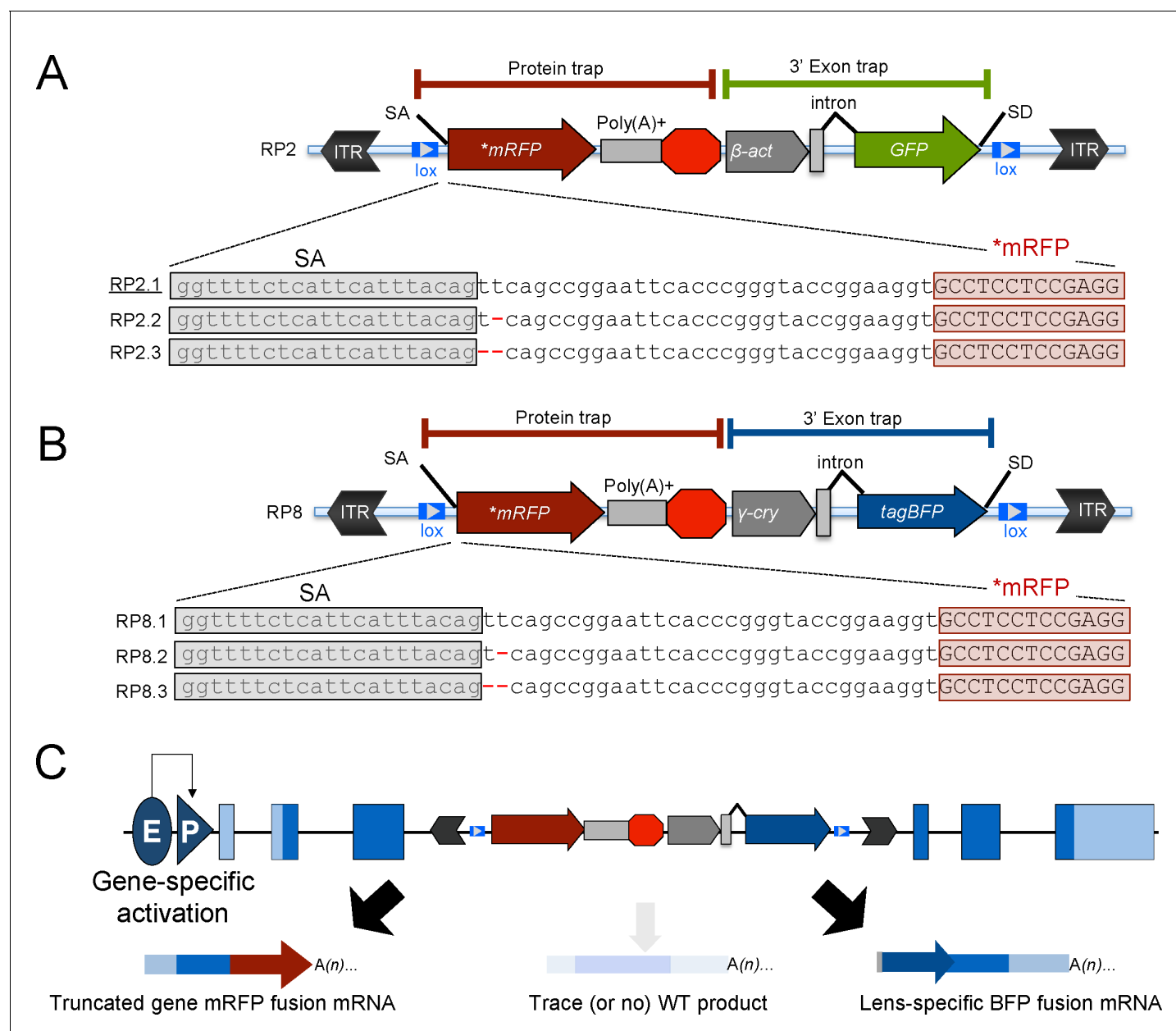


---

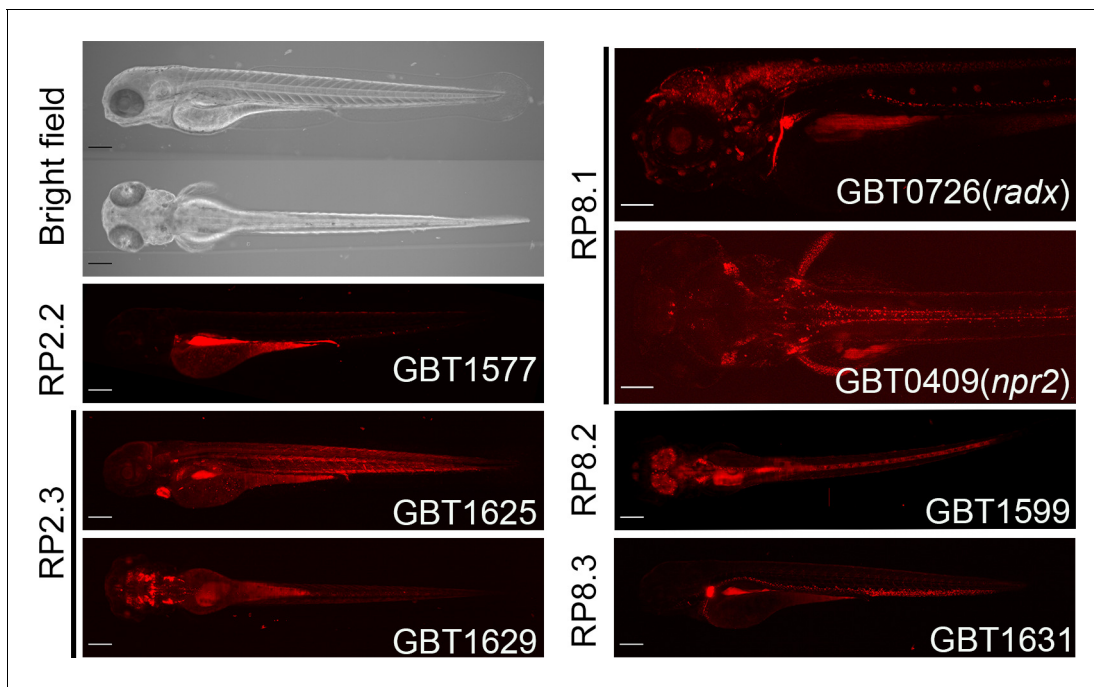
## Figures and figure supplements

Building the vertebrate codex using the gene breaking protein trap library

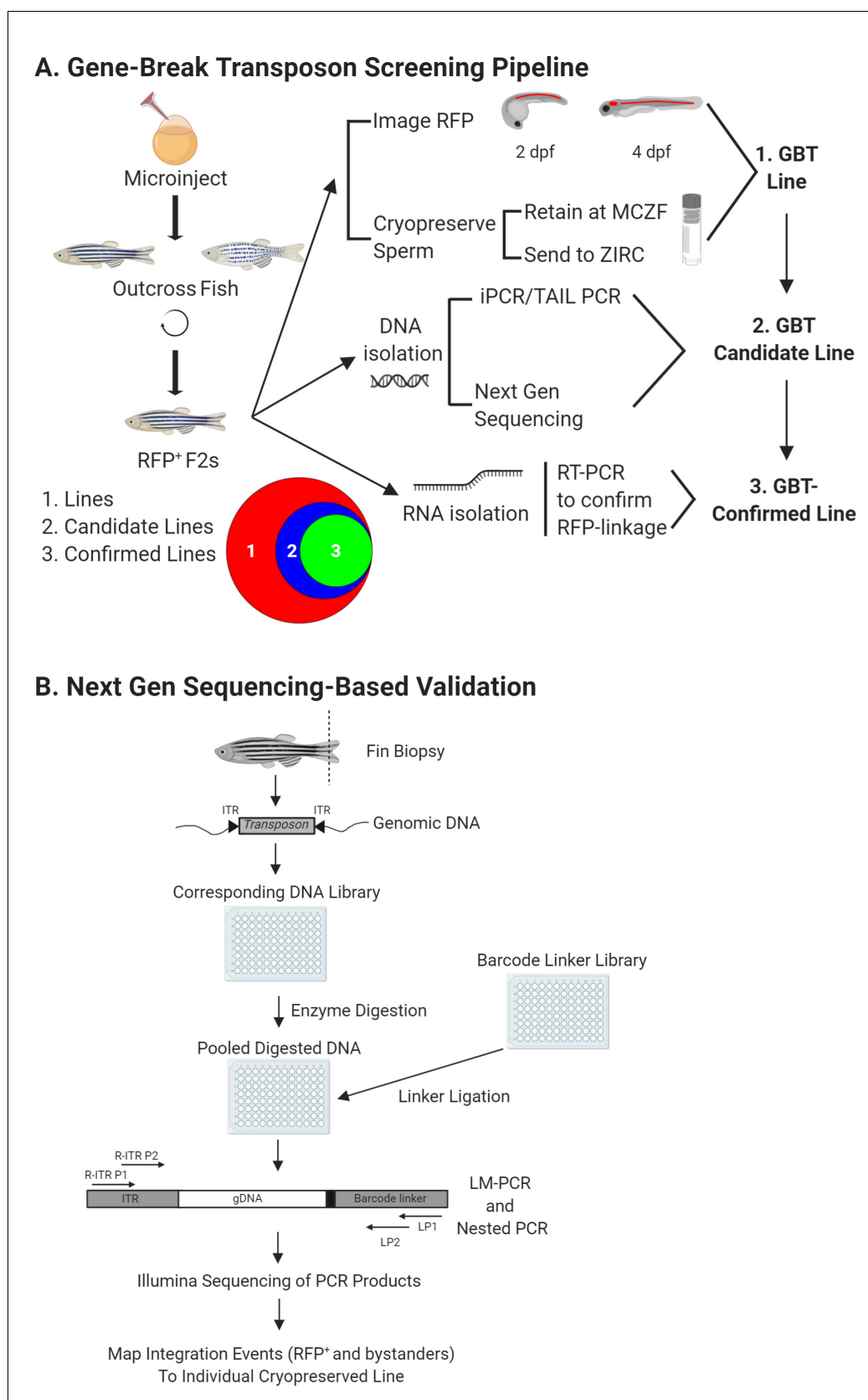
**Noriko Ichino et al**



**Figure 1.** Schematic of the RP2 and RP8 gene-break transposon (GBT) system with all three reading frames of AUG-less mRFP reporter. (A–C) Schematic of the GBT system, RP2 and RP8 incorporate a protein-trap cassette fused with three reading frames of AUG-less mRFP reporter and a 3' exon trap cassette with GFP or tagBFP reporters, respectively. (A) RP2 series (RP2.1, RP2.2 and RP2.3). Underline: Previously published vector construct (B–C) RP8 series (RP8.1, RP8.2 and RP8.3) with a schematic RP8 insertion event showing expected transcription off of a locus below (C). ITR: inverted terminal repeat, SA: splice acceptor, lox: Cre recombinase recognition sequence, \*mRFP: AUG-less mRFP sequence, poly (A)+: polyadenylation signal, red octagon: extra transcriptional terminator and putative border element,  $\beta$ -act: carp beta-actin enhancer,  $\gamma$ -cry: gamma crystalline promoter, SD: splice donor, E: enhancer, P: promoter, and WT: wild-type.



**Figure 1—figure supplement 1.** Representative expression patterns of mRFP fusion protein integrated all reading frames of RP2 and RP8. Lateral and dorsal views of representative bright field images at four dpf and lateral or dorsal views of RFP expression patterns at four dpf in GBT1577 integrated RP2.2, GBT1625 and GBT1629 integrated RP2.3, GBT0409 (*npr2*) and GBT0726 (*radx*) integrated RP8.1, GBT1599 integrated RP8.2 and GBT1631 integrated RP8.3. Scale bars = 200  $\mu$ m.

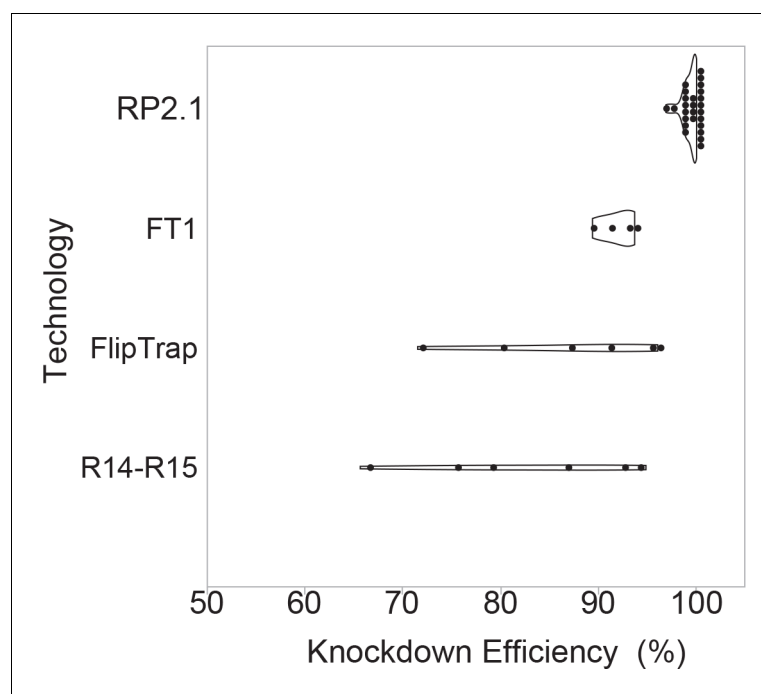


**Figure 2.** GBT screening pipeline. (A) Overview of GBT screening pipeline. Wild-type embryos at 1 cell were co-injected with RP plasmid and Tol2 transposase mRNA to create F0 founders. These F0 larvae were screened for

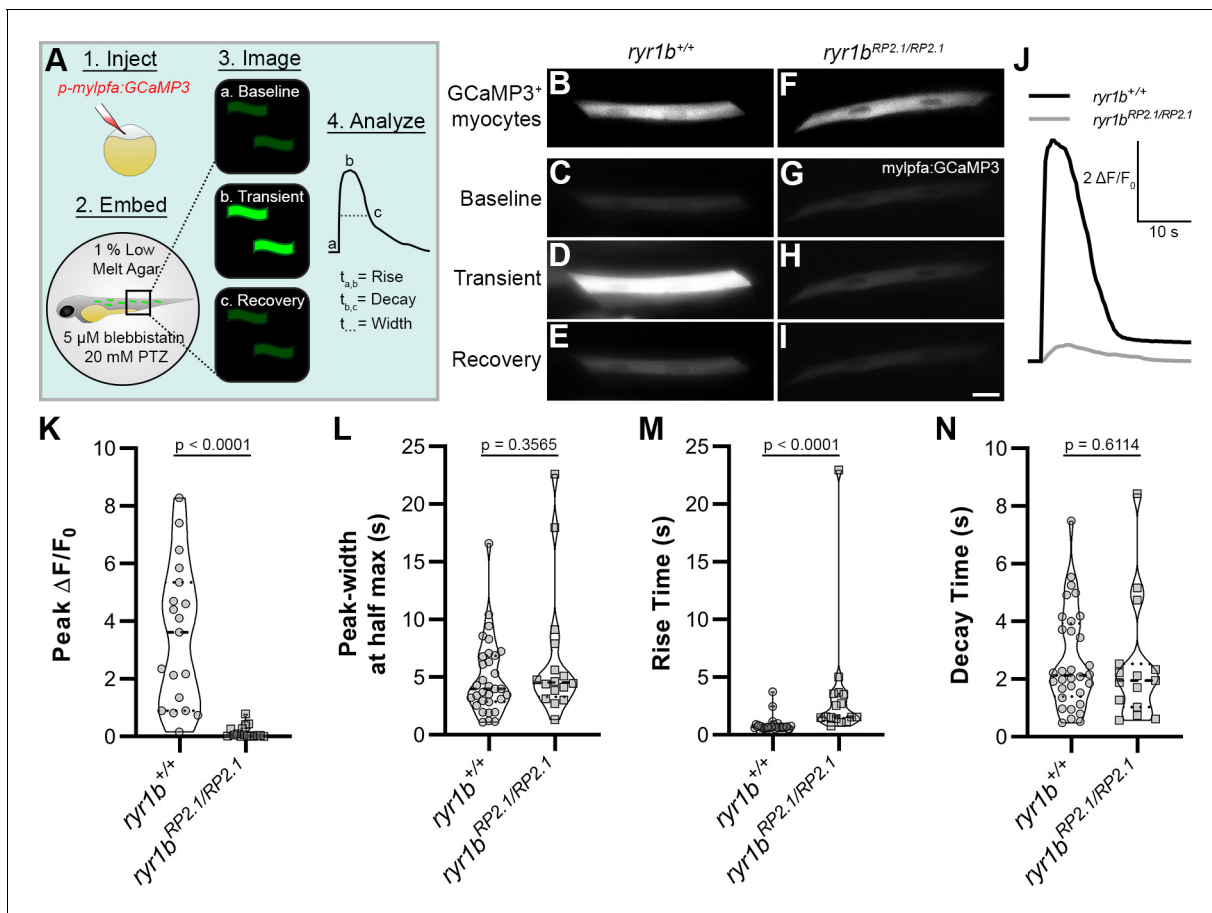
*Figure 2 continued on next page*

*Figure 2 continued*

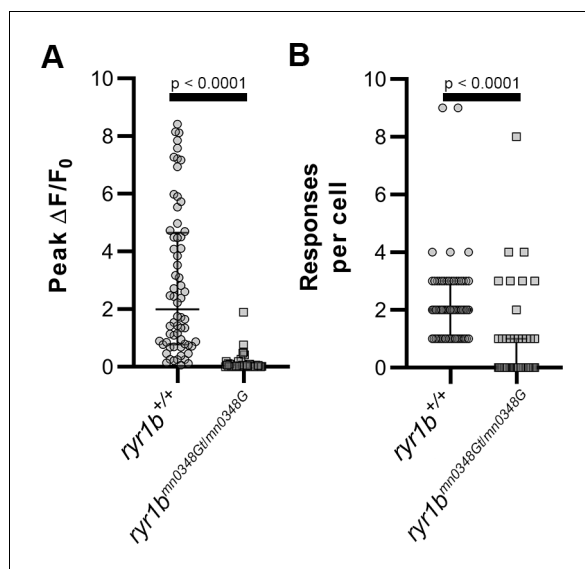
non-mosaic RP expression, raised, and outcrossed for two generations. Then, mRFP<sup>+</sup> F2 heterozygous larvae were 3-dimensionally imaged at 2 and 4 dpf and this imaging data were uploaded to zfishbook (<http://www.zfishbook.org/>). Sperm from four F2 males in over 1200 robust mRFP expressing lines were cryopreserved using the Zebrafish International Resource Center (ZIRC) standard protocol and stored at both ZIRC and Mayo Clinic Zebrafish Core Facility (MCZF). DNA and RNA isolated from these four F2 males with cryopreserved sperm was utilized to perform next-generation sequencing and to confirm RFP linkage of candidate lines by manual PCRs (iPCR, TAIL-PCR, 5' RACE and 3' RACE). Venn diagram illustrates current library of over 1,200 GBT lines with 204 GBT-confirmed lines out of 348 molecularly analyzed GBT-candidate lines. **(B)** Next generation sequencing based validation for GBT integration loci. Fin biopsies from four F2 males were utilized as DNA source for the validation process to identify GBT integration loci. Extracted genomic DNA was fragmented, pooled in 96-wells plate, and ligated with barcode linker to identify each single male with cryopreserved sperm. Linker-mediated (LM) PCR with the primers, R-ITR P1 and LP1 and nested PCR with the primers, R-ITR P2 and LP2 were conducted to perform Illumina sequencing the final PCR products. The integration events of individual sperm-cryopreserved male were mapped on zebrafish reference genome sequence with bioinformatics analysis. This figure was created with BioRender.com. The area proportional Venn diagram was produced using BioVenn (<http://www.biovenn.nl/>).



**Figure 3.** Knockdown efficiency of RP2.1 compared with previous gene-trap systems. Violin plots comparing percent knockdown efficiency in the analyzed individual lines generated by four protein trap systems. All plots show median. The data of previous protein trap systems were converted from the data in the original articles, R14-R15, our initial R-series protein trap vectors (n = 6), (Clark et al., 2011a; FlipTrap, FlipTrap vectors (n = 6), Trinh et al., 2011; FT1, FT1 vector (n = 4), Ni et al., 2012; RP2.1, RP2.1 vector (n = 26), Clark et al., 2011a; Ding et al., 2013; Ding et al., 2017; El-Rass et al., 2017; Westcot et al., 2015 and unpublished data) (Figure 3—source data 1). The graph was made in JMP14 (SAS, Cary, NC).

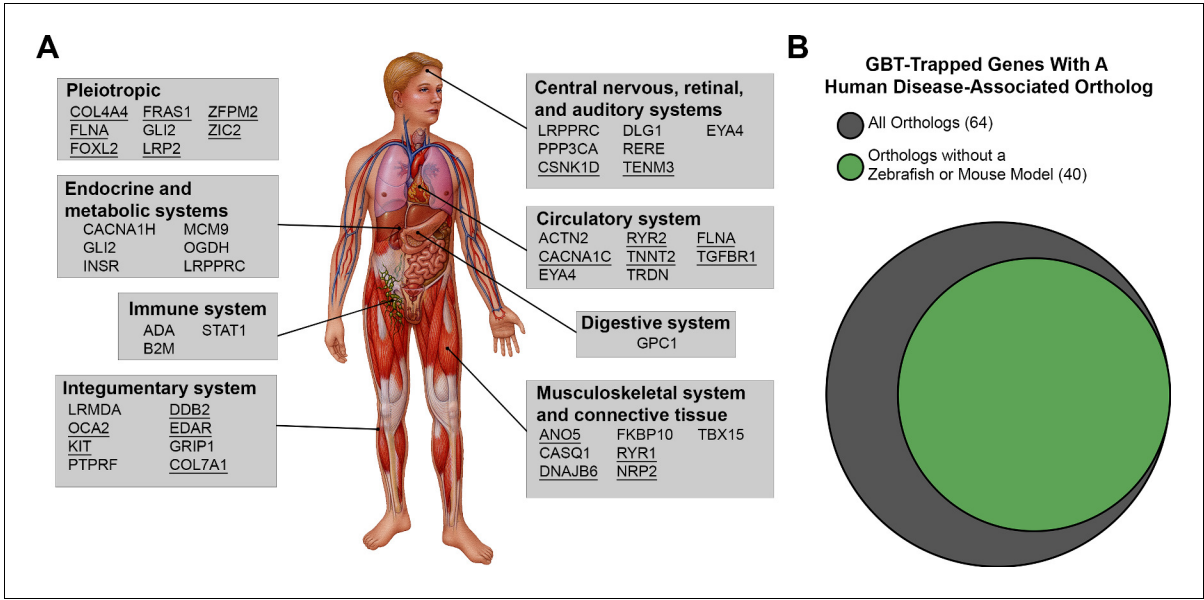


**Figure 4.** GBT demonstrates that neural disinhibition mediated  $\text{Ca}^{2+}$  transients in *mylpfa*<sup>+</sup> myocytes require the ryanodine receptor *ryr1b* in vivo. (A) Cartoon showing approach to assay  $\text{Ca}^{2+}$  transients in zebrafish myocytes through (1) injection of *p-mylpfa:GCaMP3* (Baxendale et al., 2012) at the single cell stage, (2) embedding in 1% low melt agar/20 mM pentylenetetrazole (PTZ)/5  $\mu$ M (S)-(-)-blebbistatin, (3) imaging for 3 min to record transient-associated changes in myocyte GCaMP3 fluorescence at 2 days post-fertilization, and (4)  $\text{Ca}^{2+}$  transient analysis. (B–I) Static images of GCaMP3 expressing myocytes (B, F) and representative GCaMP3 time-series images showing baseline (C, G), transient peak (D, H), and recovery (E, I) in *ryr1b*<sup>+/+</sup> (C–E) and *ryr1b*<sup>mn0348Gt/mn0348Gt</sup> (G–I) animals, respectively. Scale bar = 20  $\mu$ m. (J) Representative  $\Delta F/F_0$  traces of  $\text{Ca}^{2+}$  transients from *ryr1b*<sup>+/+</sup> (black) and *ryr1b*<sup>mn0348Gt/mn0348Gt</sup> (gray) myocytes. (K–N) Violin plots comparing transient peak  $\Delta F/F_0$  (averaged within fish) (K),  $\text{Ca}^{2+}$  transient peak-width (L),  $\text{Ca}^{2+}$  transient rise (M) and decay (N) time between *ryr1b*<sup>+/+</sup> and *ryr1b*<sup>mn0348Gt/mn0348Gt</sup> animals. All plots show median with interquartile range. For (K)  $n_{\text{ryr1b}+/+} = 19$  animals,  $n_{\text{ryr1bmn0348Gt/mn0348Gt}} = 16$  animals. For (L–M)  $n_{\text{ryr1b}+/+} = 32$  cells,  $n_{\text{ryr1bmn0348Gt/mn0348Gt}} = 16$  cells. For (N)  $n_{\text{ryr1b}+/+} = 32$  cells,  $n_{\text{ryr1bmn0348Gt/mn0348Gt}} = 15$  cells. Data are compiled from four independent experiments containing at least two animals in each group. p-values determined using the Mann-Whitney U test. Effect size (Cohen's d)=1.829 (K) and 0.866 (M). Source data can be found in **Figure 4—source data 1** (K, L, M, N) and **Figure 4—source data 2** (J).

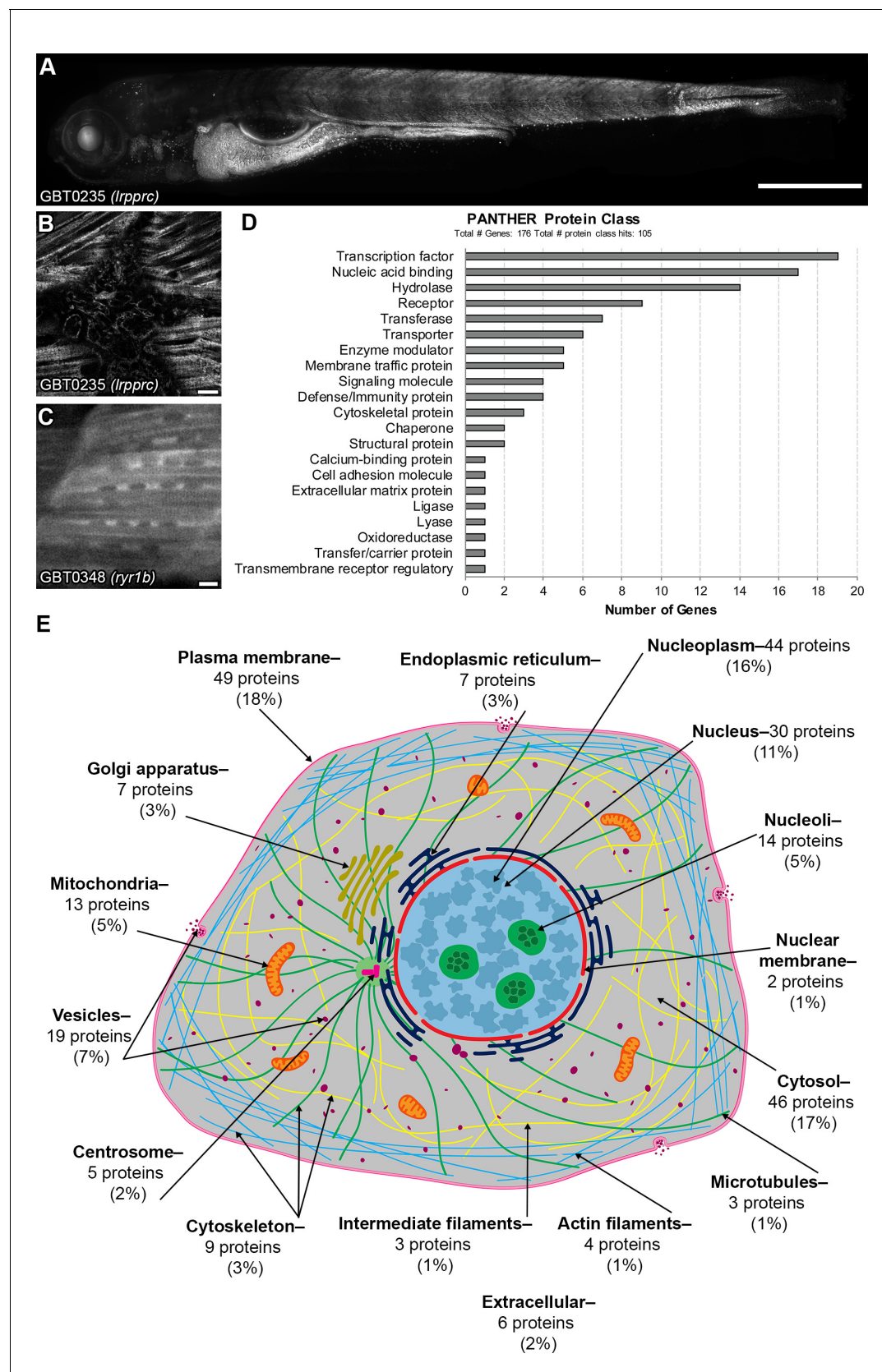


**Figure 4—figure supplement 1.**  $\text{Ca}^{2+}$  transients in *ryr1b*<sup>+/+</sup> myocytes have higher peak amplitude and are more frequent than in *ryr1b*<sup>mn0348Gt/mn0348Gt</sup> myocytes. (A) Dot plot comparing peak  $\Delta F/F_0$  responses (averaged within cell) between *ryr1b*<sup>+/+</sup> and *ryr1b*<sup>mn0348Gt/mn0348Gt</sup> animals. (B) Dot plot representing the number of responses per cell ( $\geq 0.05 \Delta F/F_0$ ) recorded during the 3 min imaging window in *ryr1b*<sup>+/+</sup> and *ryr1b*<sup>mn0348Gt/mn0348Gt</sup> animals. Plots show median with interquartile range.  $n_{\text{ryr1b}+/+} = 64$  cells,  $n_{\text{ryr1bmn0348Gt/mn0348Gt}} = 48$  cells. Data were compiled from four independent experiments containing at least two animals in each group. p-values determined using the Mann-Whitney U test. Effect size (Cohen's d)=1.445 (A) and 0.931 (B). Source data can be found in **Figure 4—figure supplement 1—source data 1**.

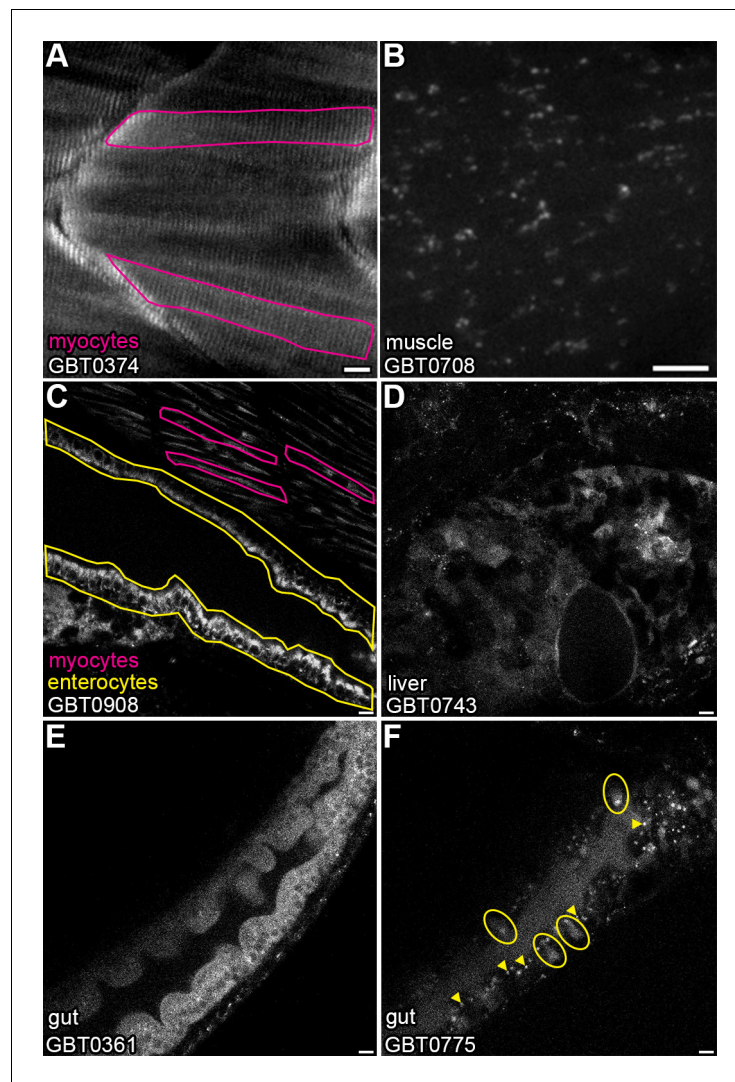




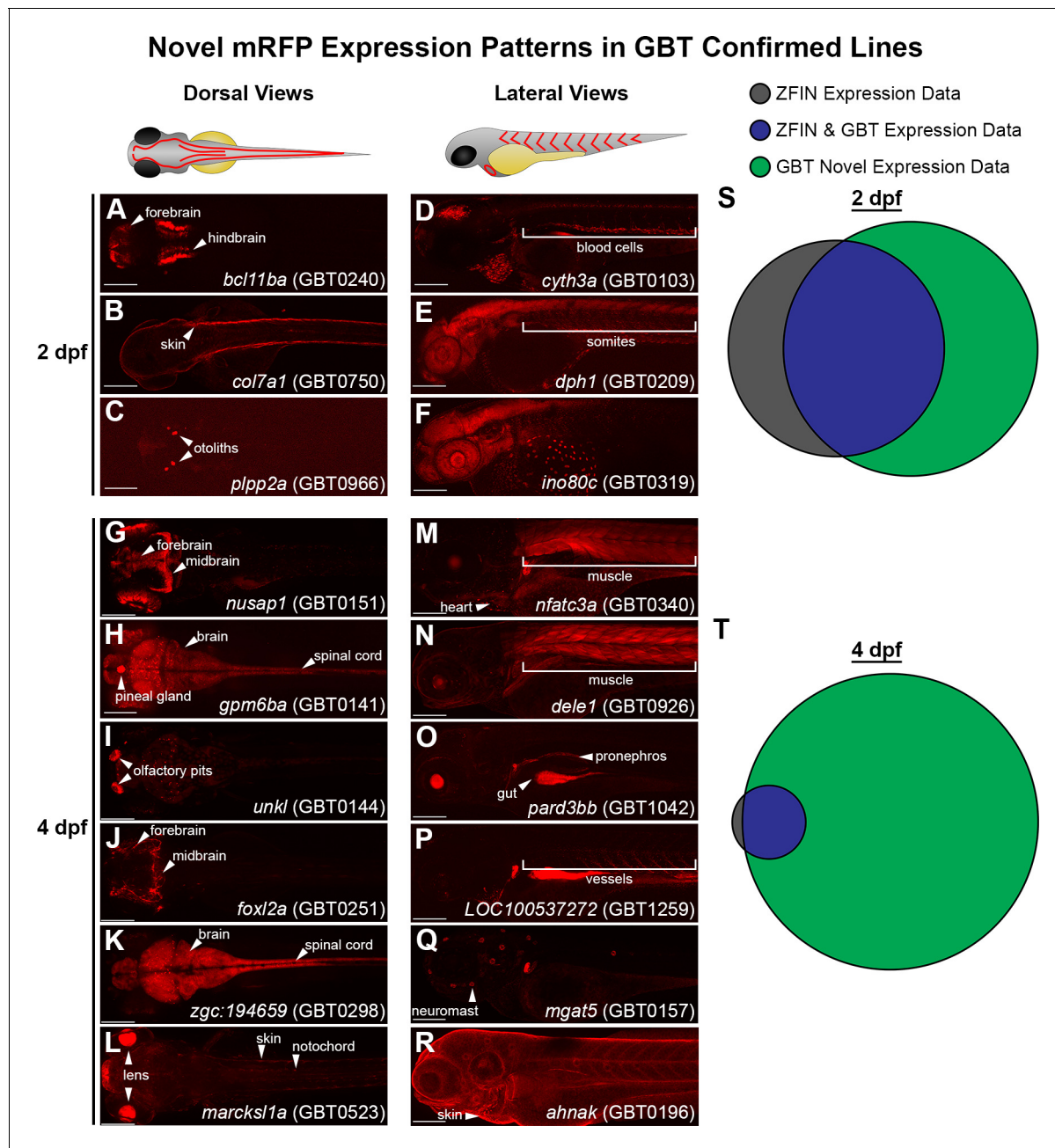
**Figure 5.** Disease-associated human orthologs of the GBT trapped genes are implicated in human genetic disorders of multiple organ systems. **(A)** Representative human orthologs of the GBT-tagged genes are associated with genetic disorders in multi-organ systems. Image provided by Mayo Clinic Media Services. Underline: Disease causative genes with documentations of established disease model in mouse or zebrafish **(B)** Area proportional Venn diagram of 64 human orthologs tagged that are associated with human genetic disorders. 40 human orthologs of GBT-tagged genes are associated with human genetic disorders without an established disease model in zebrafish or mouse. Area proportional Venn diagram was produced using BioVenn (<http://www.biovenn.nl/>).



**Figure 6.** GBT-confirmed lines illuminate and disrupt genes encoding proteins with diverse functions and subcellular localizations.



**Figure 6—figure supplement 1.** GBT protein traps illuminate diverse subcellular protein localizations. (A–B) Confocal images demonstrating patterns of subcellular localization seen in muscle with strong banding in GBT0374 (candidate gene = unannotated transcript) (A) and large, diffuse puncta in GBT0708 (candidate gene = *adam15*) (B). Scale bars = 10  $\mu$ m. (C) Confocal image of GBT0908 with ubiquitous, cytoplasmic expression. Note apical enrichment in enterocytes. (D) Confocal image of GBT0743 (candidate gene = *pole4*) with variegated expression in the liver. (E–F) Confocal images of gut expression with cytoplasmic, pan-enterocyte labeling in GBT0361 (E) in contrast with endomembrane puncta and enrichment of mRFP signal in a subset of enterocytes in GBT0775 (candidate gene = *cd83*) (F).



**Figure 7.** GBT protein trap elucidates novel gene expression patterns in embryonic and larval zebrafish. (A–C) Dorsal views of 2 days post-fertilization (dpf) embryos with GBT protein trap mRFP expression patterns ranging from *bcl11ba* in the forebrain and hindbrain (A), to *col7a1* in the skin (B), and *plpp2a* in the otoliths (C). (D–F) Lateral views of 2 dpf embryos with GBT protein trap mRFP expression patterns ranging from *cyth3a* in blood cells (D), to *dph1* in somites (E), and *ino80c* around the yolk (F). (G–L) Dorsal views of GBT protein trap mRFP expression patterns in 4 dpf larvae including *nusap1* in the forebrain and midbrain (G), *gpm6ba* in the brain, spinal cord, and pineal gland (H), *unkl* in the olfactory pits (I), *foxl2a* in the forebrain and midbrain (J), *zgc:194659* in the brain and spinal cord (K), and *marcks1a* in the lens, skin, and notochord (L). (M–R) Lateral views of GBT protein trap mRFP expression patterns in 4 dpf larvae including *nfatc3a* in heart and muscle (M), *dele1* in muscle (N), *pard3bb* in the gut and pronephros (O), *LOC100537272* in vessels (P), *mgat5* in neuromasts (Q), and *ahnak* in skin (R). Scale bars = 200  $\mu$ m. (S–T) Area proportional Venn diagrams of 193 genes trapped in GBT-confirmed lines comparing the ZFIN-assembled database with mRFP expression in GBT lines available through zfishbook at two dpf (S) and four dpf (T). 67 (35%) and 174 (90%) of 193 genes trapped in GBT-confirmed lines have no description about wild-type expression at 2 dpf and 4 dpf, respectively.