox recombination. (*C,G*) Expression of mCherry upon Cre-lox recombination. (*D*) Membrane-cerulean counterstain. (*H*) Expression of Cerulean-cre localized to the nucleus *Tg(bactin2:cerulean-cre)*^{ct5000} in embryos. (*E,I*) Merges of *B–D* and *F–H*, respectively, demonstrate that expression of either Cre (*D*) or Cerulean-Cre fusion protein (*I*) leads to the recombination and conversion of the Citrine to mCherry fusion protein. The arrowhead points to Citrine-positive nuclei that have not undergone Cre-lox recombination in the somatic tissue of progenies from a *Gt(hmga2-citrine)*^{ct29a} adult crossed to a *Tg(bactin2:cerulean-cre)*^{ct5000} adult. Bar, 20μm. (*J*) Quantification of relative wild-type transcripts in wild-type (blue) and homozygous Cre-induced mutant embryos as determined by RT-qPCR. Relative levels of transcripts in wild-type siblings and homozygous mutant embryos were normalized to *glyceraldehyde 3-phosphate dehydrogenase* (*gapdh*) transcripts. Transcript levels of the trap gene in wild-type siblings are expressed as 100%, while transcript levels in homozygous mutant embryos are represent as a percentage relative to wild-type siblings. Error bars represent standard deviations from triplicate RT-qPCR experiments performed on five to 10 embryos per line. The trapped gene and designated alleles are listed in the *X*-axis. (*K*) Bright-field image of wild-type sibling (*top* of image) and homozygous mutant embryos (*bottom* of image) for *Gt(tpm4a-mCherry)*^{ct31aR} at 48 hpf. Homozygous mutant embryos exhibit defects in cardiac contraction and pericardial edema (arrow). See also Supplemental Figures S5 and S6.

line is sufficient to induce a mutant phenotype, as the homozygous embryos for *Gt(tmp4a-mCherry)*^{ct31aR} exhibit clear defects in cardiac contraction and pericardial edema (Fig. 4K).

As a test of the efficacy of Cre-induced mutagenesis, we compared the phenotype of the Cre-recombined allele from the integrant into *clathrin heavy chain a* (*cltca*) with a previously identified mutant zebrafish (Amsterdam et al. 2004). The previous retroviral insertion mutant *cltca* (hi2462Tg) exhibits defects in head, jaw, and cardiovascular development, as well as general necrosis (Amsterdam et al. 2004). In the *Gt(cltca-citrine)*^{ct116a} line, the FlipTrap vector has inserted between exons 27 and 28 of 31 exons, fusing Citrine between the sixth and seventh CLH (clathrin heavy chain repeat homology) domains of Cltca (Fig. 5A). The Cltca-Citrine fusion protein is expressed throughout the embryo, with heightened expression in the vasculature (Fig. 5C). The Cre-induced mutant allele *Gt(cltca-mCherry)*^{ct116aR} should delete the seventh CLH domain in Cltca (Fig. 5B). RT-qPCR for wild-type transcripts in the homozygous *Gt(cltca-mCherry)*^{ct116aR} mutant embryos reveals that only 5% of the normal level of wild-

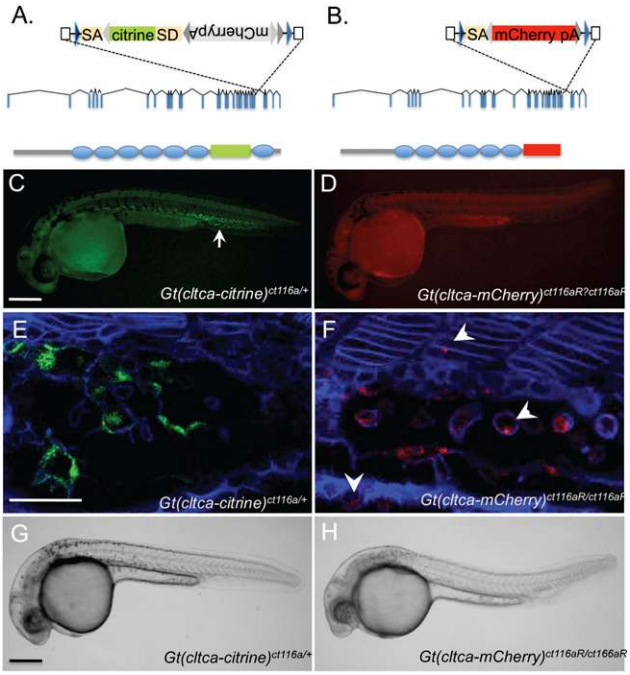

Figure 5. Cre-induced mutant phenotype in FlipTrap of *cltca*. (*A*,*B*) Schematic of *cltca* with FlipTrap insertion before (*A*) and after (*B*) Cre recombination. Predicted protein domains in the *cltca* gene (below) as determined by Simple Modular Architecture Research Tool (SMART) with Citrine (*A*, green bar) and mCherry (*B*, red bar) protein fusion. CLH domains are shown as blue ovals. In the cre mutant allele *Gt(cltca-mCherry)*^{ct116aR}, the terminal CLH domain is deleted. (*C*,*D*) Wide-field fluorescence image of *Gt(cltca-citrine)*^{ct116a} (*C*) and *Gt(cltca-mCherry)*^{ct116aR} (*D*) embryos taken with YFP and RFP filters, respectively. Cltca-Citrine is expressed ubiquitously but is heightened in the tail vasculature (arrow), while in the cre mutant allele *Gt(cltca-mCherry)*^{ct116aR}, mCherry is expressed uniformly. (*E*,*F*) Confocal image of the tail vasculature in *Gt(cltca-citrine)*^{ct116a} (*E*) and *Gt(cltca-mCherry)*^{ct116aR} (*F*) in which the cell membranes are labeled with a membrane-localized Cerulean protein (blue). In the cre mutant allele, mCherry expression in the tail vasculature is at levels similar to surrounding cell types (arrowhead). (*G*,*H*) Bright-field image of *Gt(cltca-citrine)*^{ct116a} (*G*) and homozygous *Gt(cltca-mCherry)*^{ct116aR} (*H*) embryos at 32 hpf. Homozygous *Gt(cltca-mCherry)*^{ct116aR} embryos have a smaller forebrain and midbrain, body curvature, and dilated caudal vein. Bars: *E*,*F*, 20 μm; *C*,*D*,*G*,*H*, 50 μm. See also Supplemental Figure S7.

type *cltca* transcripts remain (Fig. 4J). In contrast to the Cltca-Citrine fusion protein, the Cltca-mCherry appears more uniformly distributed in both heterozygous and homozygous mutant embryos (Fig. 5E). Homozygous mutant *Gt(cltca-mCherry)*^{ct116aR} embryos display general necrosis and have a smaller forebrain and midbrain, body curvature, and a dilated caudal vein (Fig. 5H; Supplemental Fig. S7, higher magnification of mutant embryos), similar to the hi2462Tg mutant (Amsterdam et al. 2004). Comparison of the Citrine and mCherry expression suggests that the last CLH domain in Cltca is essential for maintaining the wild-type level of Cltca expression in the vasculature and for the normal development of the forebrain, midbrain, and vasculature. These results demonstrate that the FlipTrap lines provide an in vivo approach for assessing the structure–function of the proteome.

Disruption of protein localization in the Cre-induced mutant allele

The Citrine fusions provide a direct readout of the subcellular localization of the trapped gene products, permitting comparisons with the mCherry truncated forms to give insights into the domains responsible for subcellular localization. In the *Gt(col1al-citrine)*^{ct98a} line, the FlipTrap vector has integrated between exons 6 and 7 of the nine-exon *collagen α-1(IX) chain-like* (*col1al*) gene (Fig. 6A). Col1al is a novel collagen protein in which the FlipTrap insertion has created a Citrine fusion that is bracketed by a thrombospondin and a collagen domain (Fig. 6A). Col1al-Citrine is expressed in the extracellular matrix of the developing tail fin in the striated rays starting at 26 h post-fertilization (hpf) (Fig. 6C). The Cre-induced mutant allele *Gt(col1al-mCherry)*^{ct98aR} deletes the collagen domain (Fig. 6B). Heterozygous *Gt(col1al-mCherry)*^{ct98aR} mutant embryos appear wild type (data not shown). However, homozygous *Gt(col1al-mCherry)*^{ct98aR} mutant embryos display defects in the tissue integrity of the tail mesenchyme, with large holes appearing in the tail mesenchyme (Fig. 6E,F, arrowhead). Consistent with this phenotype, the Col1al-mCherry mutant protein is expressed in the cytoplasm of the tail mesenchyme cells (Fig. 6F). The different localizations of full-length Col1al-Citrine and truncated Col1al-mCherry proteins suggest that *col1al* is expressed by the tail mesenchyme and exported into the extracellular matrix to form arrays of Col1al as a structural component of the tail. In situ hybridization for the *col1al* supports this model, as *col1al* transcript is expressed in the tail mesenchyme, while antibody to Citrine reveals that the Col1al-Citrine protein is localized to the extracellular matrix (Fig. 6G–I). Therefore, the combined ability to assess both full-length and truncated fusion proteins provides a direct means to study the molecular mechanisms of protein expression, trafficking, and function.

Versatility of FlipTraps through targeted genetic manipulation of the trap locus

In answer to the absence of homologous recombination technology in zebrafish, we designed the FlipTrap vector to enable targeted genetic manipulation of each FlipTrap line. Heterotypic FRT sites (FRT and FRT-F3) are placed inside the two TE sequences in the FlipTrap vector (Figs. 1A, 7A), permitting any FlipTrap to be excised and replaced by another DNA cassette. In the presence of Flp recombinase and exogenous DNA containing the same FRT sites, recombination replaces the FlipTrap cassette with the exogenous DNA, allowing for targeted genetic manipulation at each FlipTrap locus. To test the functionality of the FRT sites, we replaced the inserted FlipTrap cassette in the *Gt(desma-citrine)*^{ct122a} line with another cassette containing a splice acceptor followed by mCherry coding sequencing and a polyA signal (Fig. 7A). Similar to

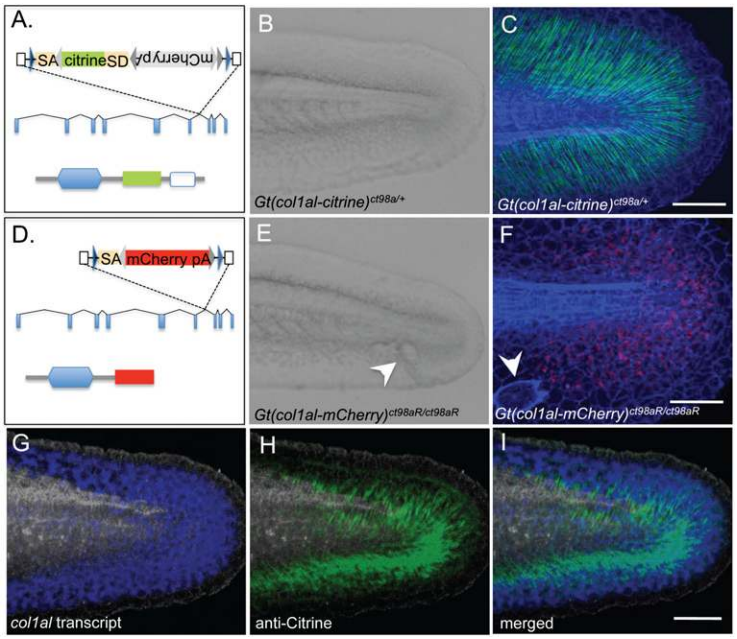

Figure 6. Mutant allele disrupts localization of novel collagen protein. (*A*,*D*) Schematic of *col1al* with FlipTrap insertion before (*A*) and after (*D*) Cre recombination. Predicted protein domains in genes as determined by SMART analysis with Citrine (*A*, green bar) and mCherry (*D*, red bar) protein fusion are shown as follows: The thrombospondin N-terminal-like domain is a blue hexagon, and the collagen domain is a blue-outlined rectangle. In the Cre-induced mutant allele, the collagen domain is deleted from the Col1al-mCherry protein. (*B*,*E*) Bright-field image of tail region of *Gt(col1al-citrine)*^{ct98a} (*B*) and homozygous *Gt(col1al-mCherry)*^{ct98aR} (*E*) embryos at 28 hpf. Holes are present in the tail mesenchyme of homozygous *Gt(col1al-mCherry)*^{ct98aR} embryos (arrowhead). (*C*,*F*) Projection of confocal *Z*-stacks of the tail region of *Gt(col1al-citrine)*^{ct98a} (*C*) and homozygous *Gt(col1al-mCherry)*^{ct98aR} (*F*) counterstained with phalloidin conjugated to Alexa-633 (blue). (*C*) Col1al-citrine fusion proteins (green) are localized as striated bands in the extracellular matrix of the tail mesenchyme. (*F*) The truncated Col1al-mCherry fusion protein (red) localizes as puncta within the cytoplasm of the tail mesenchyme. (*G–I*) Projection of confocal *Z*-stacks of *col1al* transcript (*G*, blue) and Col1al-citrine protein (*H*, green) expression in the tail region with antibody staining to β-catenin (white) as counterstain. *col1al* is expressed in the tail mesenchyme cells, while Col1al-Citrine protein localizes to the extracellular matrix. Bar, 20 μm.

the Cre-rearranged FlipTrap, this cassette will truncate the trapped gene at the insertion, fusing mCherry in place of the final exon of the trapped gene. *Gt(desma-citrine)*^{ct122a} is a trap of *desmin a* (*desma*), a component of the myoseptal junctions in cardiac and skeletal muscles. The truncation of Desma-mCherry does not result in morphological or subcellular localization defects (data not shown). However, the distinct subcellular localization of Desma in the myoseptal junctions provides a clear readout of cassette exchange occurring at the *desma* locus. We found that injecting the replacement mCherry cassette with *flp* mRNA into *Gt(desma-citrine)*^{ct122a} embryos leads to expression of Desma-mCherry in the cardiac and skeletal muscles cells that normally express Desma-Citrine. This conversion was observed in 94% of the injected embryos (N = 122), indicating that cassette exchange in the FlipTrap line is efficient and can be used to generate variants for each line.

Discussion

The FlipTrap is a versatile and multifunctional vector, enabling both the visualization of the functional proteome and the interrogation of the proteome's function through a single integration event. To visualize the functional proteome, the FlipTrap creates fluorescent tags of full-length proteins at their endogenous loci. This is in contrast to previous gene trap approaches in vertebrates in which the trapped protein is truncated (Skarnes et al. 1992; Kawakami et al. 2004; Clark et al. 2011). We show from expression analyses of the Citrine fusion proteins that the FlipTrap lines can provide protein expression information at cellular and subcellular resolutions. The subcellular localization can be predictive of protein function, as we show that some Citrine fusion proteins exhibit differential localization in different cell types or within the same cells as they differentiate. Cre-lox recombination converts the FlipTrap lines to mutant alleles that tag only the 5′ end of the trapped loci to study protein function. Additionally, the FlipTraps enable targeted genetic manipulation of the trap locus through Flp-FRT-mediated cassette exchange, providing a means to make variants from each line.

Our results show that the FlipTrap vector can efficiently tag full-length proteins in zebrafish. We isolated a library of FlipTrap lines that have the endogenous full-length proteins tagged with Citrine. Annotation of these lines reveals novel genes (12%) as well as provides experimental data for the expression of predicted genes (17%). The majority (94%) of the Citrine fusion proteins generated by the FlipTrap vector do not exhibit mutant phenotypes as homozygous, suggesting that they are functional. Multiple genome-wide studies have shown that exon boundaries correlate with functional protein domains (Doolitte 1995; Liu and Grigoriev 2004; Zhang and Chasin 2004; Gudlaugsdottir et al. 2007). Consistent with this correlation, mapping of the integration sites show that the majority of the FlipTrap insertions are between exons that do not span functional protein domains (89.8%). Additionally, protein domains within single coding sequences are independently folding modules that can be expressed in isolation (Pawson 1995). This makes it possible for a fully functional fusion protein to be generated by the insertion of the artificial citrine exon between endogenous exons. Finally, the screen criteria for the presence of a fluorescence signal excludes all integration events that lead to out-of-frame fusions and misfolded proteins, ensuring that the

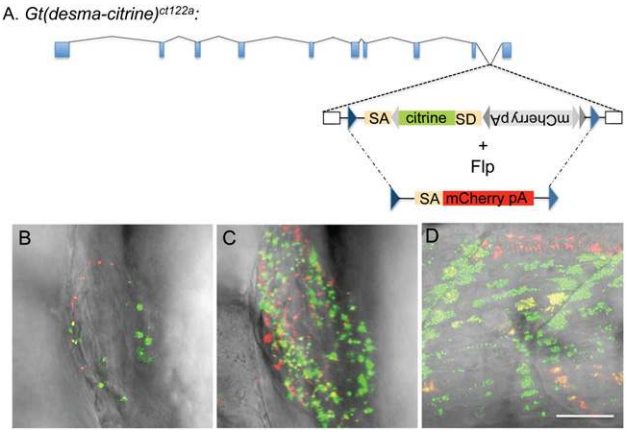

Figure 7. FRT-Flp-mediated cassette exchange in FlipTrap. (*A*) Schematic of cassette exchange system applied to a FlipTrap of the *desma* gene *Gt(desma-citrine)*^{ct122a}. The exogenous DNA cassette contained FRT sites (dark-blue triangles), a splice acceptor (SA, peach), and a 3′ exon encoding mCherry (red). In the presence of Flp recombinase and exogenous DNA, the FlipTrap cassette is replaced and converted from a Citrine fusion to a mCherry fusion protein. (*B–D*) Confocal image of a *Gt(desma-citrine)*^{ct122a} embryo injected with Flp recombinase and the conversion of Desma-Citrine expression to Desma-mCherry after Flp-FRT recombination. (*B*) Confocal image of the heart tube in a *Gt(desma-citrine)*^{ct122a} embryo after FRT-Flp-mediated cassette exchange showing a portion of the myocardial cells have exchanged the citrine exon for the mCherry cassette (red). (*C*) Projection of confocal *Z*-stack of the entire heart of the embryo in *B*. (*D*) Confocal image of trunk muscle cells in *Gt(desma-citrine)*^{ct122a} embryos after cassette exchange in a few muscle fibers as indicated by expression of Desma-mCherry (red). Confocal images were overlaid with a Nomarski image. Bar, 50 μm.

full-length proteins isolated through the screen are biased toward being functional.

In the few cases where the Citrine fusion protein leads to mutant phenotypes in the homozygous embryos, the FlipTrap had typically inserted between exons that together encode a functional domain. These cases provide lessons for the rational design of fluorescent fusion proteins in which N-terminal or C-terminal tags are nonfunctional. In *S. cerevisiae*, 18% of the essential genes show impaired protein function when tagged on the C terminus (Gavin et al. 2002). Given that a number of motifs must occur at the N terminus or C terminus of a protein—including KDEL for ER retention, WRPW for interaction with the Groucho family of repressors, and the ligand of PDZ-binding domains (Munro and Pelham 1987; Wainwright and Ish-Horowicz 1992; Doyle et al. 1996)—terminal tagging of these proteins would be expected to disrupt their function. The FlipTrap insertion data suggest that internal tagging of these classes of proteins such that the fluorescent tag is placed between exons that do not span protein domains would offer strategic advantages.

The ability to fluorescently tag full-length proteins as they are expressed from their endogenous loci provides a number of advantages over conventional transgenesis and traditional gene trapping, in which only the 5′ ends of proteins are tagged. The FlipTraps retain all regulatory sequences of the endogenous gene, including untranslated regions (UTRs), which ensures the retention of transcriptional and translational regulatory elements that would affect the expression levels, localization, and function of the protein. Because the full-length Citrine fusion proteins are expressed from their endogenous loci with all of their interacting domains, the FlipTrap lines can provide an accurate assessment of spatial and temporal gene expression patterns that neither transgenes nor tagging of only the 5′ fragment are able to do reliably. This is nicely reflected in the trap lines that are ubiquitously expressed but show heightened levels of Citrine fusion protein or unique subcellular localization in specific tissues (Fig. 1F–H).

The retention of localization signals by the tagging of the full-length protein provides subcellular localization data that can be informative of protein function. We show several examples in which splice components and transcription factors localize to the nucleus, while structural components in the muscle or extracellular matrix localize in predicted subcellular compartments. Additionally, the FlipTraps circumvent possible overexpression phenotypes and create the possibility of quantifying in vivo protein expression levels with fluorescent imaging. Furthermore, the presence of the Citrine tag provides a label that avoids the need for antibodies, which might cross-react, perturb function, or be in limited supply. Finally, the ability to detect full-length proteins in the embryos without fixation and staining provides an avenue for the assessment of protein dynamics in vivo to directly visualize the biological function of the trap gene.

To interrogate the proteome, the Citrine fusion proteins can be converted to mutant alleles by Cre recombination, bringing Cre-mediated conditional mutagenesis to a vertebrate in the absence of homologous recombination. The Cre mutant allele efficiently disrupts production of wild-type transcripts. The conversion of the FlipTrap from a wild-type to mutant allele is visible through the switching of fluorescent fusion protein expression from a full-length Citrine fusion to a truncated mCherry fusion protein. We demonstrate that Cre recombination can lead to mutant phenotypes for the *cltca* gene, matching an insertional mutant phenotype (Amsterdam et al. 2004). The truncated mCherry fusion protein can disrupt localization of the trapped protein (Fig. 6C,F), underscoring the importance of being able to visualize both the full-length Citrine fusion and truncated mCherry mutant proteins. The ability to generate mutant alleles from the FlipTrap lines by Cre-*lox* recombination enables conditional mutagenesis in zebrafish and any system to which the FlipTrap vector can be applied. Cre alleles of FlipTrap lines with 3′ insertions may be hypomorphic, as the mutant allele will retain most of the endogenous protein. Whether a FlipTrap line will lead to a null, hypomorph, or neomorph protein will depend on the integration site of the vector. To ensure null mutant alleles following Cre-mediated recombination, next-generation FlipTrap vectors can include a protein degradation sequence such as PEST to degrade the trapped protein.

While Cre recombination has been a powerful genetic mutagenesis tool in mice, it has only been available in a few select model organisms that have the technology for homologous recombination in embryonic stem cells (cf. mice and rats) and in single-cell organisms such as yeast and bacteria. In systems without experimentally manipulable homologous recombination, such as zebrafish, Cre recombination has been used mainly to convert transgenic reporters such as GFP or RFP in specific tissues for cell-tracking experiments and does not target the endogenous locus for mutagenesis (Hans et al. 2009; Collins et al. 2010). The FlipTrap approach enables Cre-mediated conditional mutagenesis of endogenous loci in the absence of traditional homologous recombination technology. FlipTraps thus open up the potential for Cre conditional alleles in any species with exons (which includes most eukaryotes) and in which a transposon will work. The Tol2 and Ac transposons have already been shown to mobilize in species ranging from *Drosophila* (Urasaki et al. 2008), zebrafish, and *Xenopus* to chickens, mice, and humans (Emelyanov et al. 2006; Kawakami 2007; Sato et al. 2007).

The ability to perform cassette exchange by Flp-FRT-mediated recombination permits conversion of the FlipTrap lines into any variant of choice, enabling targeted genetic manipulation of each trap locus. Directed gene manipulation has been a powerful tool for engineering specific gene mutations to dissect mechanisms of gene function. While this technology exists in a limited number of model organisms, it has been elusive to studies in zebrafish. Cassette exchange in the FlipTrap lines adds a level of versatility that provides endless possibilities for each FlipTrap line. It permits conversion of the tissue-specific FlipTrap lines into driver lines or conversion of any FlipTrap to a fusion of a different color to simplify imaging of different gene products in a single embryo. Additionally, FlipTrap lines with insertions at the C terminus that would be hypomorphic upon Cre recombination can be targeted for degradation through the addition of protein degradation signals. These multiple capacities of the FlipTrap lines offer the ability to assess the proteome by fluorescently tagged full-length protein to visualize biological function of genes at the subcellular level in the living embryo and simultaneously dissect their function through conditional mutagenesis and targeted genetic manipulation.

Materials and methods

Zebrafish maintenance and strains

Adult fish and embryos were maintained as described (Westerfield 1994). The AB strain was used to generate the mosaic F_0 adults for screening. A mix genetic background AB/TL strain was bred to mosaic F_0 AB adults for screening. The FlipTrap lines were maintained in a mix genetic background of AB/TL.

Construction of the FlipTrap vector

The FlipTrap vector was generated by fusion PCR. Briefly, PCR fragments of citrine, mCherry, *α-tubulin* polyA, *rassf8* splice acceptor and donor, lox, and FRT sequences were amplified independently. The fragments were linked by fusion PCR using the following pairs of primers: F3-F (5′-AGTCTACGCCCCCAACT GAGAGAACTCAAAGGTTACCCCAGTTGGGGCACTACAT CGATTCAGGAACCTCACAGACTGC-3′) and F3-R (5′-TCCC GGGTGAATGTGTAGCGACCAGTTCG-3′), F4-F (5′-TCGCTA CACATTCACCCGGGATAACTTCGTATAGCATACATTATAC GAAGTTATCCGGAGTGAGCAAGGGCGAGGAGCTG-3′) and F4-R (5′-ACCTGTAGCCCAATGTTTGAAGGACCGGTCTTG TACAGCTCGTCCATGC-3′), F5-F (5′-TTCAAACATTGGGCT ACAGGTGAGTGTACCTGAAGAAAGAACACAATTTACTT ACACTTACATTTCAAACAGAT-3′) and F5-R (5′-AAAAGTA GATTAGTTACGTAATAACTTCGTATAAAGTATCCTATACG AAGTTATGGATCCCTTTGGGTAGTTTACAATA-3′), F6-F (5′-TACGTAACTAATCTACTTTTCCTTCATCCCAAGTGGTTG GAAGCCAAACCTGCCAAAGATTAGTAAGCAGA-3′) and F6-R (5′-TAAGATATCAATATCCACTAAATGTCTAAAACTG-3′), F7-F (5′-TAGTGGATATTGATATCTTACTTGTACAGCTCGTC CA-3′) and F7-R (5′-ATGATAATATGGCCACAACCATGGTGA GCAAGGGCGAGGA-3′), and F8-F (5′-CTCACCATGGTTGTG GCCATATTATCATC-3′) and F8-R (5′-CATACATTATACGAAG TTATCTAGATAGCTGACTGAGCCCCTCTCCCTCCCCCCC CCCTAACGTTAC-3′).

The fused FlipTrap fragment was subsequently cloned into either the T2KXIGδ-in (Kawakami et al. 2004) or pMDS-eGFP (Emelyanov et al. 2006) at XhoI/BlgII and XhoI/SnaBI, respectively. The full sequence of the FlipTrap vector is available through NCBI (accession no. JN564735).

Microinjections and screening of FlipTrap lines

To generate FlipTrap lines, 2.3 nL of a stock solution containing 20 pg/nL FlipTrap construct and 20 pg/nL *tol2* or *Ac* mRNA was injected into the one-cell stage embryo. The injected F_0 embryos were raised and crossed to wild type for screening. The progeny, the F_1 embryos, were screened for Citrine expression with a fluorescent dissecting scope using a YFP filter (excitation: 490–500 nm; emission: 515–560 nm; dichromatic mirror: 505 nm). Citrine-positive embryos were counterstained with the vital stain bodipy TR methyl ester by incubating embryos in a 1:50 dilution of bodipy TR methyl ester for 1 h (Invitrogen). Subcellular localization of FlipTraps was determined by confocal microscopy. Citrine expression was excited with a 514-nm laser and detected with a 520- to 550-nm bandpass filter, while mCherry expression or bodipy TR methyl ester was excited with a 561-nm laser and detected with a 575- to 630-nm bandpass filter on a Zeiss LSM 510 laser-scanning confocal microscope.

RACE

For 5′ RACE, we adapted the GeneRACER kit from Invitrogen. Briefly, total RNA from a pool of 20–30 embryos per FlipTrap line was extracted with 100 μL of Trizol and dephosphorylated with 10 U of calf intestinal phosphatase (CIP). The mRNA cap structure was removed with 0.5 U of tobacco acid pyrophosphatase (TAP) before ligation with GeneRacer RNA oligo (5′-CGACUG GAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAA A-3′). cDNA was synthesized using a random hexamer primer or primers specific for citrine (5′-GAAGCAGCACGACTTCTTC A-3′). Nest PCR with primers to citrine and the GeneRacer sequence were used to amplify the 5′ end of the trapped gene. For first-round amplification, we used the citrine5RACE2 (5′-GCA GATGAACTTCAGGGTCA-3′) and GeneRacer5 (5′-CGACTG GAGCACGAGGACACTGA-3′) primers. For second-round amplification, we used the citrine5RACE (5′GACGTAAACGGCC ACAAGTTCAG-3′) and GeneRacer-nested (5′-GGACACTGA CATGGACTGAAGGAGTA-3′) primers. PCR conditions were

as follows for the first round of PCR: one cycle of 2 min at 94°C, followed by five cycles of 30 sec at 94°C and 3 min at 72°C; five cycles of 30 sec at 94°C and 3 min at 70°C; 20 cycles of 30 sec at 94°C, 30 sec at 65°C, and 3 min at 68°C; and one cycle of 10 minu at 72°C. PCR conditions were as follows for the second round of PCR: one cycle of 1 min at 94°C; 32 cycles of 30 sec at 94°C, 30 sec at 64°C, and 3 min at 68°C; and one cycle of 10 min at 72°C. PCR products were gel-purified and subcloned into pGEMT-easy for sequencing.

For 3′RACE, we extracted total RNA from single embryos with 50 μL of Trizol. cDNA was synthesized with reverse transcriptase and an oligo dT_{15} adaptor (5′-GTAATACGACTCACT ATAGGGCACGCGTGGTCGACGGCCCGGGCTGG$(T)_{15}$-3′). Nested PCR with primers to citrine and the adaptor were used to amplify the 3′ end trap gene. For first-round amplification, we used the citrine3RACE2 (5′-GACAACCACTACCTGAGCTAC C-3′) and AP1 (5′-GTAATACGACTCACTATAGGGC-3′) primers. For second-round amplification, we used the citrine3RACE (5′-ACATGGTCCTGCTGGAGTTC-3′) and AP2 (5′-ACTATAGGG CACGCGTGGT-3′) primers. PCR conditions were as follows for the first round of PCR: one cycle of 2 min at 94°C, followed by five cycles of 30 sec at 94°C and 3 min at 68°C; five cycles of 30 sec at 94°C and 3 min at 66°C; 20 cycles of 30 sec at 94°C, 30 sec at 64°C, and 3 min at 72°C; and one cycle of 10 min at 72°C. PCR conditions were as follows for the second round of PCR: one cycle of 1 min at 94°C; 30 cycles of 30 sec at 94°C, 30 sec at 68°C, and 3 min at 72°C; and one cycle of 10 min at 72°C. Parallel 3′ RACE was performed on three Citrine-positive and three wild-type siblings per FlipTrap line to reduce isolation of false-positive RACE products. Only PCR products that appear in all three reactions of Citrine embryos and not in the wild-type siblings were isolated for cloning into pGEMT-easy vector and sequenced.

RT-qPCR

Two-step qPCR was performed using ABI's SYBR Green RT–PCR system (Applied Biosystems) in accordance with the manufacturer's recommendations. Briefly, RNA was extracted from wild-type and homozygous embryos obtained from incrosses of F_1 heterozygous Cre-recombined adults using RNAqueous kit (Ambion) for cDNA synthesis with reverse transcriptase (SuperScript II RT, Invitrogen) using a random hexamer for priming. Cre-recombined adults carried the Cre-recombined allele in all cells, as only adults that contained the Cre mutant allele in their germline were used to maintain and propagate the mutant line. Embryos were sorted based on the level of mCherry expression. Homozygous embryos were twice as bright as heterozygous embryos. Reverse transcription reactions were diluted in series (onefold to 10,000-fold) and 1 μL was amplified in triplicate on a 7000 Sequence Detection System (Applied Biosystems); quantification was performed using the ΔΔCt method (Livak and Schmittgen 2001).

Determination of genomic integration

To determine the genomic integration site of the FlipTrap lines that trapped novel or hypothetical genes, we performed splinkerette PCR as previously described (Uren et al. 2009) with the following modifications. Briefly, genomic DNA was extracted from the tail of FlipTrap adults and digested with AluI or HaeIII (2 μg of genomic DNA per reaction). Splinkerette adaptors were ligated to the digested genomic DNA, and nested PCR was performed on the ligated DNA. The following primer pairs were used to amplify the 5′ end of the FlipTrap insertion: FT-5′-long-P1-F (5′-ATTAAACTGGGCATCAGCGCAATTC-3′) and SP-1 (5′-CG AAGAGTAACCGTTGCTAGGAGAGACC-3′), followed by FT-5′-long-P2-F (5′-CAAGGGAAAATAGAATGAAGTGATCTC CA-3′) and SP-2 (5′-GTGGCTGAATGAGACTGGTGTCGAC-3′). For amplification of the 3′ end of the FlipTrap insertion, FT-3′HaeIII-long-P1-R (5′-GCGTGTACTGGCATTAGATTG TCTGTC-3′) and FT-3′-AluI-long-P1-R (5′-CAAGAATCTCTAGTTTTCTTTCTTG CTT-3′) were used with SP-1, followed by FT-3′HaeIII-long-P2-R (5′-GCAGGATAAAACCTTGTA TGCATTTCATT-3′) and FT-3′-long-P2-R (5′-TTCTTTCTTGCTTTTACTTTTACTTCCTT-3′) with SP-2. For first-round PCR, conditions were as follows: one cycle of 3 min at 94°C; 29 cycles of 15 sec at 94°C, 30 sec at 62°C, and 3 min at 72°C; and one cycle of 5 min at 72°C. For second-round PCR, conditions were as follows: one cycle of 15 min at 94°C, and 25 cycles of 15 sec at 94°C, 30 sec at 57°C, 5 min at 72°C, and 5 min at 72°C.

Flp-FRT cassette exchange

For cassette exchange by Flp-FRT recombination, FlipTrap embryos were injected with 2.3 nL of a stock solution containing 25 pg/nL replacement construct and 80 pg/nL *flp* mRNA at the one-cell stage embryo. Injected embryos were raised at 28°C until the onset of gastrulation, at which time they were shifted to 25°C.

In situ hybridization and immunohistochemistry

In situ hybridization was performed as previously described in Thisse et al. (1993) with the following modifications. Embryos were not subjected to proteinase K treatment, but were instead incubated overnight in PBST (0.5% Tween-20) before hybridizing with antisense probe to permeate the embryo but avoid digestion of proteins by proteinase K. Following in situ hybridization, embryos were subjected to immunohistochemistry with an antibody to GFP to detect Citrine fusion protein expression. Antibody staining was performed in PBDT (1% BSA, 1% DMSO, 0.1% Triton X-100 in PBS at pH 7.3). We used the following primary antibodies: rabbit anti-β-catenin (1:500; Sigma), rabbit anti-GFP (1:500; Torrey Pines BioLabs), rabbit anti-α-catenin (1:200; Thermo Scientific), and mouse anti-LaminB (1:100; Affinity BioReagents). Alexa-fluor-conjugated antibodies from Invitrogen were used as secondaries.

Acknowledgments

We thank P. Pantazis for comments on the manuscript, and members of the Fraser laboratory for helpful discussions. We are grateful to LeighAnn Fletcher for fish care. The T2KXIGδ-in and pMDS-eGFP plasmids were provided by K. Kawakami and S. Parinov, respectively. This work was supported by NHGRI Center of Excellence in Genomic Science grant P50HG004071.

References

Amsterdam A, Nissen RM, Sun Z, Swindell E, Farrington S, Hopkins N. 2004. Identification of 315 genes essential for early zebrafish development. *Proc Natl Acad Sci* **101:** 12792–12797.

Anderson KV. 2000. Finding the genes that direct mammalian development: ENU mutagenesis in the mouse. *Trends Genet* **16:** 99–102.

Balciunas D, Wangensteen KJ, Wilber A, Bell J, Geurts A, Sivasubbu S, Wang X, Hackett PB, Largaespada DA, McIvor RS, et al. 2006. Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet* **2:** 1715–1724.

Berger J, Berger S, Hall TE, Lieschke GL, Currie PD. 2010. Dystrophin-deficient zebrafish feature aspects of the Duchenne

muscular dystrophy pathology. *Neuromuscul Disord* **12:** 826–832.

Branda CS, Dymecki SM. 2004. Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev Cell* **6:** 7–28.

Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, Skora AD, Nystul TG, Ohlstein B, Allen A, et al. 2007.The Carnegie protein trap library: a versatile tool for *Drosophila* developmental studies. *Genetics* **175:** 1505–1531.

Clark K, Balciunas D, Pogoda HM, Ding Y, Westcot SE, Bedell VM, Greenwood TM, Urban MD, Skuster KJ, Petzold AM, et al. 2011. In vivo protein trapping produces a functional expression codex of the vertebrate proteome. *Nat Methods* **8:** 506–515.

Clyne PJ, Brotman JS, Sweeney ST, Davis G. 2003. Green fluorescent protein tagging *Drosophila* proteins at their native genomic loci with small P elements. *Genetics* **165:** 1433–1441.

Collins RT, Linker C, Lewis J. 2010. MAZe: a tool for mosaic analysis of gene function in zebrafish. *Nat Methods* **7:** 219–223.

Cronin A, Mowbray S, Dürk H, Homburg S, Fleming I, Fisslthaler B, Oesch F, Arand M. 2002. The N-terminal domain of mammalian soluble epoxide hydrolase is a phosphatase. *Proc Natl Acad Sci* **100:** 1552–1557.

Doolitte RF. 1995. The multiplicity of domains in proteins. *Ann Rev* **64:** 287–314.

Doyle DA, Lewis LA, Kim J, Sheng E, Mac M, Kinnon R. 1996. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **85:** 1067–1076.

Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, Stainier DY, Zwartkruis F, Abdelilah S, Rangini Z, et al. 1994. A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123:** 37–46.

Emelyanov A, Gao Y, Naqvi NI, Parinov S. 2006. Trans-kingdom transposition of the maize dissociation element. *Genetics* **174:** 1095–1104.

Friedrich G, Soriano P. 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5:** 1513–1523.

Gavin A, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415:** 141–147.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature* **425:** 737–741.

Golic KG, Lindquist S. 1989. The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell* **59:** 499–509.

Gossler A, Joyner AL, Rossant J, Skarnes WC. 1989. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244:** 463–465.

Gudlaugsdottir S, Boswell DR, Wood GR, Ma J. 2007. Exon size distribution and the origin of introns. *Genetica* **131:** 299–306.

Hans S, Kaslin J, Freudenreich D, Brand M. 2009. Temporally-controlled site-specific recombination in zebrafish. *PLoS ONE* **4:** e4640. doi: 10.1371/journal.pone.0004640.

Heikal A, Hess ST, Baird GS, Tsien RY, Webb WW. 2000. Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: Coral red (dsRed) and yellow (Citrine). *Proc Natl Acad Sci* **97:** 11996–12001.

Hrabé de Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D, Marschall S, Heffner S, Pargent W, Wuensch K, Jung M, et al. 2000. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* **25:** 444–447.

Huang PL, Dawson TM, Bredt DS, Synder SH, Fishman MC. 1993. Targeted disruption of the neuronal nitric oxide synthase gene. *Cell* **75:** 1273–1286.

Huh W, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. *Nature* **425:** 686–690.

International Gene Trap Consortium. 2004. A public gene trap resource for mouse functional genomics. *Nat Genet* **36:** 543–544.

Kawakami K. 2007. Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol* **8:** S7. doi: 10.1186/gb-2007-8-S1-S7.

Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, Mishina M. 2004. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev Cell* **7:** 111–144.

Kerrebrock AW, Moore DP, Wu JS, Orr-Weaver TL. 1995. Mei-S332, a *Drosophila* protein required for sister-chromatid cohesion, can localize to meiotic centromere regions. *Cell* **83:** 247–256.

Kim Y, Yakunin AF, Kuznetsova E, Xu X, Pennycooke M, Gu J, Cheung F, Proudfoot M, Arrowsmith CH, Joachimiak A, et al. 2004. Structure- and function-based characterization of a new phosphoglycolate phosphatase from *Thermoplasma acidophilum*. *J Biol Chem* **279:** 517–526.

Lakso M, Sauer B, Mosinger B Jr, Lee EJ, Manning RW, Yu SH, Mulder KL, Westphal H. 1992. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc Natl Acad Sci* **89:** 6232–6236.

Le X, Langenau DM, Keefe MD, Kutok JL, Neuberg DS, Zon LI. 2007. Heat shock-inducible Cre/Lox approaches to induce diverse types of tumors and hyperplasia in transgenic zebrafish. *Proc Natl Acad Sci* **104:** 9410–9415.

Lippincott-Schwartz J, Cole N, Presley J. 1998. Unravelling golgi membrane traffic with green fluorescent protein. *Trends Cell Biol* **8:** 16–20.

Liu M, Grigoriev A. 2004. Protein domains correlate strongly with exons in multiple eukaryotic genomes-evidence of exon shuffling. *Trends Genet* **20:** 399–403.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $22^{-\Delta\Delta CT}$ method. *Methods* **25:** 402–408.

Luchi S. 2001. Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci* **58:** 625–635.

Morin X, Daneman R, Zavortink M, Chia W. 2001. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc Natl Acad Sci* **98:** 15050–15055.

Mullins MC, Hammerschmidt M, Haffter P, Nusslein-Volhard C. 1994. Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol* **4:** 189–202.

Munro S, Pelham HRB. 1987. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48:** 899–907.

Orban PC, Chui D, Marth JD. 1992. Tissue- and site-specific DNA recombination in transgenic mice. *Proc Natl Acad Sci* **89:** 6861–6865.

Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. 1996. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273:** 1392–1395.

Pan X, Li Y, Stein L. 2005. Site preferences of insertional mutagenesis agents in *Arabidopsis*. *Plant Physiol* **137:** 168–175.

Pawson T. 1995. Protein modules and signalling networks. *Nature* **373:** 573–580.

Rijli FM, Mark M, Lakkaraju S, Dierich A, Dollé P, Chambon P. 1993. A homeotic transformation is generated in the rostral branchial region of the head by disruption of Hoxa2, which acts as a selector gene. *Cell* **75:** 1333–1349.

Rosahl TW, Geppert M, Spillane D, Herz J, Hammer RE, Malenka RC, Südhof TC. 1993. Short-term synaptic plasticity is altered in mice lacking synapsin I. *Cell* **75:** 661–670.

Sato Y, Kasai T, Nakagawa S, Tanabe K, Watanabe T, Kawakami K, Takahashi Y. 2007. Stable integration and conditional expression of electroporated transgenes in chicken embryos. *Dev Biol* **305:** 616–624.

Sauer B, Henderson N. 1998. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci* **85:** 5166–5170.

Schnütgen F, Doerflinger N, Calléja C, Wendling O, Chambon P, Ghyselinck NB. 2003. A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat Biotechnol* **21:** 562–565.

Senecoff JF, Bruckner RC, Cox MM. 1985. The FLP recombinase of the yeast 2-um plasmid: characterization of its recombination site. *Proc Natl Acad Sci* **82:** 7270–7274.

Skarnes WC, Auerbach BA, Joyner AL. 1992. A gene trap approach in mouse embryonic stem cells: the lacZ reporter is activated by splicing, reflects endogenous expression, and is mutagenic in mice. *Genes Dev* **6:** 903–918.

Thisse C, Thisse B, Schilling TF, Postlethwait JH. 1993. Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. *Development* **119:** 1203–1215.

Urasaki A, Mito T, Noji S, Ueda R, Kawakami K. 2008. Transposition of the vertebrate *Tol2* transposable element in *Drosophila melanogaster*. *Gene* **425:** 64–68.

Uren AG, Mikkers H, Kool J, van der Weyden L, Lund AH, Wilson CH, Rance R, Jonkers J, van Lohuizen M, Berns A, et al. 2009. A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nat Protoc* **4:** 789–798.

Wainwright SM, Ish-Horowicz D. 1992. Point mutations in the *Drosophila* hairy gene demonstrate in vivo requirements for basic, helix–loop–helix, and WRPW domains. *Mol Cell Biol* **12:** 2475–2483.

Weil CF, Kunze R. 2000. Transposition of maize Ac/Ds transposable elements in the yeast *Saccharomyces cerevisiae*. *Nat Genet* **26:** 187–190.

Westerfield M. 1994. *The zebrafish book*. University of Oregon Press, Eugene, OR.

Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA. 2005. High-resolution genome-wide mapping of transposon integration in mammals. *Mol Cell Biol* **25:** 2085–2094.

Zhang, XHF, Chasin LA. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* **18:** 1241–1250.

A versatile gene trap to visualize and interrogate the function of the vertebrate proteome

Le A. Trinh, Tatiana Hochgreb, Matthew Graham, et al.

Genes Dev. 2011, **25:**
Access the most recent version at doi:10.1101/gad.174037.111