CHAPTER IV

DEVELOPMENT AND OPTIMIZATION OF CELL-SPECIFIC PROFILING STRATEGIES IN C. ELEGANS

Introduction

The whole animal microarray experiments discussed in Chapter III were insufficiently sensitive to detect changes in transcript levels in only 5% of cells. A higher resolution strategy, mRNA-tagging, was introduced at the 2001 International Worm Meeting by Peter Roy. In this method, an epitope-tagged poly-A binding protein (FLAG::PAB-1) is driven by a cell-specific promoter in the cell-type of interest. Cell-specific transcripts are then isolated by immunoprecipitation of the epitope tag. mRNA-tagging was initially used to identify muscle-enriched transcripts (Roy et al. 2002).

I learned of this method while performing my whole-animal microarray experiments at Stanford. Dr. Roy provided his unpublished vector and protocol for our effort to isolate VA-specific mRNAs. This chapter will describe the successful optimization of the mRNA-tagging method and our use of the technique to create gene-expression profiles for the entire nervous system of *C. elegans* as well as for VA motor neurons. Additionally, I will describe our efforts to find downstream UNC-4 targets by comparing transcripts isolated from VA motor neurons in wildtype and *unc-37* mutant animals.

For these experiment, we opted to use the newly available Affymetrix *C. elegans* arrays. The Affymetrix array is produced differently than a printed array. Instead of generating DNA (usually by PCR amplification) and then spotting it on glass, Affymetrix

synthesizes 11mer oligos directly on a solid substrate using a patented photolithographic manufacturing process (see www.affymetrix.com for more details). In this way, the spot size and amount of nucleic acid/spot are almost invariant between chips, thus reducing the experimental error from chip to chip. Additionally, since only one RNA species is hybridized per chip, the amount of baseline mRNA needed is reduced. Moreover, a major benefit of switching to this platform enabled us to perform all experiments at Vanderbilt in the Vanderbilt Microarray Shared Resource (VMSR).

Materials and Methods

Nematode strains

Nematodes were grown as described (Brenner 1974). Strains used to isolate transcripts via mRNA-tagging were N2, SD1241 (gaIs153, *F25B3.3*::FLAG::PAB-1), NC694 (wdEx257, *unc-4*::3XFLAG::PAB-1), NC714 (*unc-37 (e262)*; wdEx257), and NC717 (*unc-4 (e120);* wdEx257).

Molecular biology

To create a vector to express 3XFLAG::PAB-1 in *unc-4* neurons, I used PCR to amplify the 3XFLAG from p3XFLAG-CMV7.1 (Sigma). The primers used added a KpnI site and an XhoI site, allowing me to exchange the 1XFLAG in pPRSK9 (*myo-3*::FLAG::PAB-1, Peter Roy) for the 3XFLAG to generate pSV14. Sequencing of pSV14 confirmed that the 3XFLAG and PAB-1 were in-frame. The 3.2 kb 3XFLAG::PAB-1 fragment was subcloned from pSV14 into the cloning vector pSL1180 (Amersham) using

KpnI and SacI to create pSV15. pSV9-TOPO contains the 3 kb PCR-amplified (primers U4p49-F and U4p43-R, added NheI and KpnI sites) *unc-4* promoter ligated into pCR2.1-TOPO. Using the aforementioned sites, the *unc-4* promoter was subcloned into pSV15 to generate pSV16. This *unc-4*::3XFLAG::PAB-1 vector was made suitable for microparticle bombardment (see below) by ligating the *unc-119* minigene from plasmid MM051 (gift from D. Pilgrim), digested with HindIII (filled in with T4 Polymerase) and BamHI, into the EcoRV and BamHI sites. The resulting plasmid (pSV17) contained the *unc-119* minigene in the opposite orientation to the *unc-4*::3XFLAG::PAB-1 fragment.

Biolistic transformation

Microparticle bombardment was conducted (Praitis et al. 2001) in a BioRad Biolistic PDS-1000/He equipped with the Hepta Adapter. Gold beads (1 micron) were coated with DNA at 1 ug/ul. 100 mm NGM plates were seeded with a monolayer of ~100,000 L4/adult *unc-119 (ed3)* animals. For each construct, 1 'shot' was performed using a 1550 psi rupture disk at 28 inches of Hg vacuum. After a 1 hr recovery period, animals were washed from the plates with 7 ml M9 buffer and transferred to 7 NGM plates (1 ml/plate). Animals were grown at 20 ºC for 1 week. To pick transgenic animals, one-half of the plate was 'chunked' and added to a new 100 mm NGM plate; animals with wildtype movement were picked to 60 mm NGM plates and allowed to self. Worms derived from separate plates were considered independent lines; at least 2 lines were obtained for each construct.

Generating synchronized populations of L2 larvae

Strains were grown to starvation on ten 60 mm NGM plates at 25ºC. Half of a 60 mm plate ("chunked" into four pieces) was placed on a 150 mm 8P plate. These twenty 8P plates were incubated at 25ºC until a majority of the food was gone and many animals were gravid adults (a "line" of worms is usually found at the retreating edge of the bacteria). The worms were washed off the 20 x 8P plates using ice-cold M9 buffer and transferred to four 50 ml conical tubes. Animals were collected by centrifugation (3000 rpm, 3 min, 4 ºC in Beckman 7.5 rotor) and pooled. Worms were washed with ice cold M9 buffer until the supernatant was clear of bacteria. A sucrose float (30 ml ice cold M9 buffer, 20 ml cold 70% sucrose) was performed to create an axenic nematode suspension. Animals were washed 2X in ice cold M9 buffer, then resuspended in 75 ml bleach solution (15 ml Chlorox, 3.75 ml 10N NaOH, 56.25 ml water). Worms were transferred to a 125 ml glass beaker with a stir bar and incubated for 5-6 min while stirring fast (solution turns a dark yellow when nearing completion). When a majority of adults burst, the solution was passed through a 53 µm nylon mesh (Fisher #08-670-201) to separate the embryos from worm carcasses. The filtrate was split into four 50 ml conicals and brought up to 50 ml with M9 buffer. Worms were collected by centrifugation (clinical centrifuge, 2500 rpm, 2.5 min). Embryos were pooled and washed 3x with M9 buffer. The suspension was brought up to 50 ml with RT M9 buffer, then split equally into two 50 ml conical tubes. Embryos were incubated on a nutator for 12-16 hours at 20 ºC to allow L1 larvae to hatch and 'hang up.'

L1 larvae were collected by centrifugation (3000 rpm, 3 min, 4 ºC in Beckman 7.5 rotor). The supernatant was discarded and animals were resuspended in 1 ml RT M9

buffer. Each ml of L1s was split over three 150mm 8P plates. The tubes were washed with 0.5 ml M9 and added to the 150 mm plates. L1s were grown at 20 ºC for 22-25 hours to reach mid-L2, as evidenced by the appearance of the post-deirid sensory organ (usually all animals are L2s, with ≥ 80% having a post-deirid). L2s (~0.3 ml - 1 ml) were harvested from 8P plates and sucrose floated as above. Worms were resuspended in 30 ml cold M9.

Generating worm lysates

To fix worms, 0.75 ml 20% formaldehyde [weigh out 400-500 mg paraformaldehyde, resuspend in 4.5X(weight in mg) µl 5 mM NaOH, and incubate at 65ºC until dissolved] was added to 30 ml synchronized L2 larvae and incubated for 1 hour on nutator at 4 ºC (final formaldehyde concentration is 0.5%). Animals were collected by centrifugation (2500 rpm, 2.5 min, in clinical centrifuge). L2s were washed with 50 ml M9, then equilibrated with 30 ml homogenization buffer (HB) [50 mM HEPES, pH 7.6; 150 mM NaCl, 10 mM $MgCl_2$; 1 mM EGTA pH 8.0; 15 mM EDTA pH 8.0; 0.6 mg/ml Heparin; 10% glycerol; treat with 1% DEPC to make RNase-free]. On the day of use, add vanadyl ribonucleoside complex (Sigma) to 8 mM (final), DTT to 1 mM; 50U rRNasin (Promega)/ml HB, 1/2 protease inhibitor tablet (EDTA-free, Roche)]. The supernatant was removed, and animals were resuspended in 2-3 ml HB (if packed worms are ~1 ml, use 3 ml HB; otherwise use 2-2.5 ml HB). Animal carcasses were disrupted by passing the suspension through a mini French press cell (Thermo-Spectronic #FA-003) 3X at 6000 psi. Worms were further homogenized (30X in baked glass Dounce homogenizer). The lysate was centrifuged at 5000 rpm (~4000 x*g*) for 6 min at 4ºC. The

resulting supernatant was then centrifuged at 11,750 rpm (~20,000 x*g*) for 20 min at 4ºC. The final supernatant (cell-free extract) was collected and stored in 0.4-0.5 ml aliquots (0.1 ml was removed for total RNA extraction, see below). Lysates were flash-frozen in liquid nitrogen and stored at -80ºC until needed for immunoprecipitation.

Extracting total RNA from worm lysates

In this step, RNA is extracted from a small aliquot of the worm lysate in order to estimate the amount of total RNA in the preparation. Thaw 0.1 ml worm lysate. Add 4X vol (0.4 ml) TriZOL; incubate at RT for 5 min. Vortex, then add 1X vol (0.1 ml) chloroform; incubate at RT for 10 min. Vortex, then centrifuge at max speed (14000 rpm) for 15 min at 4ºC. Remove upper, aqueous layer (should be clear) to new centrifuge tube and add an equal volume of chloroform. Centrifuge at max speed for 10 min at 4ºC. Remove upper, aqueous layer to new centrifuge tube and add equal volume isopropanol. Incubate at RT for 10 min, then centrifuge at max speed for 25 min at 4ºC. Wash 2X 1 ml 70% ethanol. Use aspirator to carefully remove the supernatant and let RNA pellet air dry. Resuspend in 30 µl DEPC-water. Make a 1:50 dilution in 98 µl of DEPC-water and place in a microcuvette to determine the concentration using the Miller lab spectrophotometer (A260). Calculate the amount of total RNA/µl lysate (Example: if RNA = 1 µg/µl, then in the 30 µl there is 30 µg total RNA. Therefore 30 µg total RNA/100 µl lysate = 0.3 µg RNA/µl lysate).

Co-immunoprecipitation

Equilibrate ~100 µl anti-FLAG Sepharose EZView beads (Sigma) by washing 2X

with 1 ml 0.1M glycine (pH 3.5) and 4X with 1 ml HB (all spins are at 3500 rpm for 3 min). Thaw the amount of lysate that is equivalent to 200 μg of total RNA (e.g., if lysate = 0.3 μg RNA/μl, then 200 μg total RNA = 667 μl lysate) and add to equilibrated beads. Add HB (to 1 ml), 8 μl vanadyl ribonucleoside complex and 10 μl rRNasin. Incubate on nutator for 2 hours at 4ºC. Pellet beads and perform 3x quick washes with 2 ml low-salt homogenization buffer (lsHB) [20 mM HEPES, pH 7.6; 25 mM NaCl; 1 mM EGTA, pH 8.0; 1 mM EDTA, pH 8.0; 0.6 mg/ml Heparin; 10% glycerol. Treat with 1% DEPC. On the day of use add vanadyl ribonucleoside complex to 8 mM, DTT to 1 mM, and 50U rRNasin/ml lsHB]. Perform longer washes 3X by incubating beads in 2 ml lsHB on nutator for 30 min at 4ºC. Add 125 μl elution buffer [50 mM Tris pH 8.0; 10 mM EDTA, pH 8.0; 1.3% SDS; use DEPC-water to ensure solutions are RNase-free] to pelleted beads and incubate at 65ºC for 30 min (to elute protein from beads and reverse cross-link between protein and mRNA). Collect beads by centrifugation and transfer supernatant to centrifuge tube on ice. Elute again as above and pool eluates. TriZOL extract, isopropanol precipitate, and wash pellet as above, except use 1 ml TriZOL and 0.25 ml chloroform for first extraction. Resuspend dry pellet in 12 μl DEPC-water. Use spectrophotometer to determine concentration. Use Agilent Bioanalyzer (Vanderbilt Microarray Shared Resource, VMSR) to assay RNA quality.

RNA Amplification (adapted from the Affymetrix GeneChip Eukaryotic Small Sample Target Labeling Protocol)

(All incubations are performed in a PCR machine unless otherwise indicated).

First Round

Add 1 μl of Affy 100 primer (5'-GGCCAGTGAATTGTAATACGACTCACT

ATAGGGAGGCGG-$(dT)_{24}$ -3', 100 ng/µl) to 25 ng of immunoprecipitated mRNA, then bring up to 10 µl with DEPC-water. Heat at 70ºC for 10 min. Cool on ice for 2 min, then add 4 µl 5X 1st strand buffer (Invitrogen), 2 µl 0.1M DTT (Invitrogen), 1 µl 10 mM dNTP mix (NEB), and 1 µl rRNasin (Promega). Heat at 42ºC for 2 min, then add 1 µl Superscript II (Invitrogen). Incubate at 42ºC for 1 hour. Cool on ice, then add 30 µl 2nd strand buffer (Invitrogen), 3 µl 10 mM dNTP mix, 1 µl *E. coli* DNA ligase (Invitrogen), 4 µl *E. coli* DNA Polymerase I (Invitrogen), and 1 µl RNase H (Invitrogen). Add water up to 150 µl, then incubate at 16ºC for 2 hours. Add 2 µl T4 DNA Polymerase (Invitrogen) and incubate at 16ºC for 15 min. Precipitate cDNA by adding 0.6 vol cold 5M sodium acetate and 2.5 volumes cold 100% ethanol. Centrifuge at max speed for 20 min at 4ºC. Wash 1X cold 70% ethanol. Speed vac to dry pellet (can store dry pellet at -20ºC at this point). Resuspend pellet in 8 µl DEPC-water and incubate in dry block at 37ºC for 15 min. Using components from Ambion T7 Megascript kit, add 8 µl dNTP mix, 2 µl reaction buffer, and 2 µl Ambion enzyme mix. Incubate at 37ºC for 6 hours. Use Qiagen RNeasy mini-kit to clean RNA. Follow manufacturer's directions, except use 300 µl 100% ethanol and elute with 32 µl DEPC-water. Use spectrophotometer to quantitate. If the aRNA is less than 10 ng/µl, speed vac to reduce volume to 10 µl. Store at -80ºC until ready to perform the second round of amplification.

Second Round

Use 100 ng first round cRNA (or all if volume reduced to 10 µl). Add 1 µl of random primers (Invitrogen, 1 µg/µl) and DEPC-water up to 11 µl. Heat at 70ºC for 10 min, then cool on ice 2 min. Perform first strand synthesis as above. Add 1 µl RNase H;

incubate at 37ºC for 20 min and 95ºC for 5 min. Add 1 µl Affy 100 primer and incubate for 10 min at 70ºC. Cool on ice, then add 30 µl second strand buffer, 3 µl 10 mM dNTP mix, and 4 µl *E. coli* DNA Polymerase I. Incubate at 16ºC for 2 hours, then add 2 µl T4 DNA Polymerase and incubate for 15 min at 16ºC. Precipitate cDNA as above (can store dry pellet at -20ºC at this point). Resuspend pellet in 22 µl DEPC-water and incubate for 15 min in 37ºC dry block. Using components from Enzo BioArray HighYield RNA Transcript Labeling kit, add 4 µl 10X HY reaction buffer, 4 µl 10X Biotin labeled ribonucleotides, 4 µl 10X DTT, 4 µl 10X RNase inhibitor mix, 2 µl 20X T7 RNA polymerase. Incubate at 37ºC for 6 hours. Clean aRNA using Qiagen RNeasy mini-kit as above. Quantitate aRNA using spectrophotometer. Remove a portion of amplified aRNA for bioanalyzer. Fragment 15 µg aRNA by adding 6 µl fragmentation buffer (available from Affymetrix) and water up to 30 µl. Run pre- and post-fragmented aRNA on Agilent bioanalyzer (VMSR) to assay quality.

Hybridization

All hybridizations were performed by the VMSR according to standard Affymetrix protocols. results. 15 µg of aRNA was hybridized.

Data Analysis (adapted from Fox, et. al., 2005)

The commercially available *C. elegans* Affymetrix array was used for all experiments. This chip was designed using the December 2000 genome sequence. All probe set information is available at www.affymetrix.com as well as www.wormbase.org. The mRNA-tagging datasets consist of three independent immunoprecipitations from

experimental (SD1241, NC694, NC717, see nematode strains above) and baseline (N2) lysates. Hybridization intensities for each experiment were scaled in comparison to a global average signal from the same array (Hill et al. 2001). Expressed transcripts were initially identified on the basis of a "Present" call in a majority of experiments (2/3 for SD1241 immunoprecipitations (IPs), 3/4 for NC694 IPs, and 4/5 for N2 IPs) as determined by Affymetrix MAS 5.0 (using a Perl script written by Kellen Olszewski). In this approach, a Mismatch (MM) value for each feature is compared to a Perfect Match (PM) value to estimate non-specific binding. This strategy, however, tends to arbitrarily exclude low intensity signals in which PM and MM values may be comparable (Irizarry et al. 2003; Irizarry et al. 2003). To avoid this bias in the detection of transcripts that might be differentially elevated in the experimental data sets, intensity values were normalized using RMA (Robust Multi-Array Analysis) available through GeneTraffic (Iobion) in which the MM values are not considered (Irizarry et al. 2003; Irizarry et al. 2003). Comparisons of RMA normalized intensities for **experimental** (SD1241 IP, NC694 IP, NC717 IP) vs. **reference** (N2 IP) transcripts were statistically analyzed using Significance Analysis of Microarrays software (SAM, Stanford) (Tusher et al. 2001; Storey and Tibshirani 2003). A two-class unpaired analysis of the data was performed to identify genes that differ by $\geq$ 1.7-fold (*unc-37* target gene list) or $\geq$1.5-fold (profiling data) from the wildtype reference at a False Discovery Rate (FDR) of $\leq$ 1%. These genes were considered significantly enriched/upregulated.

To generate a list of all expressed genes, we subtracted genes depleted in the experimental datasets (i.e. enriched in reference) from the list of experimental genes called Present (see above). We attribute the presence of these baseline genes in the

experimental Present list due the background RNA which binds to Sepharose during the IP. We then added in genes which are considered statistically enriched by RMA/SAM analysis but not called Present by Affymetrix MAS5.0 (likely due to high mismatch signals).

Annotation of datasets (adapted from Fox, et al, 2005)

A WormBase mirror was established by downloading code and databases from www.wormbase.org. Using the acedb perl module, an annotation script was generated by Christian Shaffer that queries the WormBase mirror. Affymetrix IDs have been mapped to specific transcripts in WormBase. Text files containing Affy IDs (one per line) and cosmid names are input into the script that then searches the WormBase mirror and matches Affy ID/cosmid name to a specific transcript. Cosmid names are used for this search when Affy IDs have not been mapped in WormBase. This information is used to acquire other linked annotations (i.e. KOG, common name, RNAi phenotype, Expression data, Kim mountain data and Gene Ontology, etc.). All data presented here was annotated with WormBase release WS145.

Results

Generating a gene expression profile of the nervous system, part I

The main goal of my project was to identify UNC-4 target genes. To reach that goal we needed to create a transgenic that expressed an epitope-tagged PAB-1 in VA motor neurons (see below). However, we also needed to verify that we could utilize

mRNA-tagging to isolate neural-specific transcripts. To that end, we tested the method using a pan-neural FLAG::PAB-1 expressing strain (provided by Peter Roy) (Figure 4.1). We profiled the nervous system at mid-L2 as this is the developmental time at which UNC-4 function is require to maintain proper synaptic connections to VA motor neurons (Miller et al. 1992).

Our first efforts were promising. We co-immunoprecipitated 2-4 μg of mRNA reliably. Our reference dataset was mRNA isolated from staged wildtype L2 larvae, which contains transcripts from the average nematode larval cell. Upon amplification and fragmentation (see Materials and Methods), labeled aRNA was hybridized to the *C. elegans* Affymetrix chip. We opted to switch microarray platforms for the following reasons: 1) The hybridizations could be performed at Vanderbilt; 2) Affymetrix claimed to have reannotated the genome; thus more genes were represented on their platform vs. the Kim lab's chip; 3) Only one RNA sample is hybridized to a single chip, eliminating the need for large quantities of baseline mRNA; and 4) Variation from chip to chip is less than 2% according to Affymetrix data and our unpublished data (R. Fox, D. Miller).

After normalizing and statistically analyzing our pan-neural data (see Materials and Methods), we identified ~1200 enriched genes at a 1.7 fold cutoff and a False Discovery Rate (FDR) of less than or equal to 10.6% (FDR is an estimate of the number of false positive genes following SAM analysis). Of the genes annotated with expression patterns, 80% were explicitly stated to be in neurons. Several known neural genes were highly represented, including the pan-neural genes *goa-1* (Go-alpha) and *pkc-1* (PKC-Beta), and genes involved in synaptic vesicle release (*rab-3*, Rab GTPase; *snb-1*,

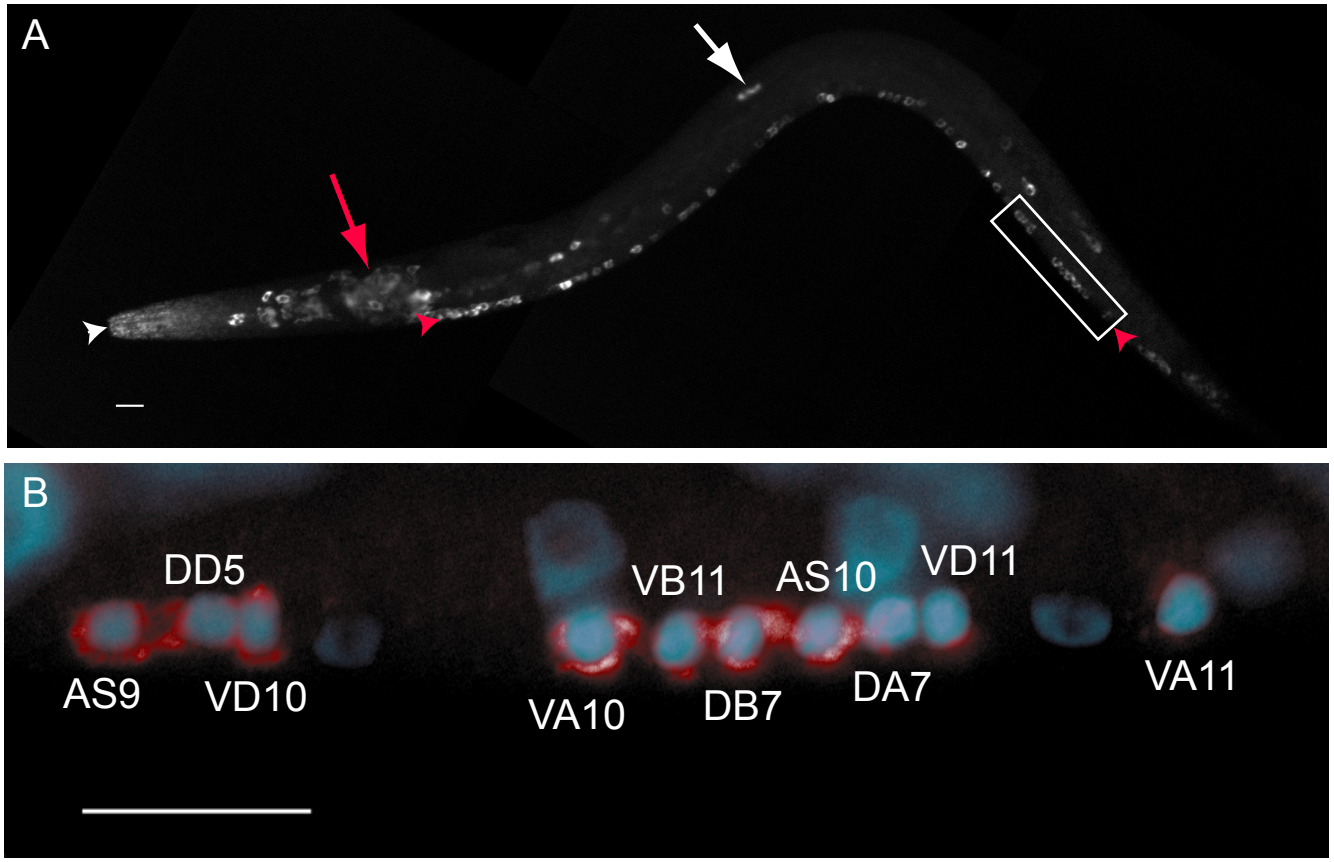

Figure 4.1. *F25B3.3*::FLAG::PAB-1 expression in L2 larvae.
A. Antibody staining detects FLAG::PAB-1 neurons (white circles) in head ganglia (red arrow), ventral nerve cord (red arrow-heads), and touch neurons (white arrow). Non-specific staining is seen at the tip of the nose (white arrowhead).
B. Close-up of ventral cord (boxed image in A), showing anti-FLAG staining (red) in cytoplasm surrounding nuclei (DAPI, blue).

Anterior is left, ventral is down. Scale bars = 10 µm

synaptobrevin; *snt-1*, synaptotagmin). However, we noted that several genes known to be highly expressed in other tissues were enriched, such as the muscle specific genes *pat-1* (Beta-integrin), *unc-15* (paramyosin), and *unc-54* (myosin). We considered two possibilities: 1). FLAG::PAB-1 was ectopically expressed in other tissues; and 2). RNA was sticking non-specifically to the anti-FLAG Sepharose beads.

Our antibody staining is largely specific to neurons, with some occasional faint expression in muscle. This result is therefore unlikely to account for the large amount of non-neural (and non-muscle) transcripts in this dataset. To assess non-specific binding to the anti-FLAG beads, my colleague Joseph Watson performed an immunoprecipitation from wildtype lysates (no expression of FLAG::PAB-1). Surprisingly, this "mock" IP pulled down a substantial amount (2-3 μg) of RNA. Bioanalysis of these control IP samples indicated that a majority of the nucleic acid was mRNA (data not shown). We therefore decided to use this control IP as a baseline in order to identify neural transcripts enriched over this background. We could also use this experiment to determine if RNAs from all tissues were sticking to the Sepharose or if a more biased sample was captured.

Analysis of the Present genes from these control IPs confirmed that transcripts from all tissue types were represented. For example, *dys-1* (muscle-specific), *ges-1* (intestine-specific), *tra-1* (intestine and gonad), *let-2* (body-wall muscle, coelomocytes, distal tip cells, spermatheca), *inx-6* (corpus of pharynx), and *unc-11* (~all neurons) are found in this list. Reanalyzing the pan-neural data using the control IP as a baseline substantially improved our results. We now identified ~650 neural enriched genes at a 1.7-fold cutoff and reduced our FDR by half (≤ 5%). This approach also eliminated the presence of the contaminating muscle genes. Furthermore, ~90% (102/108) of genes in

the dataset with known expression patterns are found in neurons. However, we felt that the method could be further improved by reducing the amount of non-specific binding to the beads. We posited that we would identify a deeper list of neural transcripts, including those which are not as highly represented (i.e. expressed in a subset of neurons). We therefore set out to optimize the mRNA-tagging method by removing this background RNA from the beads.

Optimizing mRNA-tagging

Our first efforts to remove background RNA were designed to test the idea that the RNA was sticking directly to Sepharose. We incubated wildtype lysate with "naked" Sepharose beads (without the anti-FLAG antibody attached). This experiment pulled down similar amounts of RNA (compared to our control IPs), leading us to believe that Sepharose was the basis for the non-specific binding. Unfortunately, Sepharose is the most highly used substrate for antibody-coupled beads. We researched other options but decided that our focus would be better spent on either preventing RNA from sticking to the Sepharose beads, or remove it via more stringent washes. To this end, we modified the levels of salt in the Homogenization Buffer (HB) washes used in these experiments (see Materials and Methods). We reasoned that the RNA might adhere to the Sepharose via hydrophobic interactions. It should follow that reducing the ionic strength of the homogenization buffer would inhibit these hydrophobic interactions and remove the RNA from the beads. These low-salt washes did result in a 40% reduction in background RNA. To determine if this modified procedure produced higher quality results, we hybridized these Low-Salt (LS) IP RNAs (described below).

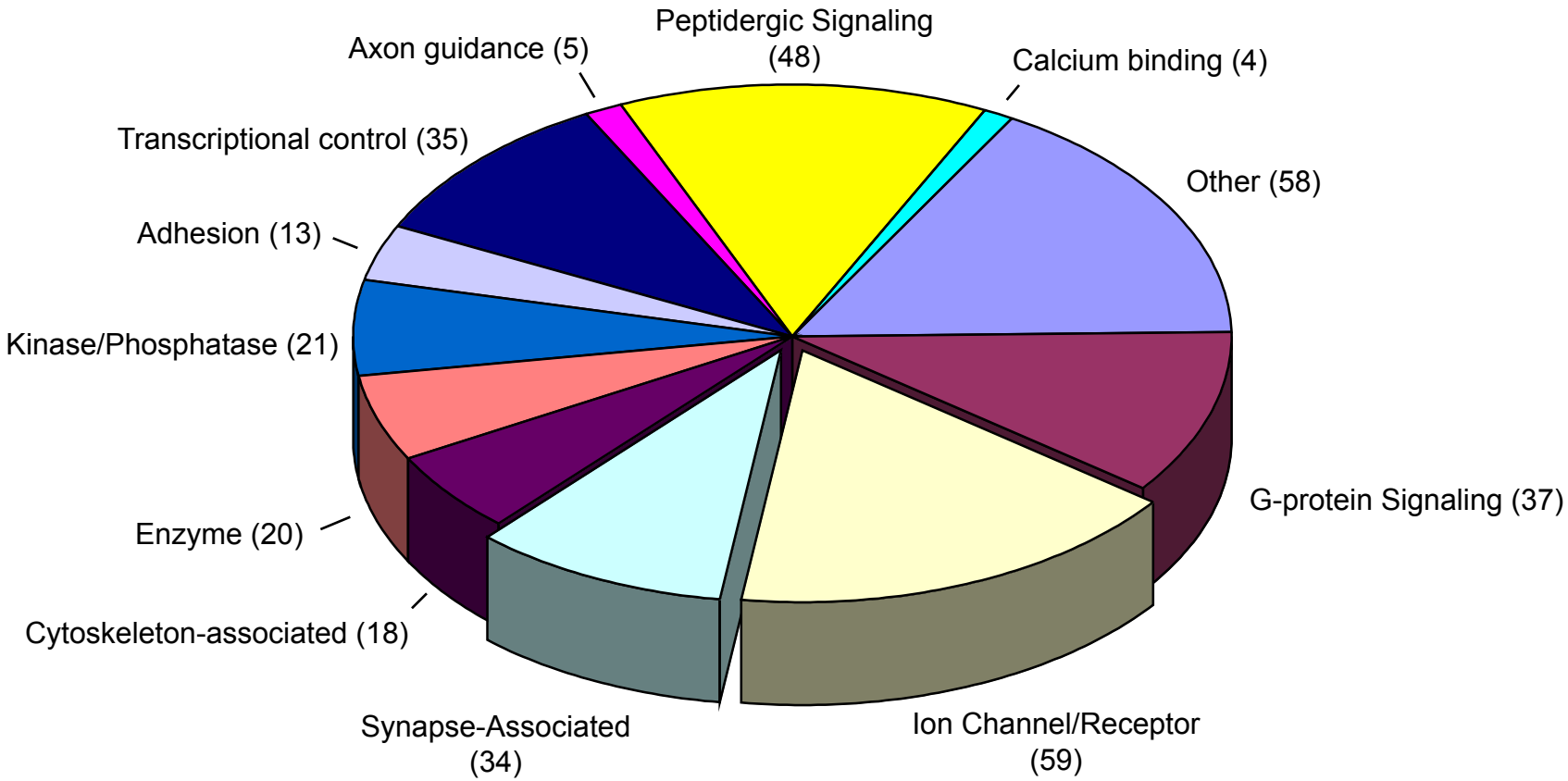

Figure 4.2 Piechart of neural-expressed genes in pan-neural enriched dataset.
In our dataset, 352 genes have been previously characterized as expressed in neurons. These were grouped according to common functions to reveal significant enrichment of functional categories known to be highly represented in neurons.. For example, there are 59 genes which are ion channel/receptors, with roles in signaling and regulating membrane potential. In addition, 34 genes are found at the synapse or belong to the synaptic vesicle pathway (see text).

Generating a gene expression profile of the nervous system, part II

In this case, ~1600 genes were significantly enriched in our LS pan-neural IP vs. LS control IP dataset, as opposed to 650 enriched genes using the earlier IP method that did not exclude high background mRNA. Furthermore, we retain the same percentage of genes known to be expressed in neurons (~90%, 352/393). Decreasing the amount of background RNA effectively deepened our dataset by ~1000 genes. The enhanced sensitivity of this analysis is also revealed by the detection of transcripts expressed in a small subset of neurons (e.g. *daf-7* is expressed in the 2 ASI neurons) (Table 4.1).

To assess the dataset, I grouped the list of genes previously identified as expressed in neuron into categories (Figure 4.2). As expected, genes involved in basic neural function (i.e. ion channel/receptors and synapse-associated components) are highly represented. In addition, the enriched transcripts reflect the diversity of neuronal types. For example, glutamate receptors (*glr-1 glr-2, glr-4, glr-5, nmr-1, nmr-2)*, acetylcholine receptors (*deg-3, des-2*), a GABA receptor (*unc-49)*, the dopamine transporter (*dat-1*), and the catecholamine transporter (*cat-1*) are all enriched in our dataset.

A striking example of the power of this profiling approach is revealed by the presence of a large number of genes involved in peptidergic signaling. The best studied invertebrate neuropeptides are the FMRFamide and related peptides (FaRPs). A subset of the 23 *C. elegans* "*flp*" (FMRFamide like peptides) genes, have known roles in cell excitability (*flp-13*), locomotion (*flp-1*), and feeding behavior (*flp-21*) (Nelson et al. 1998; Rogers et al. 2001). The majority (19/23) of the known *C. elegans flp* genes are enriched in the pan-neural dataset. *C. elegans* also contains 38 insulin-like peptides (*ins*) and 34 neuropeptide-like genes (*nlp)*. These are also highly represented in our dataset; 10

Table 4.1. List of pan-neural enriched genes that are expressed in few cells

Gene name	Fold Change	Description	Expressed in
daf-7	1.8	TGF-Beta	ASI
sro-1	2.2	7-TM receptor	ADL, SIA
ceh-10	1.8	transcription factor	AIY, CEP, RID, ALA, RME, AIN, AVJ, CAN
dat-1	6.7	dopamine transporter	eight dopaminergic neurons
del-1	2.6	DEG/ENaC subunit	11 VB, 2 SAB
unc-4	2.7	transcription factor	9 DA, 12 VA, 3 SAB, 1 I5
unc-25	4.8	glutamate decarboxylase	all 26 GABAergic neurons

ins genes and 18 *nlp* genes are enriched. Thus, the abundance of neuropeptide encoding transcripts in our dataset points to many unknown roles for peptidergic signaling in the function of the *C. elegans* nervous system (see Discussion).

In addition to ~1600 enriched genes, we also detected ~7200 'Expressed Genes' (EGs; see Materials and Methods). EGs are transcripts which are detected at some level in the pan-neural dataset and include the ~1600 enriched genes. This list of EGs contains a large number of genes involved in G-protein signaling. These data are strongly concordant with the embryonic A-class motor neuron profile generated by my colleague Rebecca Fox (see below) (Fox et al. 2005). For example, our two datasets contain most of the known components of the G-protein signal transduction pathway that functions in motor neurons to modulate acetylcholine release (Figure 4.3).

GFP reporters confirm neuronal expression of enriched transcripts in the pan-neural dataset

To further confirm our dataset, we generated or obtained 19 GFP reporter lines for uncharacterized genes. As seen in Table 4.2 and Figure 4.4. ~95% (18) of transgenics show expression in neurons. Some genes were specific to neurons. For example, *tsp-7*::GFP is expressed in all ventral cord motor neurons and many head and tail neurons; *F29G6.2*::GFP encodes a novel pan-neural gene. Other genes are expressed in neurons as well as other tissues. For instance, the TGF-β molecule, *tig-2*, is expressed in cholinergic ventral cord motor neurons as well as body wall and pharyngeal muscle. These randomly selected genes are ranked statistically high and low, further demonstrating the richness of the dataset.

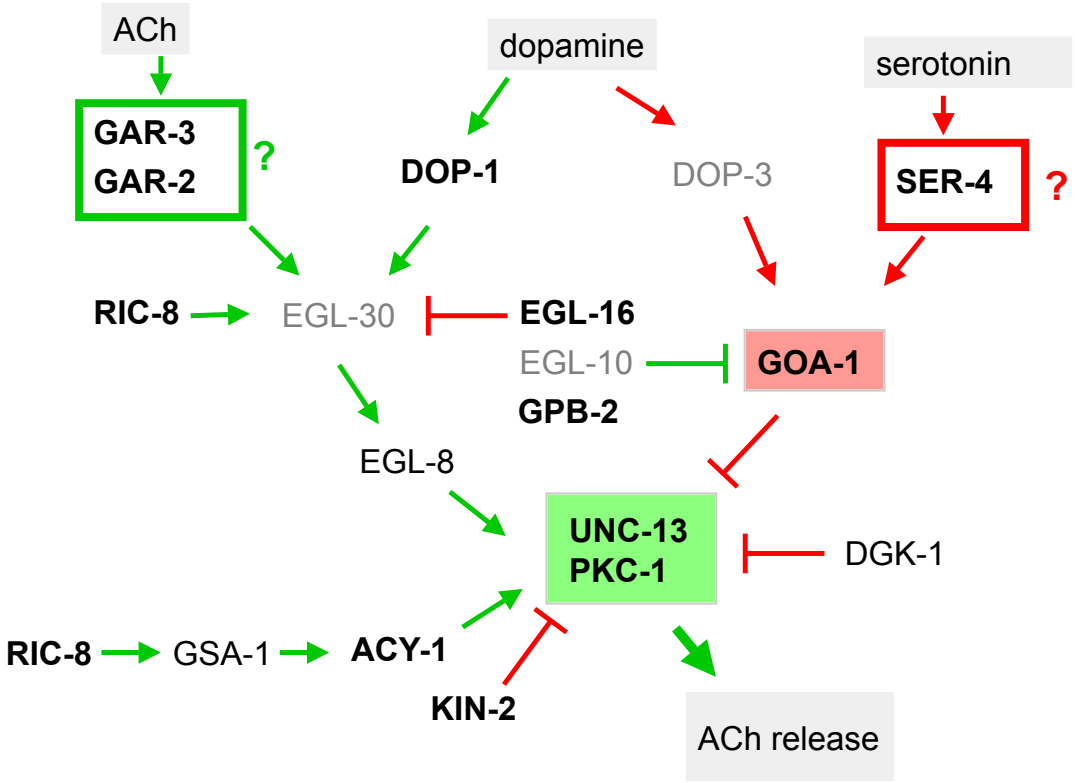

Figure 4.3. G-protein signaling pathways regulating neurotransmitter release in cholinergic motor neurons. Components shown in bold are enriched in the pan-neural dataset. All others are EGs with the exception of *dop-3*, *egl-30*, and *egl-10*, which are not detected (light gray). Green denotes interactions that promote acetylcholine (ACh) release and red marks steps that inhibit synaptic vesicle fusion. Hypothesized neurotransmitter receptors mark with question marks.Neurotransmitters are highlighted in gray boxes. This figure adapted from Fox, et al, 2005.

Generating a profile of VA motor neurons

Emboldened by our success with profiling the nervous system, the next step towards our goal of using mRNA-tagging to identify UNC-4 target genes was to generate a transgenic animal expressing epitope-tagged PAB-1 in VA motor neurons. We decided to increase the sensitivity of this approach by adding a tandem repeat of three FLAG epitopes. Of the three *unc-4*::3XFLAG::PAB-1 transgenic lines we obtained, anti-FLAG staining revealed that one of these showed highly specific expression in *unc-4* expressing neurons at the mid-L2 stage: strong staining was observed in VA (12) and SAB (3) motor neurons and AVF (2) interneurons, while weaker expression was seen in DA (9) motor neurons (Figure 4.5). The two other lines also showed ectopic expression in other cells and were discarded.

We used the A-class specific strain (NC694) to profile wildtype VA motor neurons at the mid-L2 stage. A comparison to IP control RNA revealed 415 enriched genes (1.5 fold, FDR $\leq$ 1%). Of the 100 genes in this list with known expression patterns (www.wormbase.org), 89 are found in neurons; of these, ~80% are detected in regions that also contain *unc-4* expressing neurons (retrovesicular ganglia, ventral nerve cord, pre-anal ganglia, I5 pharyngeal neuron). In addition, most of these genes encode proteins in functional categories (i.e. ion channels/receptors) that are also highly represented in the pan-neural dataset (Figure 4.6).

Additionally, we wished to verify A-class expression by making or obtaining GFP reporters for 17 randomly selected genes in our dataset. Of these, ~95% (16) were expressed in A-class neurons (Table 4.3, Figure 4.7). However, unlike *unc-4*, no GFP reporter was specific to A-class neurons. For example, the phosphatase gene *F09C3.2* is

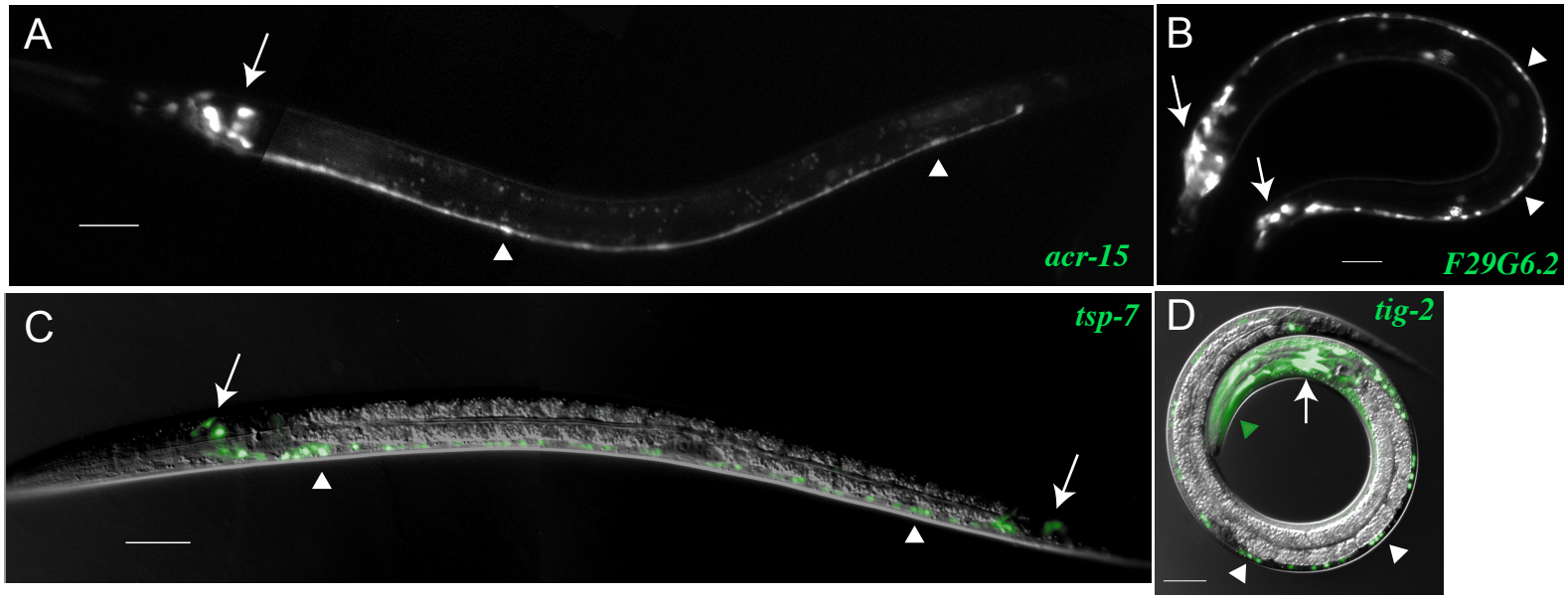

Figure 4.4 GFP reporters validate pan-neural microarray data.
Transgenic L2 animals expressing GFP reporters from representative pan-neural genes.
A. *acr-15*::GFP is expressed in head neurons (arrow) and a subset of ventral nerve cord (VNC) neurons (arrowheads).
B. *F29G6.2*::GFP is expressed in all neurons, including those in the head and tail (arrow) and VNC (arrowheads)
C. A DIC/GFP image of *tsp-7*::GFP expression in head and tail neurons (arrows) and all ventral cord motor neurons (arrowheads).
D. A DIC/GFP image showing *tig-2*::GFP expression in body wall muscle (green arrowhead), pharyngeal muscle (arrow), and cholinergic ventral cord motor neurons (white arrowheads).

Scale bars = 20 µm

Table 4.2 GFP reporters validate pan-neural dataset					
				Expression of GFP Reporters	
Rank	**Gene**	**Fold Enrichment**	**Description**	**Neurons**	**Other cells**
94	*trp-1*	2.2	Ca++ channel	ventral cord motor neurons (VNC)	
113	*acr-15*	4.9	acetylcholine receptor	VNC, head/tail neurons	
159	*syg-1*	1.5	Ig domain	VNC, other neurons	
175	*sto-4*	3.0	stomatin	VNC, head/tail neurons	
305	*nca-1*	2.3	Ca++ channel	VNC, head/tail neurons	
345	*F39B2.8*	3.5	membrane protein	VNC, head neurons	pharyngeal muscle
356	*rpy-1*	2.7	rapsyn	VNC	body wall muscle
548	*tig-2*	1.8	TGF-β	VNC, head/tail neurons, touch neurons	body wall muscle, pharyngeal muscle
600	*F29G6.2*	3.2	novel	all neurons	
636	*C04E12.7*	3.2	phopholipid scramblase	all VNC, head/tail neurons, others?	Weak in spermatheca, anal muscle, vulval muscle
725	*mec-12*	5.9	alpha-tubulin	VNC, touch neurons, head neurons	
739	*tsp-7*	3.5	tetraspanin	all VNC, head/tail neurons, touch neurons	
878	*mab-9*	1.7	T-box transcription factor	VNC, tail neurons	
959	*twk-30*	2.1	K+ channel	all VNC	
995	*mig-13*	1.8	CUB domain	anterior VNC	pharyngeal/intestinal valve, hypodermis
1048	*F55C12.4*	3.5	novel	VNC	
1061	*F09C3.2*	2.7	dehalogenase-like	VNC	intestine, hypodermis
1100	*T19C4.5*	2.0	novel		No GFP
1524	*acr-14*	1.5	acetylcholine receptor	VNC, head neurons	body wall muscle, intestine

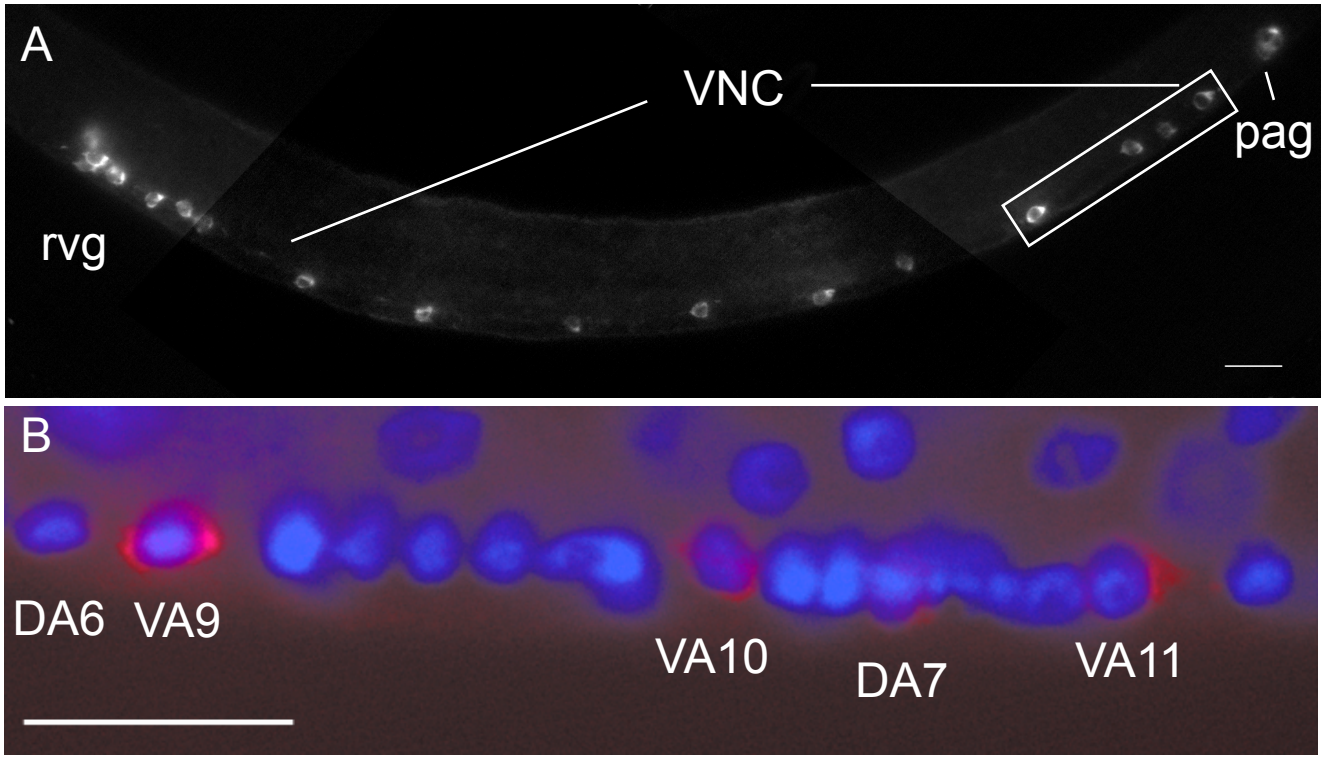

Figure 4.5 ***unc-4*::3XFLAG::PAB-1 expression in L2 larvae.**
A. Antibody staining detects FLAG::PAB-1 expressing A-class neurons (white circles) in the retrovesicular ganglian (rvg), ventral nerve cord (VNC), and pre-anal ganglian (pag).

B. Close-up of ventral cord (boxed image in A), showing anti-FLAG staining (red) in cytoplasm surrounding only A-class nuclei (DAPI, blue). Note dimmer expression in DAs vs VAs.

Anterior is left, ventral is down. Scale bars = 10 μm

expressed in all ventral cord motor neurons classes except DD. A GFP reporter for the a-tubulin *mec-12* displays fluorescence in DA, VA and VB motor neurons. The calcium channel *nca-1* is expressed in all A-class and B-class motor neurons in the ventral nerve cord. These results support the idea that a majority of the enriched transcripts in our dataset are also expressed in *unc-4* neurons.

It is especially reassuring to note that *unc-4* is the most highly enriched (~9-fold) transcript in this dataset. The O/E transcription factor *unc-3* is highly expressed in ventral cord motor neurons (i.e. DA, VA, DB, VB) and is enriched in the VA dataset (Prasad et al. 1998). I also identified mRNAs for other transcription factors with unknown roles in these motor neurons. For example, *ceh-6* (POU homeobox) and *ceh-44* (CUX class of CUT homeodomain), are enriched. This finding is interesting because overlapping homeodomain factors function in the vertebrate spinal cord to consolidate cell fate. Furthermore, POU and CUT proteins are well-established determinants of neural fate (Finney and Ruvkun 1990; Blochlinger et al. 1991). The bZip transcription factor *xbp-1* functions in the Unfolded Protein Response pathway and is important to traffic glutamate receptors from the endoplasmic reticulum; the apparent enrichment of *xbp-1* in neurons during development argues for an important role in neuronal differentiation (J Shim, T. Umemura, E. Nothstein, C. Rongo, pers. comm.). It would be interesting to determine if any of these transcription factors work with *unc-4* to specify A-type motor neurons.

Comparison of transcripts enriched in embryonic vs. larval *unc-4* neurons.

My colleague Rebecca Fox led our study profiling the embryonic DA motor neurons (Fox et al. 2005). For this work, Rebecca exploited the primary embryonic cell

culture method that our lab helped to optimize (Christensen et al. 2002). She isolated *unc-4*::GFP neurons (mainly DA motor neurons) from these cultures via Fluorescence Activated Cell Sorting (FACS). We call this technique Micro Array Profiling *C. elegans* celLs (MAPCeL). A comparison to the average nematode embryonic cell revealed ~1000 enriched genes. I worked with Rebecca to analyze these data, to research gene families enriched in these neurons, and to generate GFP reporters to validate the gene list (Fox et al. 2005). Now I have compared this embryonic DA dataset obtained by the MAPCeL approach to the VA dataset generated in this work by the mRNA-tagging method. This approach identified candidate genes for shared A-class traits as well as genes that define traits unique to each of the A-class motor neuron subtypes.

The DAs are similar to the post-embryonic VAs in that they express *unc-4*, are cholinergic, send anterior directly axons to connect to muscle, and receive inputs from the command interneurons AVA, AVD, and AVE (White et al. 1986). On the other hand, DA motor neurons also express characteristics that distinguish them from VA motor neurons including embryonic birth, dorsal inputs to muscle, and normal wiring in *unc-4* mutants (White et al. 1986; White et al. 1992).

A comparison of embryonic vs. larval A-class profiles revealed 161 shared genes. (Figure 4.8) Many of these common genes are known to be expressed in both cell types, such as those involved in synaptic vesicle signaling (*sng-1, snn-1, snt-1,* etc) or transcriptional control (*unc-3, unc-4*). Other common transcripts encode neuropeptides, acetylcholine receptors, and G-protein coupled receptors (Table 4.4). This finding further supports our speculation that *C. elegans* cholinergic motor neurons are acutely responsive to external signals as well as active in sending signals to other cells (Fox et al. 2005).

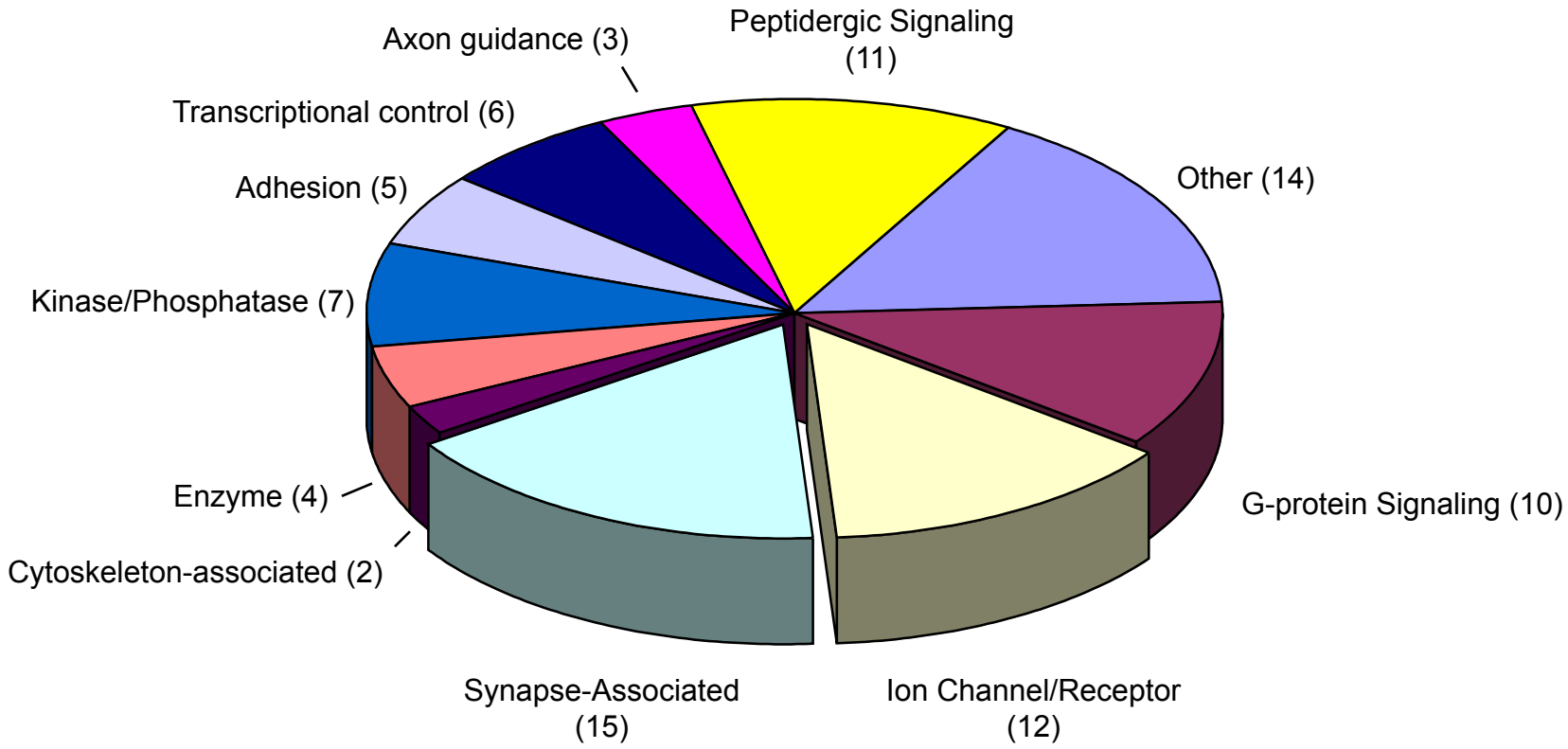

Figure 4.6 Piechart of neural-expressed genes in wildtype A-class enriched dataset.
89 genes in our dataset have been previously characterized as expressed in neurons. These were grouped according to common functions to reveal significant enrichment of functional categories known to be highly represented in neurons. For example, there are 12 genes which encode ion channel/receptors and 15 genes are found at the synapse or belong to the synaptic vesicle pathway (see text).

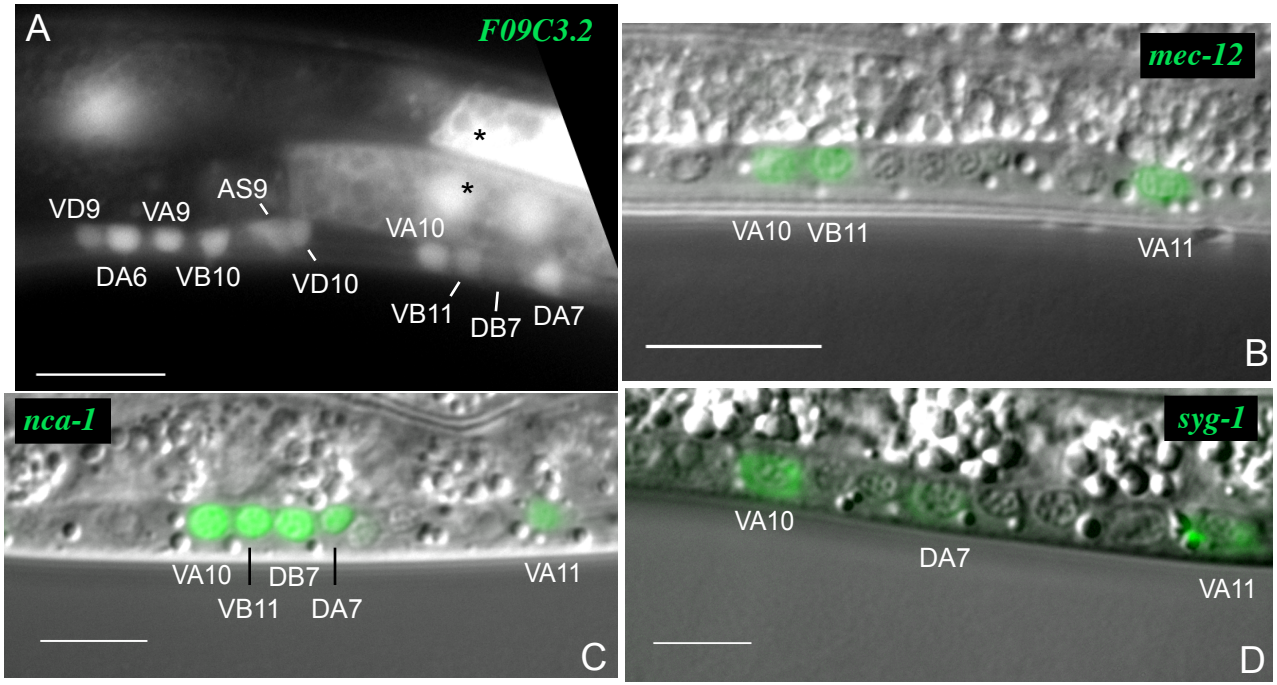

Figure 4.7 GFP reporters validate A-class microarray data.
Transgenic L2 animals expressing GFP reporters for representative A-class motor neuron genes. Anterior to the left.
A. *F09C3.2*::GFP expression in posterior ventral nerve cord (VNC) motor neurons. Astericks denote expression in the intestine.
B. A DIC/GFP image of *mec-12*::GFP expression in posterior VA and VB motor neurons.
C. A DIC/GFP image of *nca-1*::GFP expression in posterior A- and B-class motor neurons.
D. A DIC/GFP image of *syg-1*::YFP expression in posterior A-class motor neurons.
Scale bars: 10 µm in A-B, 5 µm in C-D.

Table 4.3 GFP reporters validate A-class enriched dataset					
				Expression of GFP Reporters	
Rank	Cosmid	Gene	Fold Enrichment	UNC-4 neurons	Other cells
27	ZC21.2	*trp-1*	1.9	DA, VA	DB, VB
37	E03D2.2	*nlp-9*[1]	3.5	VA	head neurons, intestine
40	F33D4.3	*flp-13*[2]	7.9	I5	ASE, ASG, ASK, BAG, DD, M3, M5, head neurons
52	F36A2.4	*twk-30*[3]	5.1	DA, VA	All ventral cord motor neurons
48	C04E12.7		1.7	DA, VA	All ventral cord motor neurons, head/tail neurons
71	C11D2.6	*nca-1*	2.2	DA, VA	DB, VB, head/tail neurons
75	F43C9.4a	*mig-13*	2.8	DA	DB, anterior VNC neurons, pharyngeal/intestinal valves, hypodermis
88	F29G6.2		1.6	DA, VA	~all neurons
98	Y47D3B.2A	*nlp-21*[4]	3.7	DA, VA	DB, VB, AS, body muscle, head neurons, intestine
117	F39B2.8		2.1		Pharyngeal muscle
126	K02E10.8	*syg-1*	1.8	DA, VA	HSN, other neurons
145	T05C12.2	*acr-14*	1.7	DA	DB, VB, AS, DD, HSN, VC4&5, AIY, head neurons, muscle intestine
239	Y71H9A.3	*sto-4*	1.6	VA	VB, AS, head/tail neurons
248	C44B11.3	*mec-12*	1.9	DA, VA	VB, touch neurons, head/tail neurons
257	T23D8.2	*tsp-7*	4.8	DA,VA	all ventral cord motor neurons, head/tail neurons, touch neurons
392	F55C12.4		2.1	DA	VB, DB, DD, VD, AS, head/tail neurons
408	F09C3.2		1.7	DA, VA	DB, VB, VD, intestine, hypodermis

1. Li, et al., 1999
2. Kim and Li, 2004
3. Salkoff et al., 2001
4. Nathoo, et al., 2001

Transcripts that are exclusively enriched in either embryonic or larval *unc-4* neurons may accounts for morphological or function differences between DA and VA motor neurons. For example, the DA motor neurons express the Netrin receptors *unc-5, unc-40,* and the Receptor Protein Tyrosine Phosphatase (RPTP) *clr-1*, function together to steer DA axons dorsally (Hedgecock et al. 1990; Chang et al. 2004). In contrast, VA motor neurons do not express these receptors and extend ventrally projecting axons. The T-box transcription factor, *mab-9*, is known to be expressed in DA but not VA motor neurons, further supporting these data (Figure 4.9). In another example, the DEG/ENaC subunit, *del-1*, is selectively enriched in the post-embryonic A-class dataset as expected from its known expression in SAB neurons after hatching (Winnier et al. 1999). These results suggest that comparisons of these datasets to other classes of motor neurons should also reveal genes with unique roles in specifying individual motor neurons types.

The dataset isolated from wildtype VA neurons is robust and contains many genes known to be present in these cells. Thus, I conclude that mRNA-tagging is sufficiently sensitive to profile a subset of neurons. I next turned my efforts to profiling VA neurons from *unc-4* and *unc-37* mutants to reach the long-sought goal of identifying UNC-4 regulated genes.

Profiling *unc-4* and *unc-37* mutant VA motor neurons

As stated in Chapter III, genes upregulated in both *unc-4* and *unc-37* mutant backgrounds will be considered strong candidate UNC-4 target genes. To identify these genes, the *unc-4*::3XFLAG::PAB-1 (3FPAB) transgene was crossed into *unc-4 (e120)* and *unc-37 (e262)* mutants. Unfortunately, for reasons that are not understood, the

combination of the 3FPAB transgene with *unc-4* mutants produced subviable animals. I crossed the 3FPAB transgene into multiple *unc-4* mutant backgrounds as well as created many new transgenic lines and crossed those into *unc-4 (e120)*, all to no avail; no combination of transgene and *unc-4* mutant was healthy. Luckily, the 3FPAB transgene was successfully crossed into the *unc-37 (e262)* mutant (Figure 4.10). Thus, using this line, I was able to generate a list of genes differentially expressed in wildtype and *unc-37* VA motor neurons.

Since UNC-4 and UNC-37 act as transcriptional repressors, genes which were significantly upregulated in *unc-37* vs. wildtype datasets were examined first. This analysis yielded ~400 genes (1.7 fold, FDR $\leq$ 5%). These results verified that mRNA-tagging is a more sensitive approach than the whole-animal experiments described in the previous chapter; two known *unc-4* regulated genes: *del-1* and *glr-4* (see Chapter II) are upregulated in this dataset (*acr-5* was not detected but this result is likely due to the insensitivity of the Affymetrix chip, as *acr-5* is also not detected in other microarray experiments in our lab). These results suggest that the *unc-37* dataset should also contain other bona fide *unc-4* targets.

I was initially surprised to see that many known A-class genes were detected as enriched in the *unc-37* dataset, including *unc-4, unc-17, cha-1, cho-1, snb-1,* and *snt-1*. To validate the dataset, I made GFP reporters for several genes (e.g. *sto-4, mab-9*) from this list using constructs generated by the Promoterome project (Dupuy et al. 2004). All of these reporter genes are expressed in DA motor neurons (Table 4.5). I reexamined the *unc-37*; 3FPAB strain by anti-FLAG staining and noticed that more DA motor neurons

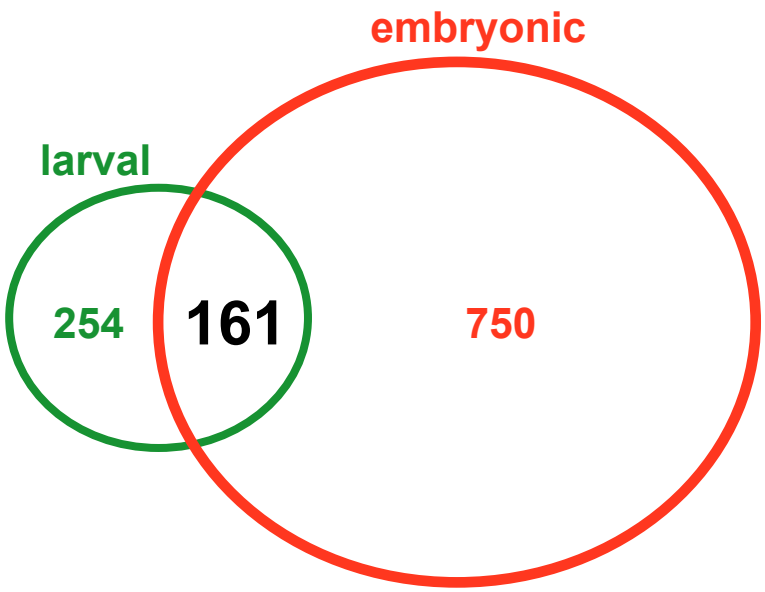

Figure 4.8 Comparison of enriched genes in VA motor neurons vs DA motor neurons.
254 genes are exclusively enriched in larval VA motor neurons while 750 genes are enriched exclusively in embryonic DA motor neurons. These similar motor neurons share 161 enriched transcripts.

Table 4.4 Examples of genes enriched in both DA and VA motor neuron profiles

Cosmid Name	Common Name	KOG (other description)
Acetylcholine Receptors		
R01E6.4	*acr-12*	Acetylcholine receptor
T05C12.2	*acr-14*	Acetylcholine receptor
K11G12.2	*acr-2*	Acetylcholine receptor
F21F3.5	*unc-38*	Acetylcholine receptor
G-protein signaling		
F08B6.2	*gpc-2*	G protein gamma subunit
C05B5.7	*rgs-1*	G protein signaling regulators
C41G11.3	*rgs-6*	G protein signaling regulators
ZC84.4		7 transmembrane receptor
F59D12.1		7 transmembrane receptor
C47E8.3		7-transmembrane receptor
F47D12.1a	*gar-2*	(Muscarinic acetylcholine receptor)
Neuropeptides		
F33D4.3	*flp-13*	(FMRFamide-like peptide)
W07E11.3a	*flp-2*	(FMRFamide-like peptide)
C18D1.3	*flp-4*	(FMRFamide-like peptide)
C03G5.7	*flp-5*	(FMRFamide-like peptide)
Y47D3B.2a	*nlp-21*	(Neuropeptide-like)
E03D2.2a	*nlp-9*	(Neuropeptide-like)
F13B12.5	*ins-1*	(Insulin-like)

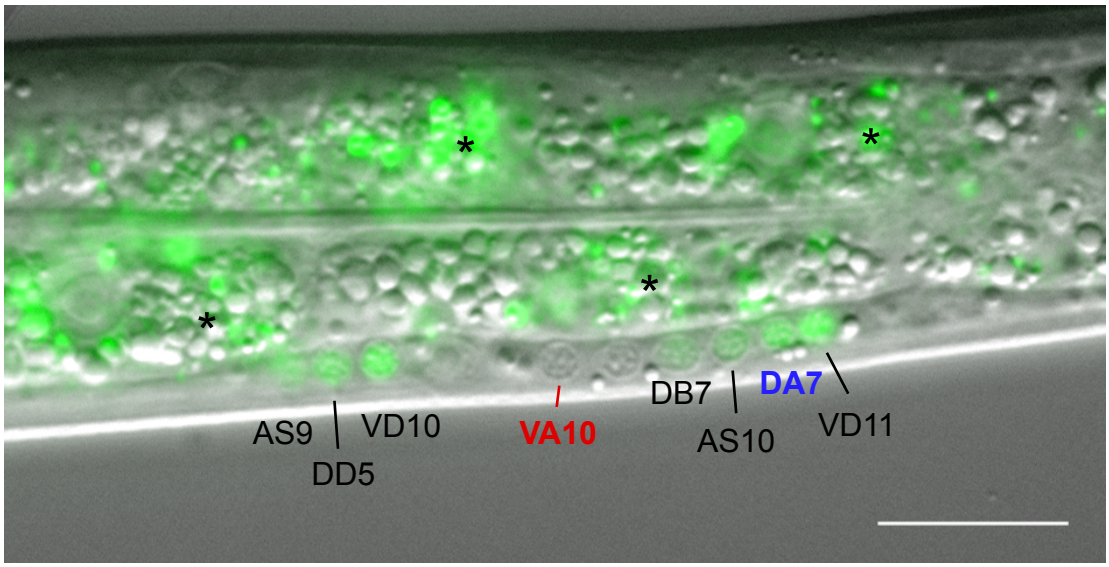

Figure 4.9 *mab-9*::GFP expression in L2 larvae.
mab-9::GFP is expressed in DA motor neurons (blue) but not VA motor neurons (red). It is also expressed in AS, DB, DD, and VD motor neurons. This transgenic was a kind gift from Roger Pocock and Allison Wollard (Oxford).
Astericks mark gut autofluorescence. Anterior is to the left, dorsal is up.
Scale bar = 10 μm.

expressed the 3FPAB transgene in mid-L2 than in the wildtype (Compare Fig 4.10 to Fig 4.5). Thus, genes which are expressed in DAs and VAs would be 'upregulated' in the *unc-37* vs wildtype analysis. I postulated that this observation explained the presence of many known A-class genes in the *unc-37* upregulated target list (and suggests that *unc-4* expression in DAs may be repressed by *unc-37* after hatching). To reduce these background "DA" genes, I eliminated any genes from my *unc-37* enriched list that were also enriched in Rebecca's embryonic DA motor neuron dataset. This adjustment decreased the number of *unc-37* regulated transcripts to ~280.

It is likely that only a subset of the *unc-37* dataset will represent *unc-4* regulated genes because Groucho co-repressor proteins are known to interact with many classes of transcription factors (Chen and Courey 2000). For example, the Even-skipped homolog *vab-7* is present in our *unc-37* target list. Mr. Watson and I performed anti-VAB-7 staining on *unc-4* and *unc-37* mutants to confirm this result. This experiment showed that VAB-7 is ectopically expressed in VA and DA motor neurons in *unc-37 (e262)*, but **not** in *unc-4 (e120)*, mutants (Figure 4.11, data not shown). Therefore, in an effort to identify the subset of *unc-37* regulated genes that are also controlled by *unc-4*, I compared my list of *unc-37* upregulated list to Rebecca's *unc-4* and *unc-37* upregulated dataset (obtained by comparing wildtype, *unc-4*, and *unc-37* FACS-isolated *unc-4*::GFP embryonic neurons). Surprisingly, only one gene, *ceh-12,* is upregulated in the list of *unc-4*/*unc-37* regulated genes detected in both the embryonic and larval datasets. This finding is very intriguing, however, because *ceh-12* is the apparent nematode homolog of HB9, a homeodomain transcription factor with a well-established role in fly and vertebrate motor neuron specification (Arber et al. 1999; Broihier and Skeath 2002). I have therefore

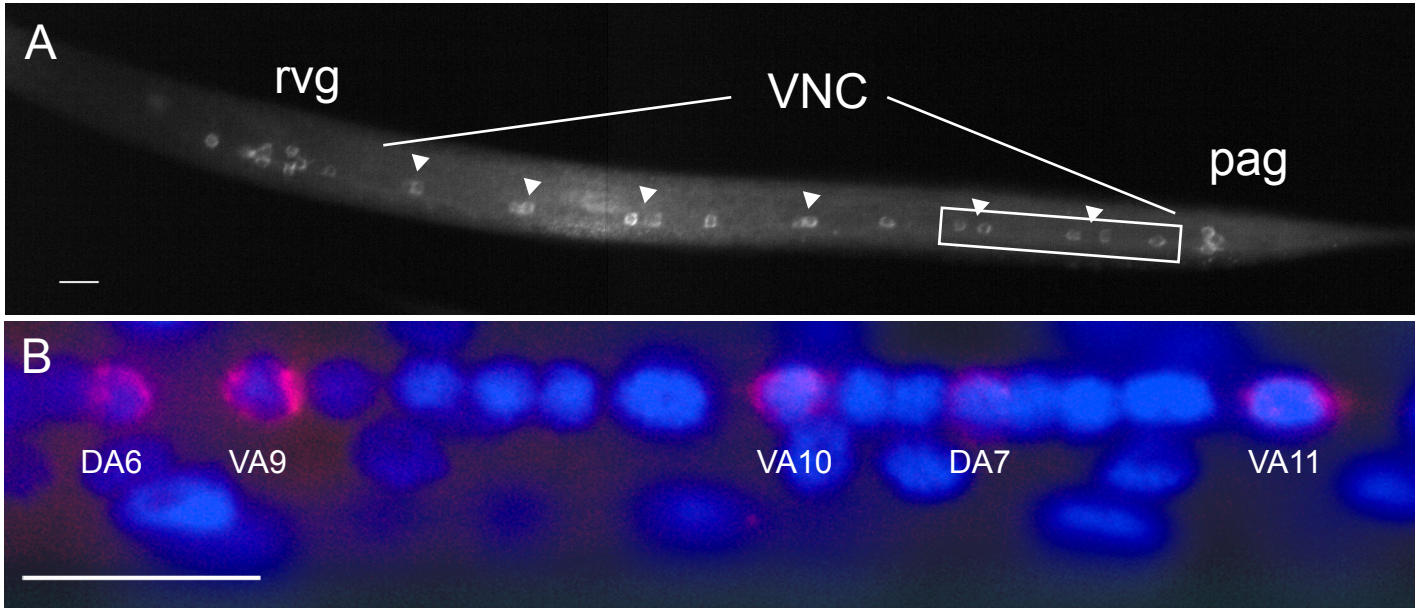

Figure 4.10 *unc-4*::3XFLAG::PAB-1 expression in *unc-37 (e262)* L2 larvae.
A. Antibody staining detects FLAG::PAB-1 expressing A-class neurons (white circles) in the retrovesicular ganglian (rvg), ventral nerve cord (VNC), and pre-anal ganglian (pag). White arrowheads point to VA and DA 'doublets.'

B. Close-up of ventral cord (boxed image in A), showing anti-FLAG staining (red) in cytoplasm surrounding only A-class nuclei (DAPI, blue). Note expression levels in DAs match that in VAs.

Anterior is to the left.. Scale bars = 10 µm

Table 4.5 List of GFP reporters from *unc-37* regulated dataset that are expressed in DA motor neurons.

Gene	Description	Fold change	Expressed in
rpy-1	Rapsyn	1.8	**DA**, VD, AS, VB, DB, body muscles
sto-4	Stomatin	1.9	**VA, DA**, VB, DB
mab-9	T-box transcription factor	2	**DA**, DB, DD, VD, AS
tig-2	TGF-Beta	2	**VA, DA**, VB, DB, body muscles, pharyngeal muscle, touch neurons
F29G6.2	Unknown	2.1	all neurons
tsp-7	Tetraspanin	2.5	all VNC motor neurons, head/tail neurons, touch neurons
F09C3.2	Unknown	2.5	**VA, DA**, VB, DB, VD, intestine, hypodermis
nlp-21	Neuropeptide	4	**VA, DA**, VB, DB, AS, body muscle, head neurons, intestine

focused on analyzing *ceh-12* to determine if it functions in the *unc-4* synaptic specificity pathway (see Chapter V).

Discussion

mRNA-tagging is a powerful strategy to profile post-embryonic neurons

A focus of this chapter was to describe the validation of a new method for isolating cell-specific mRNA from post-embryonic neurons. With the help of Joseph Watson, I optimized mRNA-tagging to generate a robust profile of the entire nervous system at the mid-L2 stage. The specificity and sensitivity of this approach were substantially enhanced by our discovery that low-salt washes reduced the background of non-specific RNAs. This outcome was evident in the increased numbers of neural enriched genes and concomitant lower False Discovery Rate (FDR). This modification and our data should convince the *C. elegans* community that this technique works well to profile post-embryonic tissues.

The pan-neural dataset was validated by database searches and GFP reporters. For example, of the 392 transcripts in our dataset with previously determined expression patterns, 352 (~90%) genes are expressed in neurons, including the pan-neural genes *snb-1* (Synaptobrevin) and *snt-1* (Synaptotagmin). A majority (~75%) of the genes, however, have not been studied. Since 54% (848) of these pan-neural enriched transcripts have a human homolog (defined by E-score $\leq x^{-10}$), this dataset is a rich resource for novel neural genes. For example, 14 unstudied enriched genes encode transcription factors. One of these genes, *T01D3.2*, is similar to *NPAS3*, a member of the bHLH-PAS (basic helix-

loop-helix, Period, Aryl hydrocarbon receptor, Simpleminded) family. Murine *NPAS3* is expressed in the developing neural tube and exclusively in the adult brain (Brunskill et al. 1999). Knocking out *NPAS3* and its closely related family member *NPAS1* results in mice with behavioral deficits, such as impaired social recognition (Erbel-Sieler et al. 2004). Further, two studies link a translocation in human *NPAS3* with schizophrenia (Kamnasaran et al. 2003; Pickard et al. 2005). Thus, it is likely that this pan-neural enriched gene exerts important roles in the *C. elegans* nervous system, and that detailed studies of its function in this model system could be useful for elucidating the molecular basis of human disease.

In addition to profiling the entire nervous system, we sought to generate a gene-expression fingerprint for a subset of ventral cord motor neurons. Because of our interest in identifying UNC-4 target genes, we created a transgenic to profile the larval VA motor neurons (Fig 4.5). This experiment revealed 415 larval A-class enriched genes. Similar to the pan-neural dataset, ~90% of genes in this dataset for which *in vivo* expression is known are in fact expressed in neurons. In addition, ~95% of GFP reporters to genes on this list confirm expression in larval A-class neurons (Table 4.3). Almost half (48%) of VA-enriched genes have human homologs, suggesting that several of the genes in this dataset could aid in the understanding of development and function of human motor circuits. Interestingly, when the VA dataset is compared to a profile of embryonic A-class motor neurons, only 161 genes are found in both cell profiles. A majority of these are genes involved in neuronal function, such as the synaptic vesicle cycle. It will be interesting to determine if any of the unstudied genes are exclusively expressed in A-

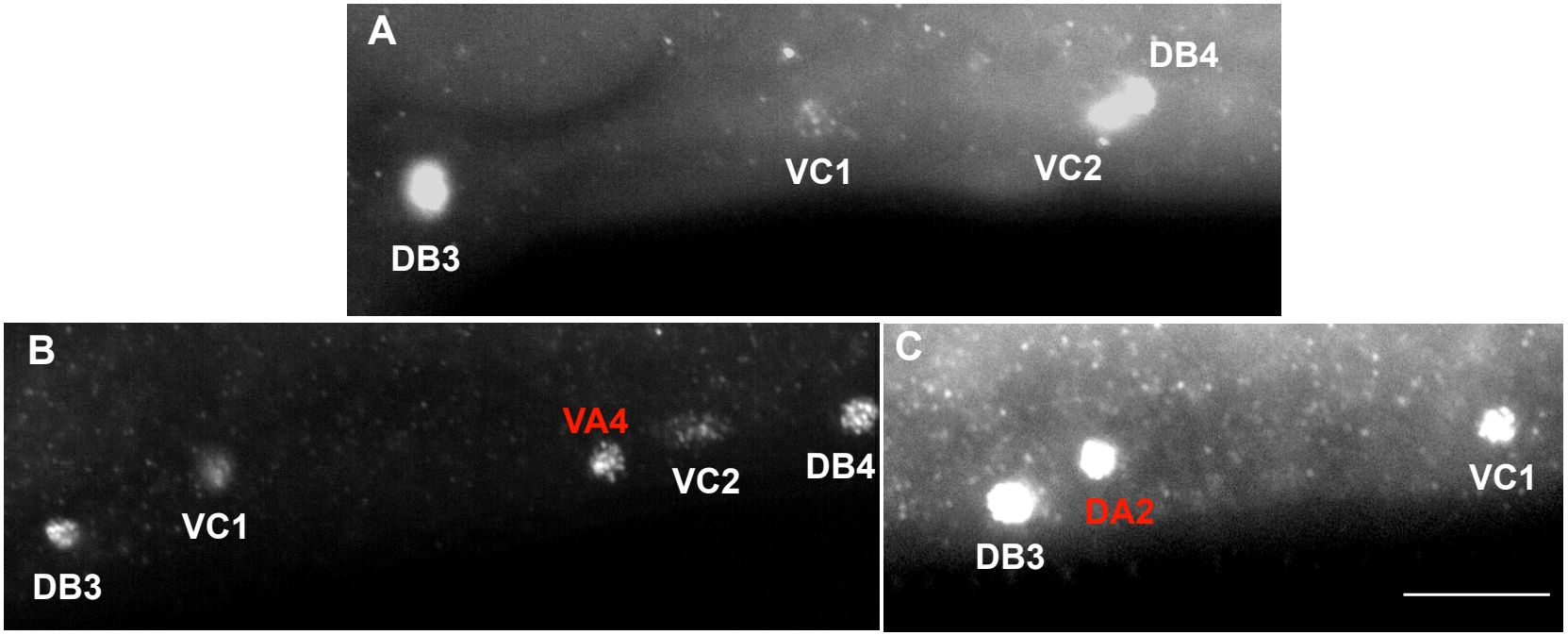

Figure 4.11 VAB-7 is ectopically expressed in A-class neurons in *unc-37 (e262)* mutants.
Anterior is to the left, scale bar = 10 µm.
A. Anti-VAB-7 staining reveals expression in DB and VC motor neurons in the anterior ventral nerve cord (VNC).
B, C. Anterior VA and DA motor neurons (red lettering) ectopically express VAB-7 in *unc-37 (e262)* mutants.

class neurons. To date, *unc-4* is the only known gene expressed solely in A-class motor neurons.

While VA and DA motor neurons are similar, they display distinctively different characteristics (e.g. axonal trajectory) that are indicative of differential gene expression. The microarray results substantiate this assumption: ~200 genes are exclusive to the VA dataset and 750 in the DA profile. Perhaps studying these genes would lend insight into the differentiation of these A motor neuron subtypes.

In summary, mRNA-tagging is a robust method to isolate cell-specific transcripts from post-embryonic tissues.

mRNA-tagging to identify transcription factor targets

Having established that mRNA-tagging is a reliable strategy for profiling neuron-specific gene expression, I set out to use this approach to identify *unc-4* regulated transcripts. With this goal in mind, I isolated transcripts from *unc-37* mutant VA motor neurons. (As noted above, we did not include mRNAs isolated from the *unc-4* background because of unexplained strongly negative interactions between the 3FPAB transgene and *unc-4* mutants.) This approach revealed ~280 upregulated genes in the *unc-37* dataset. Of these, only a subset is likely to be *unc-4* regulated (Figure 4.11). To identify these authentic UNC-4 regulated genes, I compared my *unc-37* regulated larval A-class target list to Rebecca's *unc-4*/*unc-37* regulated embryonic A-class target list. This comparison yielded one gene, the homeodomain transcription factor *ceh-12*/HB9. As discussed in Chapter V, we aggressively pursued this gene because of its known roles in

motor neuron differentiation in *Drosophila* and vertebrates (chick, mouse) (Arber et al. 1999; Thaler et al. 1999; Broihier and Skeath 2002).