Developmental Neurobiology
DOI 10.1002/dneu.22049

Title: New transgenic reporters identify somatosensory neuron subtypes in larval zebrafish

Running Title: Larval zebrafish somatosensory neuron subtypes

Keywords: zebrafish, somatosensation, trigeminal, Rohon-Beard, peripheral axon, central axon, transgenic reporter

Authors: Ana Marie S. Palanca[1], Sung-Ling Lee[1], Laura E. Yee[1,2], Carlee Joe-Wong[1,3], Le A. Trinh[4], Elizabeth Hiroyasu[1], Majid Husain[1], Scott E. Fraser[4], Matteo Pellegrini[1] and Alvaro Sagasti[1]

Affiliations

1. Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, 90095, USA
2. Current Address: Department of Biochemistry and Biophysics, University of California, San Francisco 94158, USA
3. Current Address: Applied and Computational Mathematics, Princeton University, Princeton, NJ, 08540 USA
4. Division of Biology, California Institute of Technology, Pasadena, 91125, USA

Contact Information
Corresponding Author: Alvaro Sagasti
Email address: sagasti@mcdb.ucla.edu
Phone number: 310-206-6147
Fax number: 310-206-3987

Funding: Funding was provided by an NRSA award (5F31NS064817-03) to AMSP from the NINDS and grants from the NSF (RIG:0819010) and NIDCR (5R01DE018496) to AS.

Received: Jul 02, 2012; Revised: Jul 26, 2012; Accepted: Jul 26, 2012

Abstract

To analyze somatosensory neuron diversity in larval zebrafish, we identified several enhancers from the zebrafish and pufferfish genomes and used them to create five new reporter transgenes. Sequential deletions of three of these enhancers identified small sequence elements sufficient to drive expression in zebrafish trigeminal and Rohon-Beard (RB) neurons. One of these reporters, using the *Fru.p2x3-2* enhancer, highlighted a somatosensory neuron subtype that expressed both the *p2rx3a* and *pkcα* genes. Comparison with a previously described *trpA1b* reporter revealed that it highlighted the same neurons as the *Fru.p2x3-2* reporter. To determine whether neurons of this subtype possess characteristic peripheral branching morphologies or central axon projection patterns, we analyzed the morphology of single neurons. Surprisingly, although these analyses revealed diversity in peripheral axon branching and central axon projection, PKCα/*p2rx3a*/*trpA1b*-expressing RB cells did not possess obvious characteristic morphological features, suggesting that even within this molecularly defined subtype, individual neurons may possess distinct properties. The new transgenes created in this study will be powerful tools for further characterizing the molecular, morphological, and developmental diversity of larval somatosensory neurons.

Introduction

Somatosensation is carried out by a variety of specialized populations of sensory neurons that detect different types of thermal, chemical and mechanical touch stimuli (Lumpkin and Caterina, 2007; Marmigère and Ernfors, 2007). Innervation of the skin by somatosensory neurons occurs very early in development, allowing animals to detect touch even at embryonic stages (Davies and Lumsden, 1984; Moore and Munger, 1989; Kimmel et al., 1990; Saint-Amant

and Drapeau, 1998; Sagasti et al., 2005). The degree of heterogeneity among somatosensory neurons at early embryonic and larval stages has not been fully characterized, but distinct types of neurons can be distinguished by both function and axon morphology. For example, *Xenopus* trigeminal neurons that innervate the head, fall into two functionally, anatomically and physiologically distinct types: Type I "movement" detectors inhibit swimming behavior and their peripheral arbors have many fine branches with numerous large varicosities, whereas Type II "rapid transient" detectors stimulate behavior and their peripheral arbors have few relatively straight branches with elongated varicosities (Roberts, 1980; Hayes and Roberts, 1983).

Zebrafish are an ideal model for studying early stages of somatosensory neuron development and function due to their external fertilization, rapid development, and optical clarity, but the heterogeneity of their embryonic and larval somatosensory neurons have not been well characterized. Zebrafish possess three populations of somatosensory neurons: trigeminal neurons that innervate the head, Rohon-Beard (RB) neurons that innervate the body at early larval stages, and dorsal root ganglia (DRG) neurons that innervate the body at later stages. Several observations suggest that embryonic and larval zebrafish possess multiple subtypes of somatosensory neurons. First, the cutaneous axons of trigeminal and RB neurons display a broad spectrum of peripheral axon branching patterns, which might reflect multiple subclasses with distinct morphologies (Sagasti et al., 2005). Second, populations of trigeminal peripheral axon arbors appear to "tile" the skin independently of one another: Although the territories of all arbors are limited by repulsion, individual arbors only repel a subset of others (Sagasti et al., 2005). Third, distinct patterns of varicosities can be distinguished in the central axons of individual classes of trigeminal neurons, likely reflecting distinct synaptic patterns (Pan et al., 2012). Fourth, electrophysiological and molecular analyses have revealed that sodium currents

in different populations of RB neurons rely on different sodium channels (Pineda et al., 2006). Finally, several genes, such as *protein kinase C alpha* (*pkcα*) (Slatter et al., 2005; Patten et al., 2007), the ATP-gated ion channel *p2rx3a* (Boué-Grabot et al., 2000; Norton et al., 2000; Appelbaum et al., 2007) and the chemosensory ion channel *trpA1b* (Caron et al., 2008; Pan et al., 2012) are expressed in subpopulations of zebrafish trigeminal and RB neurons. Nonetheless, co-expression between most of these molecules has largely not been examined, so these observations have yet to cohere into a clear picture of the diversity among these neurons.

Fluorescent transgenic reporters have made it possible to study somatosensory neurons in live zebrafish larvae. The most commonly used reporters for these neurons utilize enhancers cloned from genomic regions near the *islet1* and *islet2b* genes to drive expression of a fluorescent protein (Higashijima et al., 2000; Uemura et al., 2005; Pittman et al., 2008). These enhancers have generally been thought to drive expression in all zebrafish somatosensory neurons, although one report suggests that a transgenic line using an *islet1* enhancer may mark a subset of neurons (Pan et al., 2012). While these reporters have been useful for characterizing sensory neuron development, their presumed pan-neuronal labeling limits their utility for studying sensory neuron subtypes. Furthermore, several of these reporters use the Gal4-VP16/UAS system, which amplifies expression and allows combinatorial versatility (Köster and Fraser, 2001), but also attracts methylation (Goll et al., 2009), leading to variegated expression. Fluorescence expression in subsets of somatosensory neurons has been achieved with reporters for *trpA1b* and *p2rx3b* that were made by recombining fluorescent proteins into bacterial artificial chromosomes (BACs) (Kucenas et al., 2006; Pan et al., 2012), but such reporters are rare, can be laborious to create, and their relation to other reporters has yet to be fully assessed.

To initiate a systematic analysis of subtype-specific transgenes, we have created several reporters using enhancers from neurotrophin receptors and ion channels. Defining enhancer sequences that control expression of these reporters could help identify transcriptional pathways regulating somatosensory neuron development. Subtype-specific transgenic reporters will also be useful tools for characterizing somatosensory neuron diversity. These new fluorescent reporters will make it possible to test whether different axon morphologies are optimized for specific somatosensory functions and to characterize neural circuitry controlling behavioral responses to touch stimuli.

Materials and Methods

Fish strains and transgenic lines

Embryos were raised at 28.5°C on a 14 hr/10 hr light/dark cycle. All experiments were approved by the Chancellor's Animal Research Care Committee at the University of California, Los Angeles.

Expression analysis of transgenes was performed in wild-type AB or *Gt(T2KSAG)*^{j1229a} (Burgess et al., 2009) fish. Stable transgenic lines were created with the Tol2 transposase system (Kawakami, 2004). Supplementary Table 1 summarizes all the transgenic lines used in this study, as well as the transgenes used for transient transgenesis. Although transgenic lines were not mapped, expression from all lines segregated in a Mendelian manner, indicating that they integrated into single genomic loci.

Cloning/ Transgenes

Enhancer regions of somatosensory neuron specific genes were amplified from genomic DNA of the zebrafish *Danio rerio* or the pufferfish *Fugu rubripes* by PCR (Table 1). These enhancer regions were initially cloned into Gal4:GFP-pBSK vectors by PCR amplification using primers with restriction sites incorporated into the flanking sequences. Subsequent subcloning and promoter dissection of these elements was performed using the Muliti-Site Gateway Cloning System (Invitrogen, 12537-023) in combination with the Tol2 Gateway System developed by the Chien Lab (Kwan et al., 2007). Briefly, genomic sequences were PCR amplified with primers containing attB sites and recombined into pDONR P4-P1R, creating 5' DONR plasmids. The binomial elements, E1b:Gal4-VP16:pA,14XUAS:E1b and E1b:LexA-VP16:pA,4xLexAop:E1b were cloned into pDONR 221 to make middle element vectors. Reporter genes, EGFP, mCherry and KikGR were cloned into pDONR P2R-P3 to make 3' elements. Reporter transgenes were created by recombination of different sets of pDONR elements using LR Clonase II Plus (Invitrogen, 12538120). Reporter function of each transgene was tested by injection into wildtype embryos. PCR primers used for amplifying genomic fragments are listed in Supplementary Table 2.

Transient analysis of reporter transgenes and confocal imaging

Zebrafish embryos were injected at the 1-cell stage with approximately 2 nl of 50 ng/ul plasmid DNA, raised in a 28.5°C incubator and treated with phenylthiourea (PTU) at 24 hpf to block pigmentation. Larvae were screened for fluorescence between 24 and 72 hpf using a Zeiss Discovery.V12 SteREO fluorescence dissecting scope.

For confocal imaging, fish were anesthetized with 0.02% tricaine and mounted in 1.2% low melt agarose (Promega, V2111). Fluorescence was imaged with a Zeiss LSM 510 confocal

microscope using a 488 nm laser line for GFP/YFP/Citrine, 633 nm for Cy5 and 543 nm for mCherry/DsRed/Rhodamine. Images were taken with a 20x water objective (a 0.7x or 2x optical zoom was used where indicated) and projected from 20-50 optical sections of ~3 μm intervals.

Photoconversion of KikGR protein

Photoconversion of KikGR protein was performed using a Discovery.V12 SteREO fluorescence dissecting scope with a UV filter. Larvae were exposed to UV light for 5 minutes or until all neurons were photoconverted from green to red. Neurons were photoconverted in 24-hour intervals, subsequent to confocal imaging. Photoconverted larvae were raised at 28.5°C, in the dark, until imaging the following day.

Morphological analysis of peripheral axons

High-resolution (1024x1024 pixels) confocal images of single neurons were collected as described above. Using NeuroLucida tracing software, axons were traced in three dimensions; the primary axon projecting from the cell body to the first branch point in the skin was excluded from the analysis. Branch length and number for each traced neuron were acquired using NeuroLucidaExplorer software and analyzed with a Matlab program, which is described in detail in Supplementary Materials.

Retrograde labeling of Mauthner cells

Retrograde labeling of Mauthner cells was performed using rhodamine dextran (Invitrogen, 1824) according to established protocols (Fetcho and O'Malley, 1995; Volkmann and Köster, 2007). A filamented glass needle filled with a saturated concentration of rhodamine

dextran was used to create a lesion and introduce the dye within the caudal spinal cord of 48 hpf larvae. Larvae were allowed to recover for at least 24 hours at 28.5°C, in the dark, before imaging.

Whole Mount Double Fluorescent In Situ Hybridization and Antibody Staining

Embryos were fixed in 4% paraformaldehyde overnight at 4°C and permeabilized with proteinase K prior to antibody staining and/or in situ hybridization with standard protocols (Boué-Grabot et al., 2000; Slatter et al., 2005) (https://wiki.zfin.org/display/prot/Thisse+Lab+-+In+Situ+Hybridization+Protocol+-+2010+update). Antibody staining was performed with primary antibodies (PKCα, GFP/YFP/Citrine and mCherry) (1:500) incubated overnight at 4°C followed by incubation with AlexaFluor secondary antibodies (488 anti-mouse and 568 anti-rabbit for GFP, YFP and citrine expressing transgenics or 568 anti-mouse and 488 anti-rabbit for mCherry expressing transgenics) (1:1000) at 4°C overnight (Supplementary Table 3). Specificity of the PKCα antibody staining was verified with co-expression analysis, as shown in Figure 6. Supplementary Table 3 summarizes the antibodies used in this study. For fluorescent *in situ* hybridization, embryos were incubated with DIG labeled riboprobe (*p2rx3a* or *trkA*) and visualized using Cy5-Tyramide amplification (1:100) (Perkin Elmer, NEL752001KT). Embryos were mounted dorsally and imaged using confocal microscopy as described above.

Results

Zebrafish and pufferfish genomic sequences drive transgene expression in trigeminal and RB neurons

To analyze the development and diversity of somatosensory neurons in larval zebrafish, we created a library of transgenic reporters by cloning zebrafish *(Danio rerio)* and pufferfish *(Fugu rubripes*) genomic regions from upstream of neurotrophin receptors and ion channels (ATP-gated channels and Trp channels) and used them to drive expression of GFP (Table 1, Figure 1). To ensure robust reporter expression and allow for potential co-expression of effector genes, these candidate enhancers were subcloned upstream of the Gal4-VP16 transcriptional activator; on the same plasmid, multiple copies of the Gal4 upstream activation sequence (14xUAS) were used to drive GFP expression (Figure 1A) (Köster and Fraser, 2001). For comparison, we also created similar transgenes using two previously identified somatosensory-specific enhancers from the *islet1* gene, *isl1(ss)* and *CREST3* (Higashijima et al., 2000; Uemura et al., 2005). Similar to the *islet1* enhancers, several of the cloned candidate sequences (from the *trpA1a*, *Fru.trkA*, *Fru.trkB*, *Fru.trkC* and *Fru.p2x3-2* genes) drove GFP expression primarily in zebrafish trigeminal (Figure 1B-H) and RB (Figure 1B'-H') neurons. Many of these reporters were also consistently expressed in limited populations of additional cells, including muscles or other populations of neurons (Table 1). Stable transgenic reporter lines using several of these enhancers (*isl1(ss)*, *CREST3*, *Fru.trkA*, and *Fru.p2x3-2*) exhibited similar patterns of fluorescence, confirming the reliability of the transient transgenic approach for characterizing expression.

Using transient transgenesis, we analyzed the onset of fluorescence in somatosensory neurons to determine whether these reporters exhibited differences in temporal expression. The zebrafish *isl1(ss)* enhancer drove expression earliest, turning on at approximately the 13 somites (som) stage, before somatosensory neurons have projected central and peripheral axons. Expression from *CREST3* and *trpA1a* enhancer-driven reporters was first visible at the 17

somites stage, *Fru.p2x3-2* enhancer-driven expression began at the 20 somites stage, *Fru.trkA* enhancer-driven expression began at 28 hours post-fertilization (hpf) and *Fru.trkC* enhancer-driven expression began at 30 hpf (Figure 2A).

Determining when enhancer activity ceased was more difficult than determining when it turned on, since mRNA and proteins can perdure for many hours. We devised a technique using a photoconvertible fluorescent protein to address this issue and used it to characterize *isl1(ss)* and *CREST3* enhancer activity. Stable transgenic lines were made using these enhancers to drive expression of KikGR, a photoconvertible fluorescent protein that changes from green to red when exposed to UV light (Tsutsui et al., 2005). KikGR-expressing cells were photoconverted and imaged at daily intervals (Figure 2B). Expression of green fluorescence after photoconversion indicated newly synthesized KikGR; absence of green fluorescence after photoconversion indicated that the enhancer had turned off. With this approach, we found that the *isl1(ss)* enhancer drove expression until approximately 54 hpf (Figure 2C-G), while the *CREST3* enhancer continued to be active past 13 days post-fertilization (dpf) in a subset of somatosensory neurons (Figure 2C'-G' and Figure 3A).

Small regulatory regions were sufficient for transgene expression in zebrafish somatosensory neurons

Identifying smaller genomic regions sufficient to drive expression in somatosensory neurons could make enhancers easier to subclone, potentially improve expression specificity or efficiency, and makes possible the identification of conserved regulatory motifs. To isolate minimal sequences sufficient for somatosensory neuron expression, we performed sequential deletion analysis of the *Fru.trkA*, *Fru.p2x3-2*, and *trpA1a* enhancer regions. Specific regions

from within each enhancer sequence were subcloned into fluorescent reporter transgenes (Figure 4A), which were then injected into one-cell stage embryos and monitored for transient expression during the first three days of development. To estimate the relative efficiency of each enhancer, we quantified the number of embryos with fluorescence in somatosensory neurons in each injected clutch (Figure 4).

From an initial ~4 kb [-3939: -1] *Fru.trkA* enhancer region, we isolated a ~1 kb region proximal to the start site [-996: -1] capable of driving expression in somatosensory neurons (Figure 4B). This enhancer was almost twice as efficient as the original ~4 kb sequence (48% vs 25% of embryos with expression), indicating that there may be a negative regulatory element within the distal 3 kb of the original enhancer that reduced its activity. Subsequent deletions within the [-996: -1] fragment decreased its efficiency, with the two shortest sufficient elements [-996: -814] and [-238: -81] driving expression in an average of ~18% and 20% of embryos, respectively. However, when these two minimal regions [-996: -814] and [-238: -81] were combined [-996: -678, -238: -1], they were almost as efficient as the [-996: -1] fragment, indicating that each of the two small regulatory elements were sufficient for somatosensory expression, but sum to achieve optimal expression.

The ~1.6 kb [-1619: -1] *Fru.p2x3-2* enhancer was dissected into several progressively smaller fragments that drove expression in TG and RB neurons (Figure 4C). One of these fragments was 305 bps long [-1036: -731] and drove expression with comparable efficiency to the full 1.6 kb sequence (65% and 70% of embryos, respectively). A 144 bp fragment [-946: -802] was also sufficient for expression in somatosensory neurons, but was less efficient.

Dissection of the ~5 kb [-4993: -1] *trpA1a* genomic sequence indicated that the ~2 kb furthest from the start site was sufficient to drive expression (Figure 4D). By examining

overlapping fragments, we isolated a 941 bp region [-4457: -3517] that was sufficient for expression in somatosensory neurons. However, partially overlapping fragments [-4993: -4003] and [-3873: -3013] were unable to drive expression, indicating that multiple elements are required together. By creating 4 internal deletions of this 941 bp [-4457: -3517] fragment, we found two smaller regions, [-4136: -3946] and [-3946: -3517], that were required but not sufficient to drive robust expression in sensory neurons.

The Fru.p2x3-2 reporter and PKCα gene trap reporter define a somatosensory neuron subset that partially overlaps with neurons expressing a Fru.trkA reporter

PKCα is expressed in approximately 40% of RB neurons in 24 and 48 hpf larvae (Slatter et al., 2005). We obtained a gene trap line with an insertion in the gene encoding PKCα (*PKCα*^{ct7a}) to examine the relationship between the neurons marked by our reporter lines and PKCα-expressing neurons. In this line, an exon encoding the yellow fluorescent protein, citrine, is integrated between exons 3 and 4 of the PKCα gene, resulting in the expression of a PKCα-citrine fusion protein under the control of the endogenous PKCα transcriptional regulatory elements (Trinh et al., 2011). To determine if expression from *PKCα*^{ct7a} is confined to a subset of RB neurons, we crossed it to an *isl1(ss)* transgenic reporter line, *Tg(isl1(ss):Gal4-VP16,14xUAS:DsRed)*. Due to variegation from the Gal4-VP16/UAS system, DsRed in this line is expressed in approximately 74% of RB neurons (as determined by crossing it to a non-Gal4-VP16/UAS, non-variegated *isl2b:GFP* line) (Figure 5A and F). At 72 hpf, 41% of RB neurons in the *isl1(ss)* reporter line also expressed citrine from the *PKCα*^{ct7a} gene trap line, and 56% expressed DsRed only (Figure 5B and F). 3% of the labeled neurons expressed citrine only, which likely represents a population of neurons with silenced DsRed.

The *p2rx3a* gene is also expressed in a subset of somatosensory neurons (Boué-Grabot et al., 2000; Norton et al., 2000; Appelbaum et al., 2007). To create a potential reporter line for the *p2rx3a* gene, we used the *Fru.p2x3-2* enhancer (P3X3-2 is the pufferfish holomog of zebrafish P2rx3a) to drive expression of LexA-VP16 and drove expression of mCherry with the LexA operator (4xLexAop). Using the LexA-VP16/LexAop binary system has the same advantages as the Gal4-VP16/UAS system in terms of amplification and combinatorial versatility, but can be crossed to Gal4-VP16/UAS lines without cross-activating them (Lai and Lee, 2006). Crossing a *Fru.p2x3-2* reporter line, *Tg(Fru.p2x3-2[-1036,-731]:LexA-VP16,4xLexAop:mCherry)*, to the *isl2b:GFP* line revealed that the *Fru.p2x3-2* reporter labeled approximately 30% of RB neurons (Figure 5C and F). To determine the relationship between the *Fru.p2x3-2* reporter and $PKC\alpha^{ct7a}$ neurons, we crossed the two lines together (Figure 5D). Strikingly, citrine and mCherry expression almost completely coincided, with 96% of RB neurons expressing both (Figure 5F). Thus, an RB neuron subtype, comprising approximately 30-40% of the population, is marked by expression of both the *Fru.p2x3-2* reporter line and $PKC\alpha^{ct7a}$.

This subtype of RB neurons persists until at least 16 dpf (Figure 3B-E), well beyond the time when RB neurons are thought to disappear, consistent with previous reports that PKCα-expressing RB neurons are present until at least 2 weeks post-fertilization (Slatter et al., 2005; Patten et al., 2007). RB neurons in fish and amphibians degenerate during larval stages and their function is replaced by DRG neurons. It has been reported that most (Williams et al., 2000; Svoboda et al., 2001), or at least a subset (Reyes et al., 2004), of RB neurons degenerate in zebrafish between 2 and 4 dpf. We did not detect widespread RB neuron degeneration within the first five days of development with any of our reporter transgenes, and found that at least some RB neurons persisted for at least two weeks in several transgenic lines (Figure 3). This

observation is consistent with previous studies (Metcalfe et al., 1990; Slatter et al., 2005; Patten et al., 2007; O'Brien et al., 2012), suggesting that RB neuron degeneration in zebrafish might be more limited, or occur at later stages, than previously thought.

Like the *p2rx3a* gene, the *trkA* gene is also expressed in a subpopulation of larval zebrafish somatosensory neurons (Martin et al., 1995; Pan et al., 2012). We hypothesized that reporters using the *Fru.trkA* enhancer would highlight this subtype, allowing us to assess its relationship to the *Fru.p2x3-2* reporter-expressing subpopulation. We made three *Fru.trkA* transgenes using the Gal4-VP16/UAS system and one using the LexA-VP16/LexAop system to drive expression of different reporter genes. All of these lines were extensively variegated, but unfortunately, without the use of these amplification systems, the expression from this enhancer was very weak (data not shown). A line using *Fru.trkA* to drive expression of a YFP-tagged version of the light activated ChannelRhodopsin-2 (ChR2-YFP), *Tg(Fru.trkA:Gal4-VP16,14xUAS:ChR2-YFP),* was the least variegated, so we crossed it to the *Fru.p2x3-2* reporter line (Figure 5E). Fluorescence expression in the *Fru.trkA* reporter line partially overlapped with the subset labeled by the *Fru.p2x3-2* reporter line. Although variegation made it difficult to analyze neurons that were not labeled by the *Fru.trkA* reporter line, it was clear that only about half of the neurons expressing this reporter also expressed the *Fru.p2x3-2* reporter line, which is not highly variegated (see below), indicating that it marks a different, perhaps broader, set of neurons than the *Fru.p2x3-2*/*PKC*α^{ct7a} reporters (Figure 5F).

Morphological variation in peripheral and central arbors of RB neurons does not correlate with subtype reporters

Having identified PKCα/*p2rx3a*-expressing RB neurons as a distinct subtype, we hypothesized that these neurons might display a characteristic peripheral axon branching morphology optimized for their sensory properties or a distinct central connectivity pattern dictating their behavioral function. To test the former possibility, we examined confocal images of single neurons expressing different transgenic reporters (using the *isl1(ss)*, *Fru.trkA*, *Fru.trkC* or *Fru.p2x3-2* enhancers) in 72 hpf transient transgenic embryos (Figure 1B'-H'). Peripheral arbors of neurons labeled by each transgene did not exhibit dramatically distinct morphologies, like those that distinguish sensory subtypes in *Xenopus* (Hayes and Roberts, 1983) or *Drosophila* (Grueber et al., 2003) larvae. However, to assess morphology more objectively, we manually traced arbors of many single neurons arborizing over the central trunk with NeuroLucida tracing software (MicroBrightField), allowing us to extract quantitative information about each arbor. To compare arbors we created a Matlab program that generated a hierarchically clustered dendrogram reflecting the degree of morphological similarity between each pair of axons, based on branch length and number for each branch order. The algorithm reliably segregated most trigeminal and RB arbors into different groups and grouped arbors of the same neurons traced by different experimenters, giving us confidence that the method accurately distinguishes peripheral axons (Supplementary Figure 1A-A'). This program segregated RB arbors into five major clusters, but neurons labeled by different transgenes fell into multiple clusters, indicating substantial morphological diversity within each population of RB neurons labeled by these transgenic reporters (Supplementary Figure 1B-F). This quantitative analysis confirmed our qualitative impression that peripheral arbors marked by different transgenes do not exhibit dramatic morphological distinctions, but does not preclude the possibility that other more subtle features could distinguish these arbors.

The central axons of somatosensory neurons connect to downstream neurons in the central nervous system (CNS) to elicit behavioral responses. In zebrafish larvae, touch elicits a stereotypical C-start escape response (Saint-Amant and Drapeau, 1998; Drapeau et al., 2002). This response has been attributed to the direct activation of a well-characterized pair of reticulospinal interneurons that form in rhombomere 4 of the hindbrain, called Mauthner cells, by RB neurons (Korn and Faber, 2005; Kohashi and Oda, 2008). However, touch can elicit an escape response in the absence of the Mauthner cells (Liu and Fetcho, 1999; Burgess et al., 2009), or even the entire hindbrain (Downes and Granato, 2006), suggesting that RB neurons can connect directly to spinal cord circuits.

We hypothesized that subtypes of RB neurons might connect to distinct circuits, as has been suggested for trigeminal neurons (Pan et al., 2012). To characterize central axon projections of RB neurons, we imaged single neurons in transient transgenics while simultaneously visualizing the Mauthner cells by dye-filling (Fetcho and O'Malley, 1995; Volkmann and Köster, 2007) or with a transgenic line, Gt(T2KSAG)^{j1229a} (Burgess et al., 2009) (Figure 6A-B and D-D', respectively). Surprisingly, the central projections of most RB neurons failed to contact the Mauthner cell dendrite, instead terminating either in the caudal hindbrain or within the spinal cord (Figure 6B). Both Mauthner- and non-Mauthner dendrite-contacting neurons were situated all along the rostral-caudal extent of the spinal cord, though Mauthner dendrite-contacting RB neurons appeared to be more abundant in the rostral spinal cord (Figure 6C). Failure to contact the Mauthner dendrite was not due to delayed outgrowth, since these axons never grew further into the hindbrain (Figure 6D and D').

The majority of RB neurons labeled by all of the examined transgenes (using the *isl1(ss)*, *CREST3*, *Fru.p2x3-2* and *Fru.trkA* enhancers) failed to contact the Mauthner cell dendrite

(Figure 6E and data not shown). Notably, neurons labeled by the *Fru.p2x3-2* reporter displayed a similar distribution of central axon termination patterns as axons labeled by the *isl1(ss)* and *CREST3* enhancer-driven reporters, indicating that although PKCα/*p2rx3a*-expressing neurons define a distinct RB subtype, neurons within that subtype can exhibit different termination patterns.

Fluorescent transgenes reflect endogenous gene expression

To determine whether the subpopulations of neurons identified with our reporters reflect endogenous gene expression and thus truly distinct molecular subtypes of somatosensory neurons, we examined endogenous protein or mRNA expression in our reporter lines (Figure 7). Double antibody staining for PKCα protein and GFP in the *isl2b:GFP* line confirmed that PKCα-expressing neurons make up a subpopulation of RB neurons (Figure 7D and G). Staining for endogenous PKCα and for citrine in *PKCα*^{ct7a} verified that the gene trap line faithfully labels all PKCα-expressing RB neurons at 72 hpf (Figure 7A and G). As expected, PKCα and mCherry antibody staining in the *Fru.p2x3-2* reporter line, *Tg(Fru.p2x3-2[-1036, -731]:LexA-VP16,4xLexAop:mCherry)*, revealed that most PKCα-expressing neurons (89%) are labeled by the *Fru.p2x3-2* reporter line (Figure 7E, G).

Fluorescent *in situ* hybridization (FISH) for *p2rx3a* (previously reported as *p2x3*, *p2rx3* and *p2x3.1*) and antibody staining for mCherry in 72 hpf *Tg(Fru.p2x3-2[-1036, -731]:LexA-VP16,4xLexAop:mCherry)* larvae revealed that 92% of *p2rx3a*-expressing neurons also expressed the *Fru.p2x3-2* reporter, indicating that transgene expression faithfully recapitulated endogenous expression but was silenced in, at most, 10% of *p2rx3a*-expressing neurons. At 54 hpf, a larger proportion of neurons stained with the *p2rx3a* mRNA probe failed to express the

reporter (20-35%, data not shown), perhaps reflecting cross-hybridization with the *p2rx3a* ortholog, *p2rx3b*, since increasing the hybridization temperature reduced the number of neurons that did not co-express the transcript and the reporter.

To characterize *trkA* (also known as *ntrk1*) expression in the *Fru.trkA* reporter line, *Tg(Fru.trkA:Gal4-VP16,14xUAS:ChR2-YFP)*, we performed FISH for *trkA* mRNA and antibody staining for YFP at 32 hpf, when gene expression appeared to be highest. 60% of neurons expressing *trkA* mRNA also expressed YFP, but few YFP-expressing neurons failed to express *trkA* (Figure 7C and G). This result confirms that the *Fru.trkA* reporter line is variegated (YFP was silenced in ~40% of *trkA* neurons), but also that expression from the transgene faithfully reflects endogenous *trkA* expression.

To determine the relationship between the PKCα/*p2rx3a* subtype of RB neurons and expression in previously published *p2rx3b:EGFP* and *trpA1b:EGFP* BAC reporter lines, we performed double antibody staining for GFP and endogenous PKCα in 72 hpf larvae from each of these BAC reporter lines. Approximately 56% of *p2rx3b:EGFP*-expressing RB neurons also expressed PKCα (Figure 8A and B), indicating that PKCα/*p2rx3a*-expressing neurons are a subset of *p2rx3b:EGFP*-expressing neurons. This finding also suggests that *p2rx3b:EGFP* is expressed in most, if not all, RB neurons. In contrast, expression of the *trpA1b:EGFP* BAC reporter completely overlapped with PKCα-expressing RB neurons. We have thus identified a subtype of RB neurons expressing three different transgenic reporter lines (PKCα/*p2rx3a/trpA1b)* and accurately reflecting the expression of genes involved in sensory signal transduction.

Discussion

New enhancers and transgenic reporters for studying zebrafish somatosensory neurons

To sense diverse touch stimuli, vertebrate somatosensation is carried out by several functionally and molecularly distinct sensory neuron subtypes. Although less well characterized, multiple somatosensory neuron subtypes exist even at early embryonic stages. Despite having morphologically similar cutaneous endings (O'Brien et al., 2012), subpopulations of RB neurons in larval zebrafish differentially express genes such as *pkcα*, *p2rx3a* and specific sodium channels (Boué-Grabot et al., 2000; Norton et al., 2000; Slatter et al., 2005; Pineda et al., 2006; Appelbaum et al., 2007; Patten et al., 2007). To characterize the morphology, function and connectivity of larval zebrafish somatosensory neuron subpopulations we created several new reporter transgenes that express fluorescent proteins in these neurons using enhancers from neurotrophin receptors and ion channels. Some of the enhancers were cloned from the pufferfish genome, which we used because its more compact genome made it easier to identify regulatory elements. The fact that these enhancers drove expression in the expected neuronal populations indicates that they contain regulatory elements conserved at least among teleost fish. By dissecting these enhancer sequences and performing expression analysis, we identified several compact sequences (<200 bp) that are sufficient and/or required for somatosensory neuron expression, making them excellent starting points for identifying transcriptional regulatory motifs. Given their portability, strong expression, and temporal variability, these new enhancers and reporters will be useful tools for developmental and functional studies of zebrafish somatosensory neurons.

RB neurons possess diverse peripheral axon morphologies and central axon termination patterns, even within molecularly distinct subtypes

Different axon morphologies may be optimized for the sensation of particular kinds of stimuli. In *Drosophila* and *Xenopus*, the complexity of cutaneous neurite branches defines functionally distinct somatosensory neuron subtypes (Hayes and Roberts, 1983; Grueber et al., 2003). Zebrafish trigeminal and RB neurons display a spectrum of arbor morphologies, prompting us to hypothesize that branching morphology is characteristic of RB subtypes as well. Analysis of branching morphology of RB neurons labeled by reporters driven by the *isl1(ss)*, *Fru.trkA*, *Fru.trkC* and *Fru.p2x3-2* enhancers grouped axon arbors into five categories. However, most of these reporters were found in all five categories, potentially implying that branching morphology may be a stochastically determined property unrelated to sensory function.

Central axon termination patterns determine the potential neural circuits that a particular sensory neuron can activate. Touch usually elicits an escape response in larval zebrafish, but kinematically distinct patterns of escape can be distinguished (Saint-Amant and Drapeau, 1998; Liu and Fetcho, 1999; Drapeau et al., 2002; Burgess et al., 2009; Liu et al., 2012), implying that different circuits underlie them. The Mauthner cell is the central component of the classic escape response circuit (Liu and Fetcho, 1999; Korn and Faber, 2005; Kohashi and Oda, 2008). It was therefore surprising to find that the majority of RB neurons visualized with any of our transgenic reporters failed to contact the Mauthner cell dendrite, terminating instead within the spinal cord or caudal hindbrain. Touch can elicit escape responses in the absence of Mauthner cells (Liu and Fetcho, 1999; Burgess et al., 2009), or even of the entire hindbrain (Downes and Granato, 2006), implying that RB neurons must also connect to local spinal circuits. Mauthner-independent escape responses have a substantially slower latency than the classic escape response (Liu and Fetcho, 1999; Kohashi and Oda, 2008; Burgess et al., 2009), suggesting that non-Mauthner-

contacting RB cells may elicit those slow responses. If that is the case, activating individual neurons within this class might elicit distinct behaviors. The new transgenic enhancers presented here provide valuable tools for addressing this hypothesis and for comprehensively analyzing the sensory and behavioral functions of subtypes of somatosensory neuron in larval zebrafish.

A somatosensory neuron subtype is defined by PKCα and p2rx3a expression

Analysis of the *PKCα*^{ct7a} gene trap line, the *Fru.p2x3-2* reporter line and the *trpA1b:EGFP* BAC transgenic line revealed that they highlight the same population of RB neurons and faithfully reflect endogenous expression of PKCα and *p2rx3a*, thus defining a molecularly distinct somatosensory neuron subtype. To our knowledge, this is the first report of a somatosensory neuron subtype in zebrafish defined by more than one reporter or gene. In different experiments the percentage of RB neurons highlighted by these two reporters ranged from 30-50%, likely reflecting variability from animal to animal, but nonetheless consistent with previous reports that PKCα is expressed in ~40% of RB neurons (Slatter et al., 2005; Patten et al., 2007). PKCα is a kinase that functions in diverse signaling pathways, P2rx3a is an ATP-gated ion channel involved in nociception, and TrpA1b is a channel that is activated by pungent chemicals (Caron et al., 2008; Prober et al., 2008; Pan et al., 2012). This expression pattern suggests that these cells likely detect distinct nociceptive somatosensory stimuli and possess unique physiological properties. Since they likely detect distinct stimuli, the peripheral arbors of PKCα/*p2rx3a/trpA1b*-expressing neurons likely "tile" the skin independently of other RB neurons, so that animals can sense each modality throughout their body. This prediction remains to be tested, but our previous observation that each zebrafish trigeminal axon arbor repels approximately half of its neighboring arbors is consistent with this model (Sagasti et al., 2005).

Acknowledgements

We would like to thank Lindsey Mork for initial cloning of zebrafish enhancers, Holly Vu for tracing neurons, Hillary McGraw for advice on *trkA in situ* hybridization, Michael Granato for the *Gt(T2KSAG)*^{j1229a} line, Albert Pan and Alex Schier for the *Tg(trpA1b:EGFP)* line, Mark Voigt for the *Tg(p2rx3b:EGFP)* line, Matt Veldman, Ann Cavanaugh and Kevin Mouillesseaux for technical advice and members of the Sagasti Lab for comments on the manuscript. Funds were provided by an NRSA (5F31NS064817) to AMSP from the NINDS and grants from the NSF (RIG:0819010) and NIDCR (5R01DE018496) to AS.

Literature Cited

Appelbaum L, Skariah G, Mourrain P, Mignot E. 2007. Comparative expression of p2x receptors and ecto-nucleoside triphosphate diphosphohydrolase 3 in hypocretin and sensory neurons in zebrafish. Brain Res 1174:66-75.

Boué-Grabot E, Akimenko MA, Séguéla P. 2000. Unique functional properties of a sensory neuronal P2X ATP-gated channel from zebrafish. J Neurochem 75:1600-1607.

Burgess HA, Johnson SL, Granato M. 2009. Unidirectional startle responses and disrupted left-right co-ordination of motor behaviors in robo3 mutant zebrafish. Genes Brain Behav 8:500-511.

Caron SJ, Prober D, Choy M, Schier AF. 2008. In vivo birthdating by BAPTISM reveals that trigeminal sensory neuron diversity depends on early neurogenesis. Development 135:3259-3269.

Davies A, Lumsden A. 1984. Relation of target encounter and neuronal death to nerve growth factor responsiveness in the developing mouse trigeminal ganglion. J Comp Neurol 223:124-137.

Downes GB, Granato M. 2006. Supraspinal input is dispensable to generate glycine-mediated locomotive behaviors in the zebrafish embryo. J Neurobiol 66:437-451.

Drapeau P, Saint-Amant L, Buss RR, Chong M, McDearmid JR, Brustein E. 2002. Development of the locomotor network in zebrafish. Prog Neurobiol 68:85-111.

Fetcho JR, O'Malley DM. 1995. Visualization of active neural circuitry in the spinal cord of intact zebrafish. J Neurophysiol 73:399-406.

Goll MG, Anderson R, Stainier DY, Spradling AC, Halpern ME. 2009. Transcriptional silencing and reactivation in transgenic zebrafish. Genetics 182:747-755.

Grueber WB, Ye B, Moore AW, Jan LY, Jan YN. 2003. Dendrites of distinct classes of Drosophila sensory neurons show different capacities for homotypic repulsion. Curr Biol 13:618-626.

Hayes BP, Roberts A. 1983. The anatomy of two functional types of mechanoreceptive 'free' nerve-ending in the head skin of Xenopus embryos. Proc R Soc Lond B Biol Sci 218:61-76.

Higashijima S, Hotta Y, Okamoto H. 2000. Visualization of cranial motor neurons in live transgenic zebrafish expressing green fluorescent protein under the control of the islet-1 promoter/enhancer. J Neurosci 20:206-218.

Kawakami K. 2004. Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. Methods Cell Biol 77:201-222.

Kimmel CB, Hatta K, Metcalfe WK. 1990. Early axonal contacts during development of an identified dendrite in the brain of the zebrafish. Neuron 4:535-545.

Kohashi T, Oda Y. 2008. Initiation of Mauthner- or non-Mauthner-mediated fast escape evoked by different modes of sensory input. J Neurosci 28:10641-10653.

Korn H, Faber DS. 2005. The Mauthner cell half a century later: a neurobiological model for decision-making? Neuron 47:13-28.

Kucenas S, Soto F, Cox JA, Voigt MM. 2006. Selective labeling of central and peripheral sensory neurons in the developing zebrafish using P2X(3) receptor subunit transgenes. Neuroscience 138:641-652.

Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, Campbell DS, Parant JM, Yost HJ, Kanki JP, Chien CB. 2007. The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn 236:3088-3099.

Köster RW, Fraser SE. 2001. Tracing transgene expression in living zebrafish embryos. Dev Biol 233:329-346.

Lai SL, Lee T. 2006. Genetic mosaic with dual binary transcriptional systems in Drosophila. Nat Neurosci 9:703-709.

Liu KS, Fetcho JR. 1999. Laser ablations reveal functional relationships of segmental hindbrain neurons in zebrafish. Neuron 23:325-335.

Liu YC, Bailey I, Hale ME. 2012. Alternative startle motor patterns and behaviors in the larval zebrafish (Danio rerio). J Comp Physiol A Neuroethol Sens Neural Behav Physiol 198:11-24.

Lumpkin EA, Caterina MJ. 2007. Mechanisms of sensory transduction in the skin. Nature 445:858-865.

Marmigère F, Ernfors P. 2007. Specification and connectivity of neuronal subtypes in the sensory lineage. Nat Rev Neurosci 8:114-127.

Martin SC, Marazzi G, Sandell JH, Heinrich G. 1995. Five Trk receptors in the zebrafish. Dev Biol 169:745-758.

Metcalfe WK, Myers PZ, Trevarrow B, Bass MB, Kimmel CB. 1990. Primary neurons that express the L2/HNK-1 carbohydrate during early development in the zebrafish. Development 110:491-504.

Moore SJ, Munger BL. 1989. The early ontogeny of the afferent nerves and papillary ridges in human digital glabrous skin. Brain Res Dev Brain Res 48:119-141.

Norton WH, Rohr KB, Burnstock G. 2000. Embryonic expression of a P2X(3) receptor encoding gene in zebrafish. Mech Dev 99:149-152.

O'Brien GS, Rieger S, Wang F, Smolen GA, Gonzalez RE, Buchanan J, Sagasti A. 2012. Coordinate development of skin cells and cutaneous sensory axons in zebrafish. J Comp Neurol 520:816-831.

Pan YA, Choy M, Prober DA, Schier AF. 2012. Robo2 determines subtype-specific axonal projections of trigeminal sensory neurons. Development 139:591-600.

Patten SA, Sihra RK, Dhami KS, Coutts CA, Ali DW. 2007. Differential expression of PKC isoforms in developing zebrafish. Int J Dev Neurosci 25:155-164.

Pineda RH, Svoboda KR, Wright MA, Taylor AD, Novak AE, Gamse JT, Eisen JS, Ribera AB. 2006. Knockdown of Nav1.6a Na+ channels affects zebrafish motoneuron development. Development 133:3827-3836.

Pittman AJ, Law MY, Chien CB. 2008. Pathfinding in a large vertebrate axon tract: isotypic interactions guide retinotectal axons at multiple choice points. Development 135:2865-2871.

Prober DA, Zimmerman S, Myers BR, McDermott BM, Kim SH, Caron S, Rihel J, Solnica-Krezel L, Julius D, Hudspeth AJ, Schier AF. 2008. Zebrafish TRPA1 channels are required for chemosensation but not for thermosensation or mechanosensory hair cell function. J Neurosci 28:10102-10110.

Reyes R, Haendel M, Grant D, Melancon E, Eisen JS. 2004. Slow degeneration of zebrafish Rohon-Beard neurons during programmed cell death. Dev Dyn 229:30-41.

Roberts A. 1980. The function and role of two types of mechanoreceptive "free" nerve endings in the head skin of amphibian embryos. Journal of Comparative Physiology 135:341-348.

Sagasti A, Guido MR, Raible DW, Schier AF. 2005. Repulsive interactions shape the morphologies and functional arrangement of zebrafish peripheral sensory arbors. Curr Biol 15:804-814.

Saint-Amant L, Drapeau P. 1998. Time course of the development of motor behaviors in the zebrafish embryo. J Neurobiol 37:622-632.

Slatter CA, Kanji H, Coutts CA, Ali DW. 2005. Expression of PKC in the developing zebrafish, Danio rerio. J Neurobiol 62:425-438.

Svoboda KR, Linares AE, Ribera AB. 2001. Activity regulates programmed cell death of zebrafish Rohon-Beard neurons. Development 128:3511-3520.

Trinh lA, Hochgreb T, Graham M, Wu D, Ruf-Zamojski F, Jayasena CS, Saxena A, Hawk R, Gonzalez-Serricchio A, Dixson A, Chow E, Gonzales C, Leung HY, Solomon I, Bronner-Fraser M, Megason SG, Fraser SE. 2011. A versatile gene trap to visualize and interrogate the function of the vertebrate proteome. Genes Dev 25:2306-2320.

Tsutsui H, Karasawa S, Shimizu H, Nukina N, Miyawaki A. 2005. Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. EMBO Rep 6:233-238.

Uemura O, Okada Y, Ando H, Guedj M, Higashijima S, Shimazaki T, Chino N, Okano H, Okamoto H. 2005. Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression. Dev Biol 278:587-606.

Volkmann K, Köster RW. 2007. In vivo retrograde labeling of neurons in the zebrafish embryo or larva with rhodamine dextran. CSH Protoc 2007:pdb.prot4832.

Williams JA, Barrios A, Gatchalian C, Rubin L, Wilson SW, Holder N. 2000. Programmed cell death in zebrafish rohon beard neurons is influenced by TrkC1/NT-3 signaling. Dev Biol 226:220-230.

Figure Legends

Figure 1. New transgenes drive expression in zebrafish somatosensory neurons. (A) Design of transgenic reporters: Genomic sequences containing somatosensory-specific enhancers drove the Gal4-VP16 transcriptional activator; on the same plasmid 14 copies of the Gal4 upstream activation sequence (14xUAS) were used to drive GFP. The adenovirus E1b minimal promoter was placed upstream of both Gal4-VP16 and the reporter gene. An SV40 polyadenylation sequence was placed 3' to both Gal4-VP16 and the reporter gene to signal transcription termination. (B-H') Transgenes were transiently expressed in zebrafish larvae and imaged by confocal microscopy at 72 hpf. Both zebrafish and pufferfish (Fru) genomic enhancer sequences drove expression of fluorescent reporters in zebrafish trigeminal (B-H) and Rohon-Beard (B'-H') neurons. Anterior is left and dorsal is up in all images. The eye (e) and yolk (y) are indicated. Arrowheads point to cell bodies; arrows point to peripheral arbors; empty arrows point to muscle. Scale bar, 100 μm.

Figure 2. Different transgenes have distinct temporal expression patterns. (A) Onset of gene expression as determined by transient transgenesis. Transgenes with indicated enhancers driving GFP (see Figure 1) were injected into embryos at the one-cell stage. Embryos were raised at 28.5°C and staged according to somite (som) number and hours post fertilization (hpf).

Initial observation of GFP expression was recorded: Expression driven by the *isl1(ss)* enhancer was the earliest at 13 som, followed by *CREST3* and *trpA1a* enhancer-driven expression at 17 som, *Fru.p2x3-2* enhancer-driven expression at 20 som, *Fru.trkA* enhancer-driven expression at 28 hpf and *Fru.trkC* enhancer-driven expression at 30 hpf. (B) The duration of *isl1(ss)* and *CREST3* enhancer activity was determined by daily photoconversion and imaging of KiKGR expression. Green fluorescence indicated newly synthesized KikGR (active promoter, new neuron), red fluorescence indicated older/photoconverted KikGR (inactive promoter, old neuron) and yellow indicated neurons expressing both forms of KikGR (active promoter, old neuron). (C-G) The *isl1(ss)* enhancer was active from 14 hpf until approximately 54 hpf. (C'-G') Enhancer activity in sensory neurons lasted until at least 14 dpf (see Figure 3). Confocal images were taken with a 20x objective and 0.7x optical zoom; anterior is left and dorsal is up. Empty arrows indicate the trigeminal ganglia, arrowheads point to anterior RB neurons, filled arrows are DRGs and asterisks are located near a population of anterior neurons. Scale bar, 300 μm.

Figure 3. Rohon-Beard neurons persist in 2-week-old larvae. Somatosensory neurons in (A) *CREST3,* (B) *PKCα*^{ct7a}, (C) *Fru.p2x3-2,* (D) *Fru.p2x3-2/PKCα*^{ct7a} and (E) *isl2b/Fru.p2x3-2* transgenic reporter fish were visible until at least two weeks post-fertilization. Confocal images were taken with a 20x objective; anterior is left and dorsal is up. (E'-E''') Magnified images of various regions outlined in D. Trigeminal neurons are shown in D' and RB neurons in D'' and D'''. Arrows point to individual RB neuron cell bodies. Arrowheads in D'-D''' point to RB neuron peripheral arbors. Scale bar, 200 μm in A-D and 50 μm in D'-D'''.

Figure 4. Identification of minimal regulatory elements sufficient for transgene expression in zebrafish trigeminal and Rohon-Beard neurons. (A) Transgenes were constructed by cloning genomic sequences upstream of Gal4-VP16,14xUAS:GFP or LexA-VP16,4xLexAop:mCherry; the E1b promoter and SV40 pA sequence were used as indicated. (B) Two fragments of the *Fru.trkA* enhancer, each less than 200 bp, were sufficient to drive expression in sensory neurons, though both together were required to drive expression at levels similar to the ~1 kb fragment [-996: -1]. (C) A 144 bp fragment [-946: -802] of the *Fru.p2x3-2* enhancer sequence was sufficient to drive expression in somatosensory neurons, though its efficiency was lower than that of a 305 bp fragment [-1036: -731]. (D) The *trpA1a* enhancer sequence was divided into 4 functional elements. The right column indicates the percent of embryos with expression within a clutch. (+) in (D) indicates qualitatively similar expression from each transgene, though this was not quantitatively measured.

Figure 5. The PKCα^{ct7a} gene trap line and a transgenic reporter using a *Fru.p2x3-2* enhancer label the same subset of Rohon-Beard neurons. (A-E) Transgenic lines, each using the indicated enhancers to drive expression of different fluorescent proteins, were crossed to compare their expression patterns. (A) Crossing *Tg(isl1(ss):Gal4-VP16,14xUAS:DsRed)* to *Tg(-17.6isl2b:GFP)* allowed us to estimate the degree of variegation caused by the Gal4/UAS system: The *isl2b* (GFP) and *isl1(ss)* (DsRed) enhancers are thought to drive expression in all somatosensory neurons, however the *isl1(ss)* reporter line was expressed in only 74% of *isl2b:GFP* RB neurons (co-expression shown in yellow), suggesting that in the *isl1(ss)* line the Gal4/UAS system was silenced in ~25% of RB neurons. Small lateral DsRed-labeled cells (indicated by arrowheads) are not RB neurons. (B) The PKCα^{ct7a} line labeled ~40% the RB

neurons labeled by the variegated *isl1(ss)* reporter line. (C) The *Fru.p2x3-2* reporter line labeled approximately a third of all RB neurons labeled by *isl2b:GFP*. (D) The *Fru.p2x3-2* reporter line was expressed in the same population of RB neurons as the PKCα^{ct7a} line. Note that the PKCα fusion protein is excluded from the nucleus. (E) Approximately 50% of RB neurons labeled by the *Fru.trkA* reporter line overlapped with the *Fru.p2x3-2* reporter line. Dorsal confocal images were taken of the anterior spinal cord with a 20x objective; anterior is up. Scale bar, 100 μm. (F) Quantification of co-expression between transgenic lines. n = total number of RB neurons. Asterisks indicate categories that may underestimate co-expression due to variegation in some reporter lines.

Figure 6. Rohon-Beard central projections either contact or fail to contact the Mauthner cell dendrite. (A-B) Dorsal confocal images of 72 hpf larvae exhibit two patterns of central projection termination. Some RB central projections contacted the Mauthner cell dendrite (labeled by retrograde dye-filling) (Mauthner) (A), whereas others terminated caudal to the Mauthner cell dendrite (non-Mauthner) (B). Arrows indicate the rostral termini of RB central axons. The Mauthner cells are outlined. Scale bar, 50 μm. (C) The position of a neuron's cell body along the rostral-caudal axis did not dictate whether or not its central projection contacted the Mauthner cell dendrite. (D and D') Failure to contact the Mauthner cell dendrite was not due to slow RB central axon growth. Confocal images of a non-Mauthner contacting RB neuron (red) in the *Gt(T2KSAG)*^{j1229a} gene trap line taken at 3 and 4 dpf. Scale bar, 50 μm. (E) All transgenes primarily labeled non-Mauthner dendrite contacting RB neurons. n = total number of RB neurons.

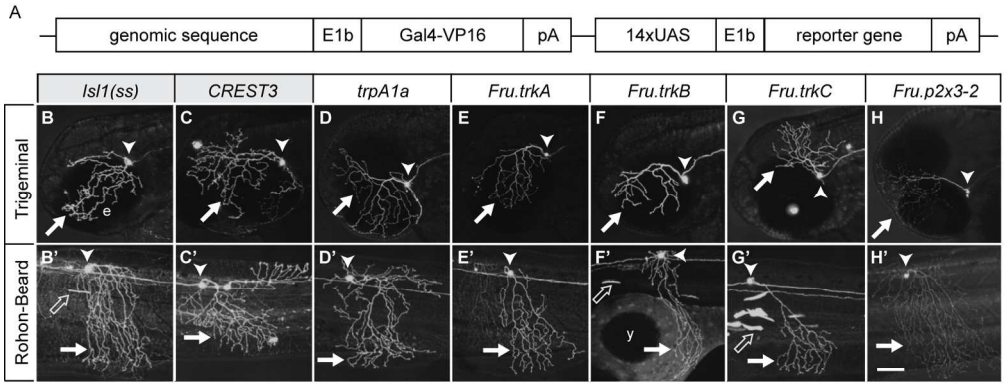
Figure 7. Fluorescent transgenic reporters accurately reflect endogenous gene expression. (A, D-F) Whole mount double antibody staining for endogenous protein expression and transgene-driven fluorescent protein expression. (B and C) Fluorescent *in situ* hybridization for endogenous mRNA and antibody staining for transgene-driven fluorescent protein expression. The label above each panel indicates the endogenous protein or mRNA visualized and the transgene enhancer used. Filled arrows indicate cells co-expressing the transgene and protein or mRNA; open arrows indicate cells expressing only one or the other. Confocal images of the dorsal spinal cord were taken with a 20x objective and a 2x optical zoom; anterior is up. Scale bar, 50 μm. (G) Quantification of co-expression between transgene and endogenous genes. The PKCα^{ct7a} line faithfully labeled all (100%) cells expressing PKCα protein, the *Fru.p2x3-2* reporter line labeled most (92%) *p2rx3a* mRNA-expressing neurons and the *Fru.trkA* reporter line labeled the majority (60%) of *trkA* expressing neurons. Note that many neurons expressing *trkA* mRNA did not express the reporter, indicating variegation. PKCα protein was expressed in 46% of the total (*isl2b:GFP*-expressing) RB neuron population. PKCα endogenous expression almost perfectly co-labeled neurons marked by the *Fru.p2x3-2* reporter line and vice versa. Expression of the PKCα protein only partially overlapped with *Fru.trkA* reporter line expression, as expected from analyzing transgene co-expression (Figure 5). Due to variegation of the transgene (indicated by the asterisk) the number of neurons expressing PKCα protein only is an overestimate. n = total number of RB neurons.

Figure 8. Transgenic lines label molecular subtypes of somatosensory neurons. (A) Antibody staining for PKCα (red) and GFP (green) in 72 hpf *trpA1b:EGFP* and *p2rx3b:EGFP* BAC transgenic lines. Confocal images of the spinal cord were taken with a 20x objective and a

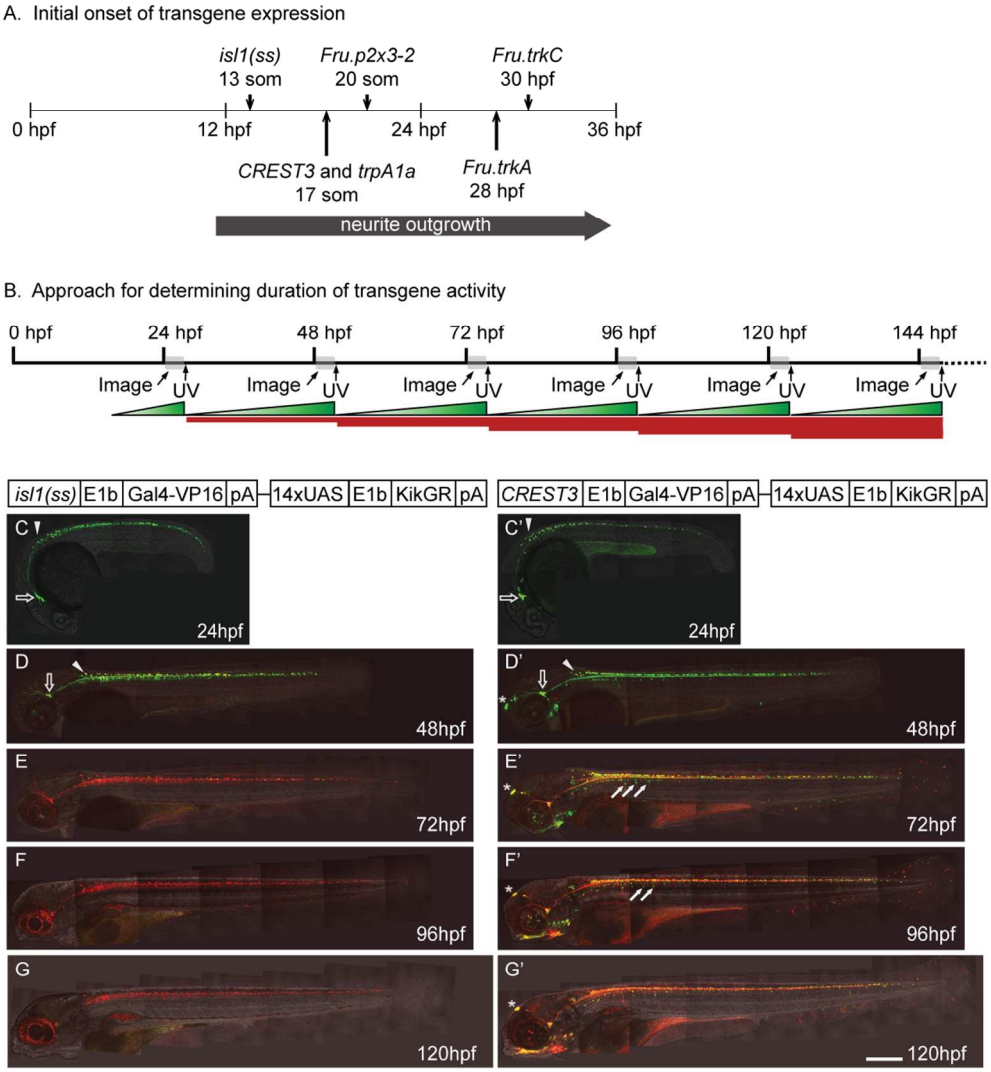
2x optical zoom; lateral images, anterior is left and dorsal is up. Scale bar, 50 μm. (B) 100% of *trpA1b:EGFP* RB neurons expressed PKCα. Approximately 56% of *p2rx3b:EGFP* RB neurons also expressed PKCα. n = number of RB neurons. (C) Model for zebrafish larval RB neuron subtypes: *isl2b* and *isl1(ss)* highlight all RB somatosensory neurons. *p2rx3b:EGFP* expression labels most, if not all, RB neurons. PKC, *p2rx3a* and *trpA1b:EGFP* are co-expressed in a distinct population of neurons that make up approximately 40% of all RB neurons. *trkA*-expressing neurons partially overlap with the PKCα/*p2rx3a* subtype.

Table 1. Enhancer regions from the zebrafish and pufferfish genomes drive reporter expression in somatosensory neurons. Enhancer regions were isolated by cloning sequences upstream of the translational start site (ATG) of indicated genes (Enhancers) of indicated lengths (Size, in base pairs) from the genomes of either zebrafish or pufferfish (Origin). All sequences drove expression in trigeminal and Rohon-Beard neurons (TG and RB); a few drove expression in additional tissues, such as muscle and other neurons (Expression). Expression analysis of all transgenes was performed with transient transgenesis, as described in the text. In some cases stable transgenic lines (*) verified expression specificity. Shaded rows indicate previously reported enhancers.

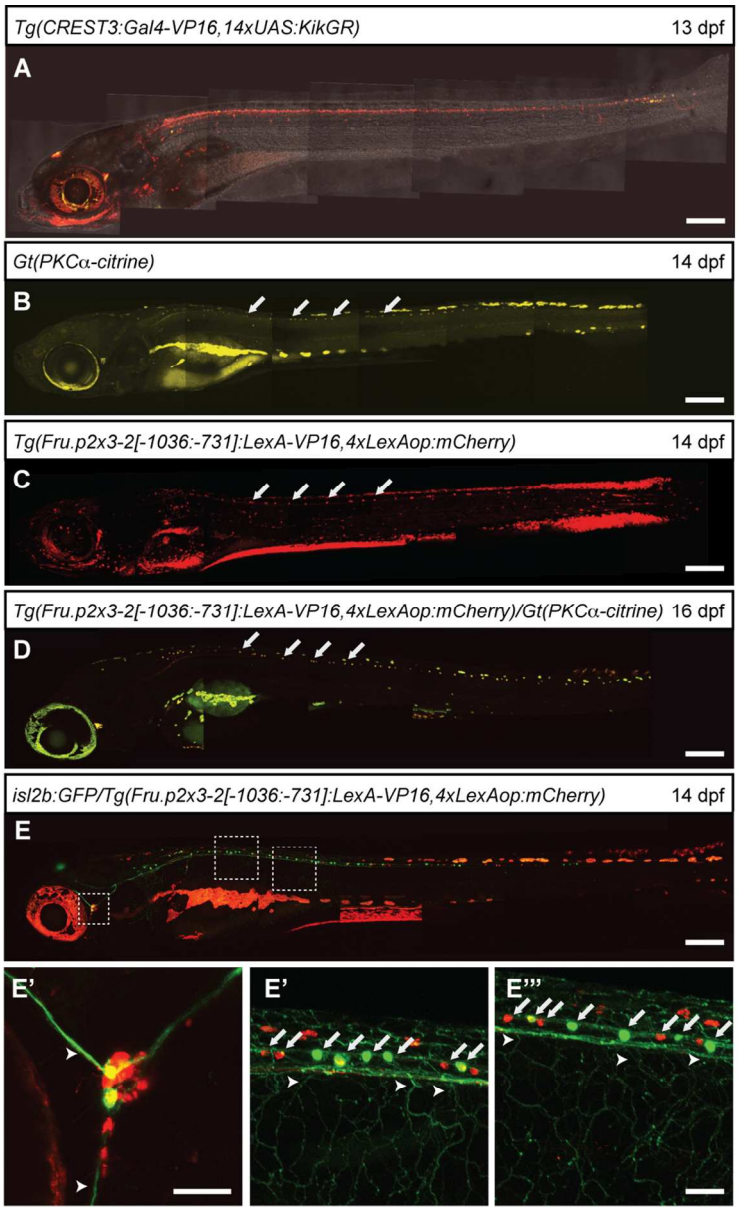
Enhancer	Size (bp)	Origin	Expression
ilsl1(ss)	4200	zebrafish	TG, RB, other neurons*
CREST3	868	zebrafish	TG, RB, DRG*
trpA1a	5019	zebrafish	TG, RB, lateral line
trkA	3939	pufferfish	TG, RB*
trkB	4017	pufferfish	TG, RB, other neurons
trkC	3936	pufferfish	TG, RB, other neurons, muscle
p2x3-2	1620	pufferfish	TG, RB, muscle*



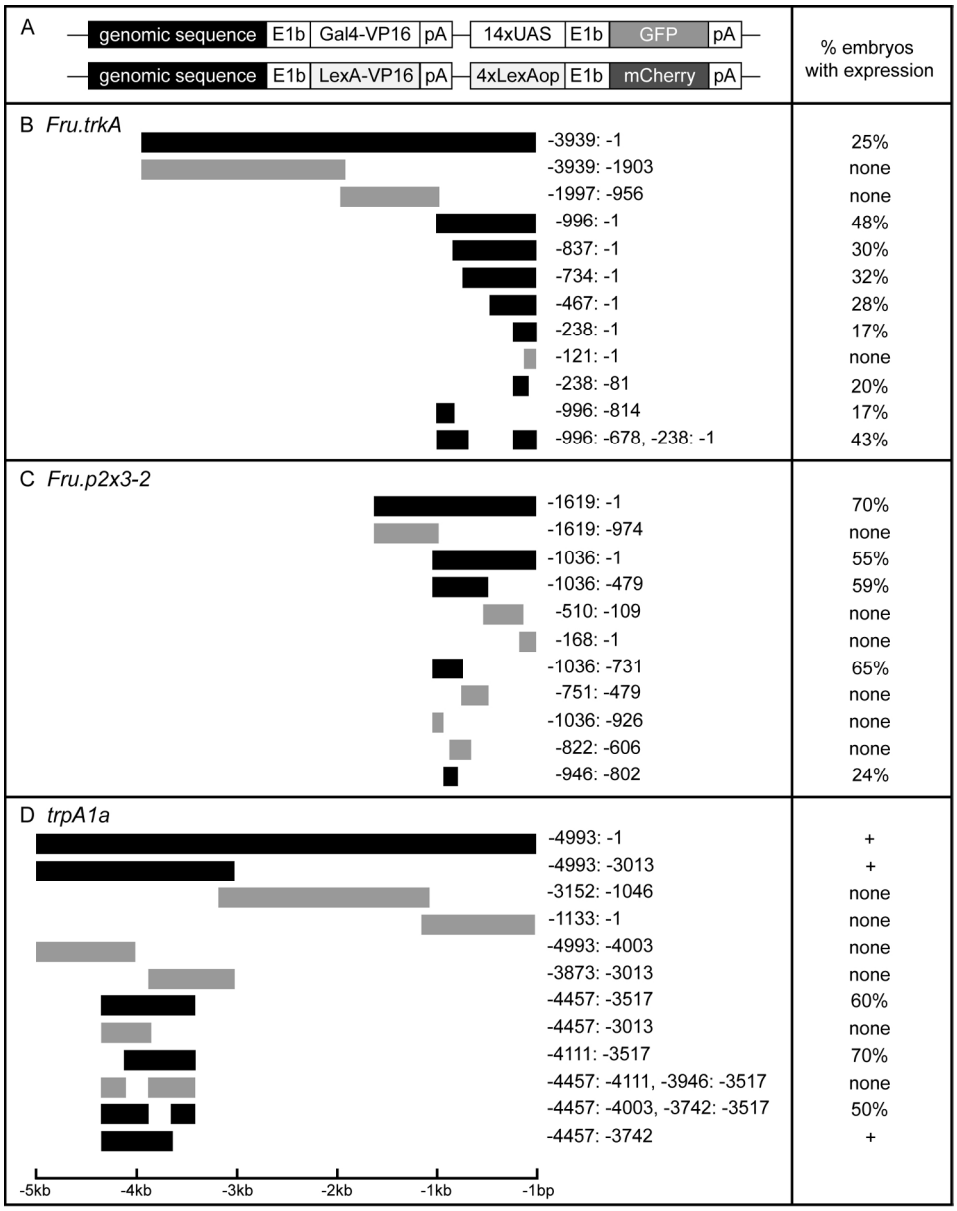
239x91mm (200 x 200 DPI)



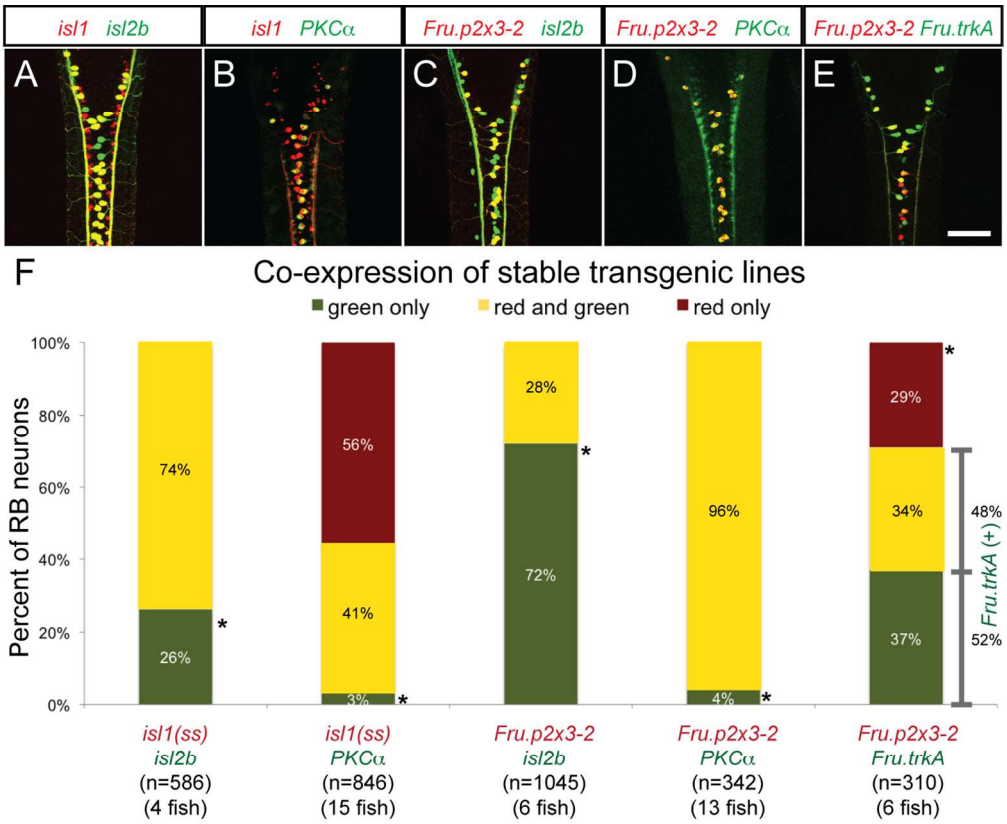
217x232mm (150 x 150 DPI)



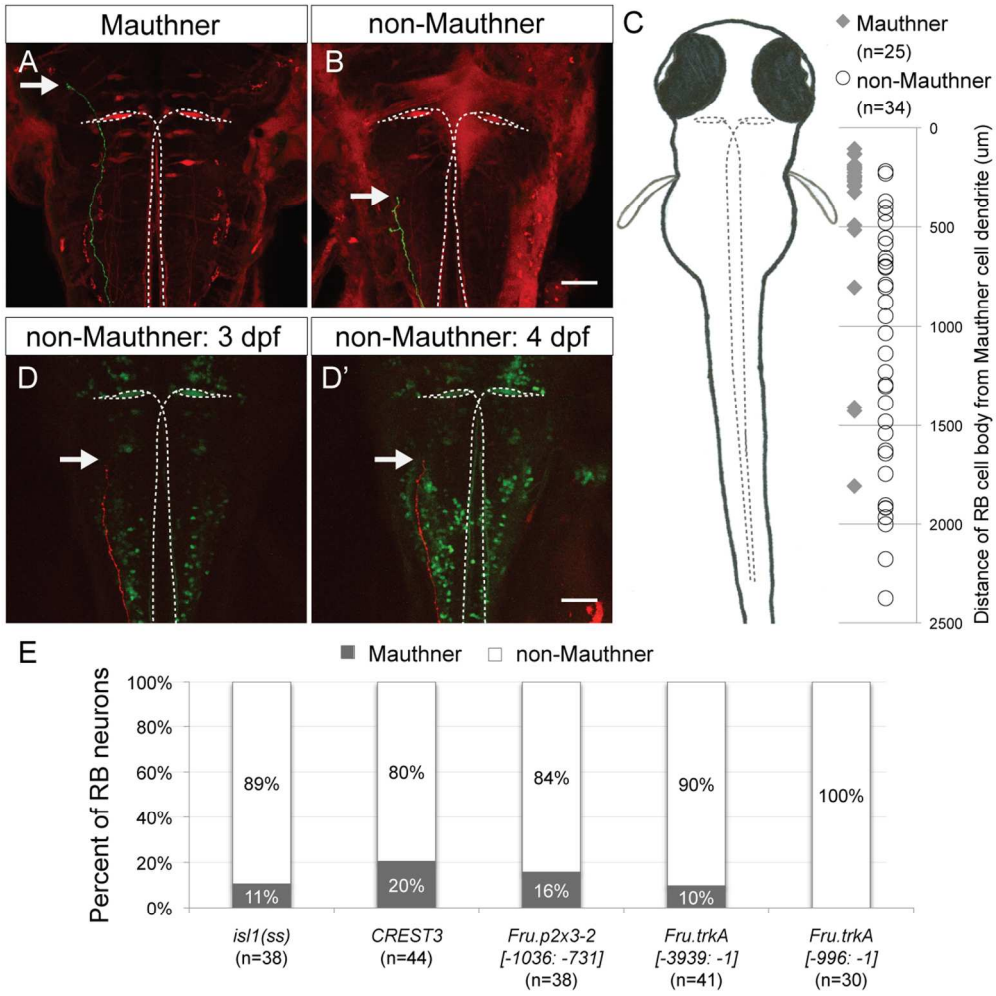
112x165mm (250 x 250 DPI)



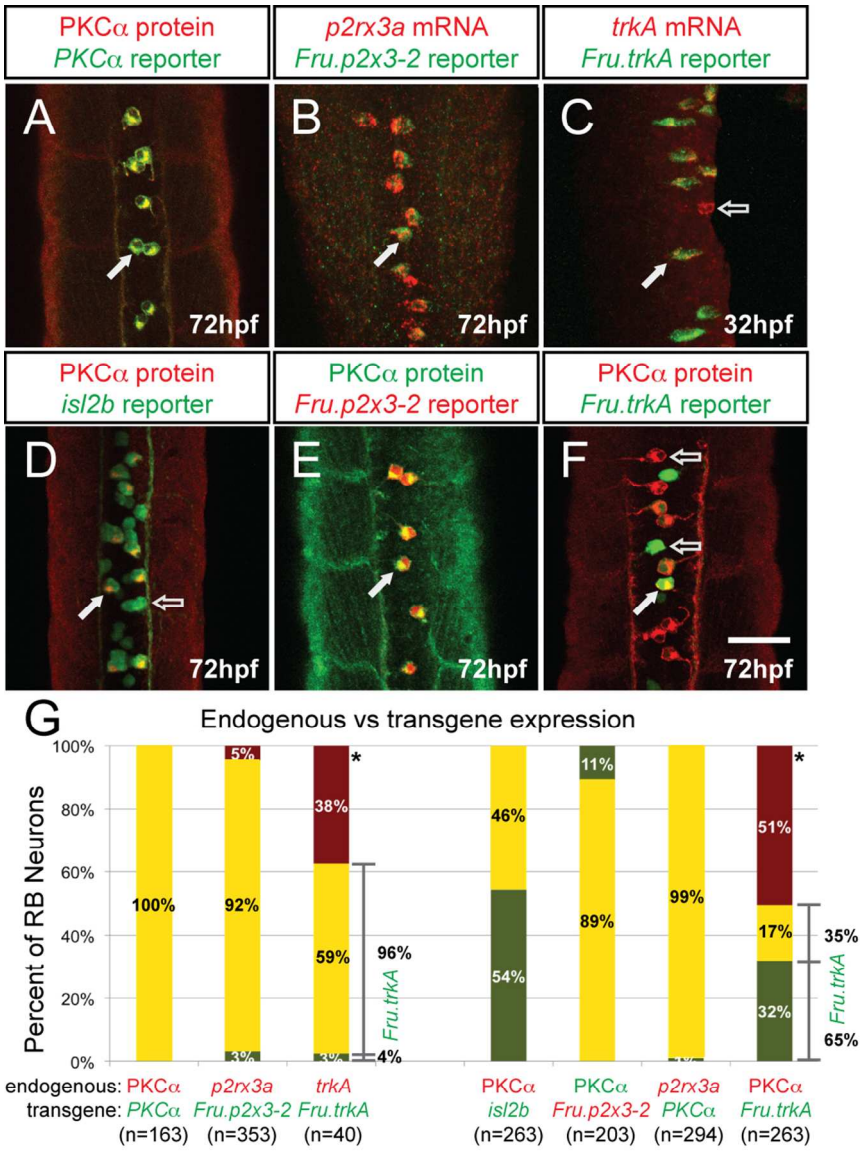
250x317mm (200 x 200 DPI)



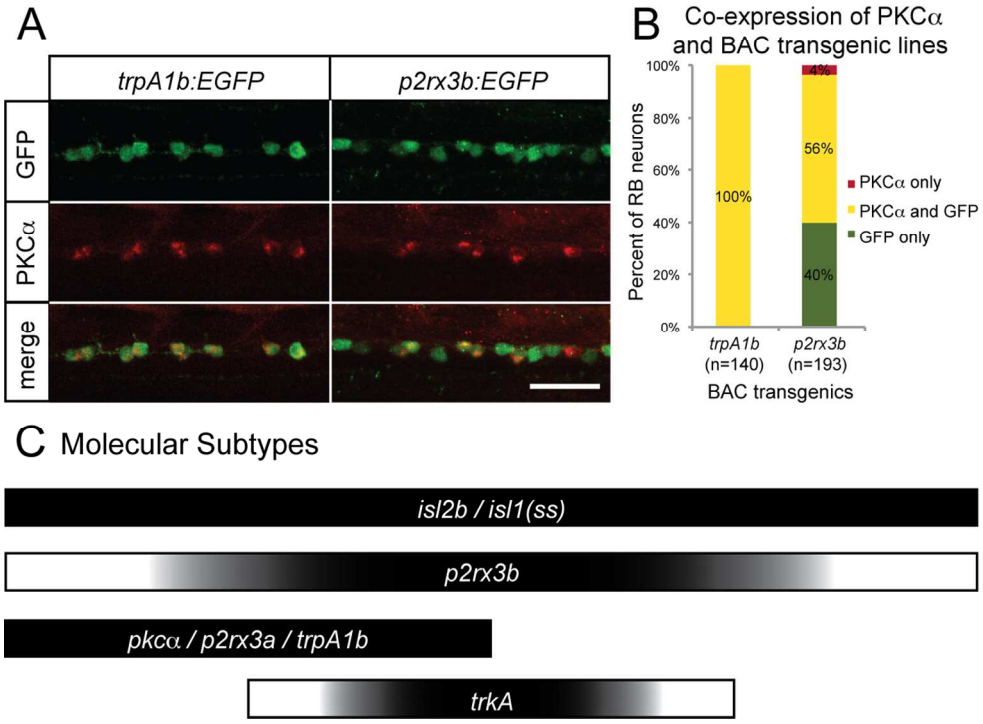
136x116mm (275 x 275 DPI)



424x427mm (80 x 80 DPI)

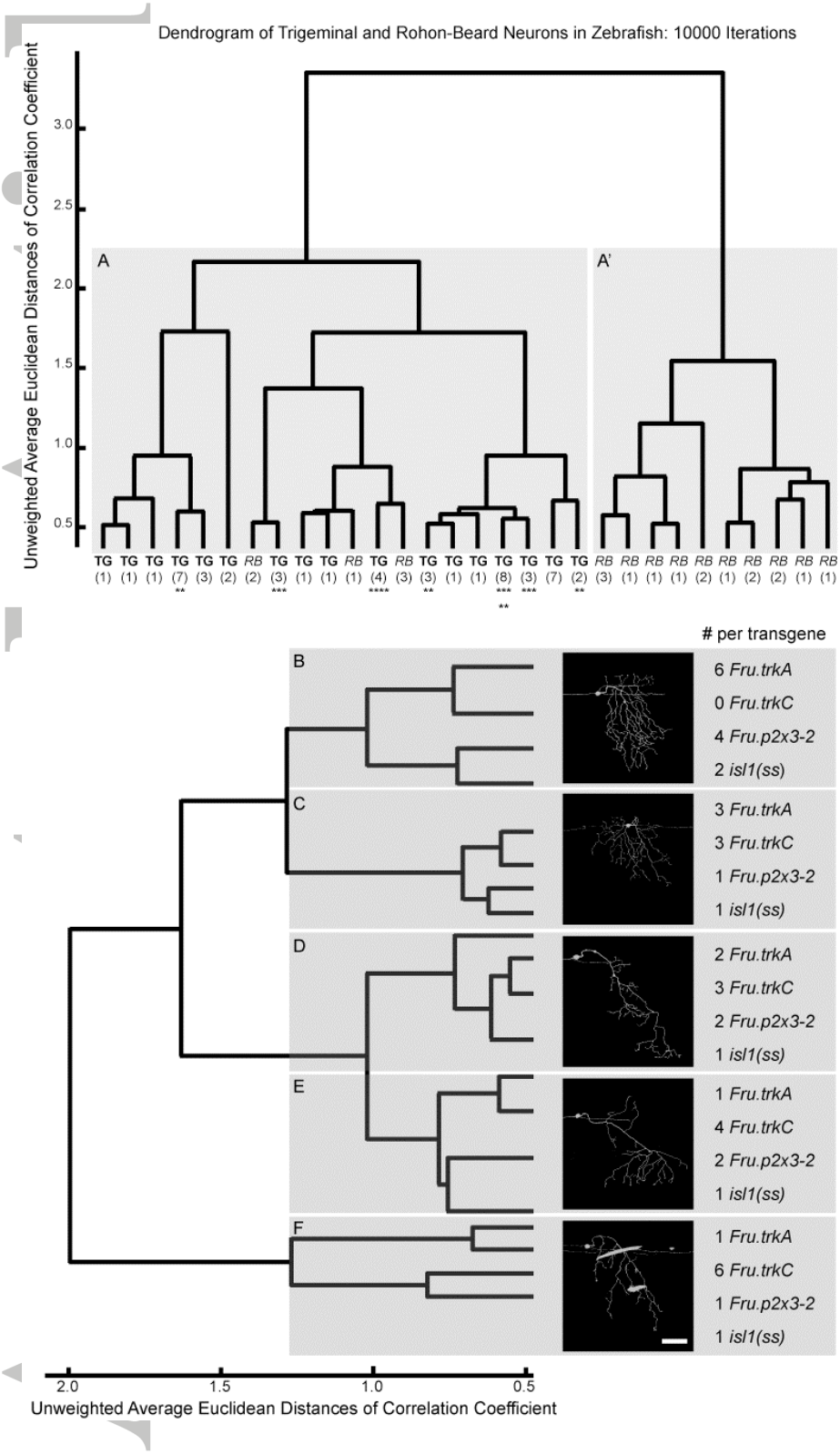


104x146mm (250 x 250 DPI)



144x111mm (250 x 250 DPI)

Supplementary Materials



Supplementary Figure 1. Categorization of peripheral axon morphology by branch number and length. Peripheral axons of single RB neurons labeled by transient transgenesis using reporter transgenes with *Fru.trkA*, *Fru.trkC*, *Fru.p2x3-2* or *isl1(ss)* enhancer sequences were hand-traced with Neurolucida Software and analyzed based on branch number and length for each branch order. Trigeminal (A) and Rohon-Beard (A') neurons cluster separately. Y-axis represents relative distance between groups. Asterisks represent the number of times the same neuron, traced by different experimenters, appears in a cluster. (B-F) Peripheral axons of individual neurons segregated into 5 main clusters from more complex (B) to less complex (F). Representative lateral confocal images of 72 hpf RB neurons are shown for each group. Dorsal is up; anterior is right. The number of RB arbors from each transgene within each cluster is reported on the right. Scale on bottom represents relative distance between categories. Scale bar, 100μm.

Supplementary Table 1: List of transgenic/gene trap lines and other transgenes.

Transgenic/Gene trap lines	Allele numbers
Tg(isl1:Gal4-VP16,UAS:EGFP), previously *sensory:GFP*	zf154
Tg(isl1:Gal4-VP16,UAS:dsRed), previously *sensory:DsRed*	Zf234
Tg(isl1(ss):Gal4-VP16,14xUAS:KikGR)	LA203
Tg(CREST3:Gal4-VP16,14xUAS:EGFP)	LA204
Tg(CREST3:Gal4-VP16,14xUAS:KikGR)	LA205
Tg(Fru.trkA:Gal4-VP16,14xUAS:ChR2-YFP)	LA206
Tg(Fru.p2x3-2[-1036:-731]:LexA-VP16,4xLexAop:mCherry)	LA207
Tg(-17.6isl2b:GFP)	zc7
Gt(PKCα-citrine), herein called *PKCα^{ct7a}*	ct7a
Gt(T2KSAG)	j1229a
Tg(trpA1b:EGFP)	a4593
Tg(p2rx3b:EGFP)	sl1
Additional Transgenes	
CREST3:LexA-VP16,4xLexAop:mCherry	
trpA1a:Gal4-VP16,14xUAS:KikGR	
Fru.trkA:Gal4-VP16,14xUAS:GFP	
Fru.trkA:LexA-VP16,4xLexAop:mCherry	
Fru.trkB: Gal4-VP16,14xUAS:GFP	
Fru.trkC:Gal4-VP16,14xUAS:GFP	
Fru.p2x3-2: Gal4-VP16,14xUAS:GFP	

Supplementary Table 2. List of primers for 5' enhancer elements. Primer name (right column) and primer sequence (left column) are provided for each enhancer sequence. Lower case letters in primer sequence indicate attB sequences, uppercase letters are specific to the enhancer sequence.

Primer name	Primer Sequence
attB4-*CREST3* Forward	ggggacaactttgtatagaaaagttgGTAACAGGATGTGACACGTCGTCTGC
attB1-*CREST3* Reverse	ggggactgctttttttgtacaaacttgGCCTGCTGCTGGTGTCATTTACTGG
attB4-*trpA1a* Forward	ggggacaactttgtatagaaaagttgAACCTATTGCACTTGTATCAGCAG
attB1-*trpA1a* Reverse	ggggactgctttttttgtacaaacttgGGGCCATGAAGAAATTCTGA
attB4-*Fru.trkA* Forward	ggggacaactttgtatagaaaagttgGTTCCTCATTGGAACAACACC
attB1-*Fru.trkA* Reverse	ggggactgctttttttgtacaaacttgACTGTCGGGAAACAGGACAG
attB4-*Fru.trkB* Forward	ggggacaactttgtatagaaaagttgTCAAGGCTTTGCTCACATGC
attB1-*Fru.trkB* Reverse	ggggactgctttttttgtacaaacttgGTTTGAGGAGCCACAACACTC
attB4-*Fru.trkC* Forward	ggggacaactttgtatagaaaagttgGACACTGTAATTGCTTCGACTG
attB1-*Fru.trkC* Reverse	ggggactgctttttttgtacaaacttgTTTTCTGCAGTGCGTCAGCAG
attB4-*Fru.p2x3-2* Forward	ggggacaactttgtatagaaaagttgCACCACTTTCGGAGGTGTCT
attB1-*Fru.p2x3-2* Reverse	ggggactgctttttttgtacaaacttgGTCAGTGTGCACCAGAGAGC

Supplementary Table 3: List of antibodies used for whole mount antibody staining.

Primary antibodies	Antigen	Source	Dilution used
PKCα	Peptide mapping at the C-terminus of human PKCα	Rabbit polyclonal antibody from Santa Cruz Biotechnologies, Inc. (sc-208)	1:500
GFP/YFP/Citrine	Full length Aequorea Victoria GFP	Mouse monoclonal antibody from Clontech (Living Colors JL-8)	1:500
DsRed/mCherry	Full-length DsRed2	Mouse monoclonal antibody from Clontech (Living Colors 632393)	1:500
Secondary antibodies	**Fluorophore**	**Source**	**Dilution used**
Goat anti-mouse IgG (H+L)	AlexaFluor 488	Molecular Probes (A11001)	1:1000
Goat anti-mouse IgG (H+L)	AlexaFluor 568	Molecular Probes (A11004)	1:1000
Goat anti-rabbit IgG (H+L)	AlexaFluor 488	Molecular Probes (A11008)	1:1000
Goat anti-rabbit IgG (H+L)	AlexaFluor 568	Molecular Probes (A11011)	1:1000

Supplementary Methods

Extracting Data from Neurolucida Explorer

Individual neurons were traced in 3 dimensions using NeuroLucida software. Data from each traced image was extracted using NeuroLucida Explorer. Running the Segment analysis generated an Excel table showing, among other characteristics, the branch order and length of each segment. Branches were ordered in this matrix so that the bottom-most branch shown was the root branch, from beginning to the first node. The next branch up was a branch of order 2, off of the root branch. The next one shown was a branch of order 3, and so on. Upon encountering a branch of order x with a normal ending (i.e., one that did not split further), the program went back to the nearest node y. The next segment displayed was the other branch of order x that originated at y. After encountering the next ending, the program went back to the nearest node from which there was an un-displayed branch and displayed that branch. The values for length and branch order of each segment were copied, preserving the original order in Excel, into a new worksheet, which was then saved as a .csv file. The data also went into a spreadsheet titled worksheet Data.xls. Both the .csv file and the spreadsheet were labeled with the name of the tracing. (The contents of the Data.xls spreadsheet do not matter; however, the name of the spreadsheet must match the name of the .csv file.) All the data was then imported into a Matlab program we developed called Comparing_Distance_Matrices3.

Creating the Distance Matrix

After importing the Excel data, a distance matrix was created for each using the Matlab function DistanceMatrixFinal. Given a branch A and its order x, the distance from branch A to a

branch B of order y (assuming without loss of generality that B is above A in the Excel table) is the sum of the lengths of the branches between B and A. The Excel table was used to extract the lengths of these branches for any given branches A and B. The start of the root branch, every node, and every ending on the neuron was counted as a point and numbered. The start of the root branch is 1, the end of the root branch 2, the ends of the second-order branches 3 and 4, and so forth. Each entry (i, j) of the final distance matrix represents the distance between points i and j in that particular tracing. A more detailed description of the methodology in creating this distance matrix follows.

When the Excel data was imported into DistanceMatrixFinal, it was flipped vertically, so that the last row of the table was the first row of the input matrix, the second-to-last row of the table was the second row of the input matrix, and so on. To find the branches between branches A and B, it was assumed first that B was below A in the input matrix (i.e., B was above A in the Excel table). First, considering the case where there was no branch of order less than A's between A and B. The lengths of the relevant branches were extracted assuming that every branch between A and B in the tracing must be between A and B in the table and have an order between x and y. The relevant branch for each of these orders is the one closest to A in the matrix. These lengths were then grouped together in a three-dimensional matrix R. Given a branch i of order x and a branch j (above i) of order y, R(i, j, :) contains the lengths of each branch of order between x and y that is closest to row i in the Excel matrix.

The matrix T that the contains the distance matrix, formed by summing the entries in each R(i, j, :). As stated above, the start of the root branch, every node, and every ending on the neuron was counted as a point, so the # of points = 1 + # of branches. Initially, the point at the end of the branch in row i was assigned to be represented by the i+1th row/column in the

distance matrix. Point 1 was the start of the root branch. The upper half of T was formed by summing the (i, j, :) entries of R and assigning that value to the (i, j+1) entry of T. The diagonal was assigned to be all zeros, because the distance between a point and itself is zero. In the case where going from point x to point y meant "backtracking" along a branch of lower order, T(x, y) was assigned to be zero, to be corrected as described below.

If there was a branch between the ith and jth rows with order less than the order of the jth row, then R(i, j, :) also represented the lengths of the branches between i and j. However, a slightly more complicated method was necessary to find these branches. The lengths of those branches between branch i and the branch of minimum order between i and j (call it branch k) were found using the same method as mentioned above for branches where no backtracking was necessary. The same was done for the branches between branches i and k. All of these branches were then entered into R(i, j, :). The sum of the entries in R(i, j, :) was then assigned to T(i+1, j+1). The upper half of T was then complete, and the lower half formed by assigning T(i, j) to equal T(j, i).

Finally, the numbers of the points in the tracing were reassigned, shuffling T. The start of the root branch was point 1. The end of the root branch was point 2, the ends of the second-order branches 3 and 4, and so forth, as described above. The new number of each point was found and assigned the old entry of T to its new coordinates.

Comparing the Distance Matrices

Comparing_Distance_Matrices3 takes the tracings and compares each pair of them to generate a correlation coefficient. All correlation coefficients were put in a matrix, where each

(i, j) entry is the correlation coefficient between matrices i and j. The correlation between a matrix and itself is 1, and all other coefficients range from 0 to the ratio of the sizes of the matrices. The matrices are numbered according to the order of their spreadsheets in Data.xls. Only the upper half of the correlation coefficient matrix was computed, since the (i, j) and the (j, i) correlation coefficients should be the same, the lower half of the matrix was assigned to be the upper half reflected over the diagonal. Each pair of matrices was compared by randomly choosing and shuffling rows and columns in a Monte Carlo simulation. To fix the primary branch (i.e., the first row and column for each matrix), we replace rTest1 in the program with rTest2. The program does this for a chosen number of iterations (this analysis was performed with 2500, 3000, 5000, 10000 iterations) and outputs the maximum global correlation coefficient. Each coefficient is then multiplied by the ratio of the sizes of the distance matrices being compared. The size of the distance matrix depends directly on the number of branches in the tracing, so this variable as well as the branch length is considered in the analysis.

Constructing the Dendrogram

Once the matrix of correlation coefficients was computed, Matlab was used to find the Euclidean distances between each pair of coefficients in the correlation coefficient matrix. Next, Matlab clustered these pairs, treating each row of the coefficient matrix as a row of observations (the coefficient when matrix i is compared with matrix 1, when compared with matrix 2, etc.). Thus, each row vector i "represents" the matrix i. Three different methods of clustering the pairs were used. The unweighted average distance used the average distance between all pairs of objects in the two clusters being combined. These clusters may be single or already-combined

row vectors. The weighted average method does the same, only using weighted instead of unweighted averages (the distances are weighted by the size of the clusters being combined). Ward's method uses the incremental sum-of-squares, meaning the increase in the total sum of squares as a result of joining the two clusters. The sum of squares is defined as the sum of the squares of the distance between all entries in the row vector and the centroid of that vector.

Each of these methods produced a matrix that Matlab converted to a dendrogram with at most thirty leaf nodes; since this analysis had seventy tracings, several nodes represented more than one distance matrix. The indexing of the dendrogram function was used to determine the names of the tracings whose distance matrices were represented by each node.

Brief Summary of Results

Given the similarity between the correlation coefficient matrices produced, the number of iterations did not appear to be important. The correlations between the matrices all rounded to 0.99, except for those involving the 5000-iteration matrix (0.93, 0.90, and 0.95) and were probably smaller because the 5000-iteration matrix was larger than the other correlation matrices (since the 5000 analysis included 5 secondary trees).

The Ward method seemed to be slightly better for producing dendrograms. Both the average and unweighted average methods clustered up to eight matrices into one node, for every number of iterations. However, the Ward method clustered a maximum of six matrices per node for the 2500 and 3000 iterations, eight for the 5000 iterations, and only five for the 10000 iterations. This difference is likely not too important, though clustering too many matrices per node indicates more random clustering, since one would expect fairly even clustering.

One of the controls to test this program's accuracy was separating Rohon-Beard and trigeminal neurons. All of the dendrograms divided the tracings overall into two classes, which generally conformed to the Rohon-Beard/trigeminal division. There was some overlap, but it appeared to be fairly consistent between all of the dendrograms.

Another control was analyzing different tracings of the same neuron to see how similar they would be. Most of these tracings were consistently clustered together; for example, KikGR_Tg-8's different tracings were clustered together in nearly all of the dendrograms, and were in adjacent clusters for the other dendrograms. KikGR_Tg-9 had similar results. Some different tracings of the same neuron did not appear together; this may be due to error on the tracer's part. For example, KikGR_Tg-6A and M always were clustered together, but KikGR_Tg-6C was always in a separate cluster.