TITLE

Members of the CUGBP Elav-Like Family of RNA-Binding Proteins are Expressed in Distinct Populations of Primary Sensory Neurons

AUTHORS

Eliza Grlickova-Duzevik[1,2]	egrlickovaduzevik@une.edu
Merilla Michael[1,2]	merilla.michael@gmail.com
Aidan McGrath-Conwell[2,3]	amcgrathconwell@une.edu
Peter Neufeld[2,3]	pneufeld@une.edu
Thomas M Reimonn[4]	Thomas.Reimonn@umassmed.edu
Derek C Molliver[1,2]	dmolliver@une.edu
Benjamin J Harrison[1,2]	bharrison2@une.edu

1. Biomedical Sciences, College of Osteopathic Medicine, University of New England, Biddeford. ME 04005
2. Centre for Excellence in the Neurosciences, University of New England. ME 04005
3. College of Arts and Sciences, University of New England, Biddeford. ME 04005
4. University of Massachusetts Medical School, Program in Bioinformatics and Integrative Biology, Worcester, MA

ACKNOWLEDGMENTS:

P20GM103643 MH116492 T32GM107000

ABSTRACT

Primary sensory Dorsal Root Ganglia (DRG) neurons are diverse, with distinct populations that respond to specific stimuli. Previously, we observed that functionally distinct populations of DRG neurons express mRNA transcript variants with different 3′ untranslated regions (3′UTR's). 3′UTRs harbor binding sites for interaction with RNA-binding proteins (RBPs) critical for targeting mRNAs to subcellular domains, modulating transcript stability and regulating the rate of translation. In the current study we sought to determine if 3′UTR-binding proteins are restricted to specific DRG neuron populations. Analysis of publicly available single-cell RNA-Sequencing (scRNA-Seq) data generated from adult mice revealed that 17 3′UTR-binding RBPs were enriched in specific populations of DRG neurons. This included 4 members of the CUGBP Elav-Like Family (CELF). CELF2 and CELF4 were enriched in peptidergic, CELF6 in both peptidergic and nonpeptidergic and CELF3 in tyrosine hydroxylase-expressing neurons. CELF4 is a known regulator of neural excitability, likely through modulation of protein synthesis via binding to interaction sites within the 3′UTRs of mRNAs. Immunofluorescence studies showed 60% of CELF4+ neurons are small diameter C fibers and 33% medium diameter myelinated (likely Aδ) fibers. Co-expression analyses using transcriptomic data and quantitative immunofluorescence revealed that CELF4 is enriched in nociceptive neurons that express GFRA3, CGRP and the capsaicin receptor TRPV1. Finally, genes with CELF4 binding motifs expressed in CELF4+ neurons are significantly associated with gene ontology (GO) terms such as "RNA-binding" and "translation". We propose that CELF4 may therefore control a novel regulon that coordinates the translation of mRNAs encoding components of the protein translation apparatus in nociceptors.

INTRODUCTION

RNA binding proteins (RBPs) are critical for orchestrating the post transcriptional fate of transcripts, including transport to sub-cellular domains for local translation, storage in / release from ribonucleoprotein complexes and recruitment of translational machinery. The 3′ untranslated region (3′UTR) of transcripts is a hub for mRNA-protein interactions. For example, RBP binding to 3′UTRs on nascent transcripts regulates poly(A) signal site selection, cleavage and polyadenylation. On mature transcripts, 3′UTRs contain "zip-code" motifs for subcellular targeting (Kislauskis et al., 1994; Kislauskis & Singer, 1992; Ross et al., 1997), anchoring points for scaffolding interactions (Berkovits & Mayr, 2015) and interaction sites for recruitment of miRNAs (Grimson et al., 2007; Majoros & Ohler, 2007) or inhibition by competition for binding sites (e.g., (Srikantan et al., 2012)).

3′UTR-protein interactions are highly dynamic and provide a regulatory layer for modulation of diverse cellular functions. By interaction with alternative cleavage sites at the distal end of nascent transcripts, RBP-3′UTR interactions generate alternate 3′UTR isoforms through a mechanism termed Alternative Polyadenylation (APA) (Harrison et al., 2019). mRNA isoforms with long 3′UTRs are predominantly expressed during development (Ji & Tian, 2009) and in the adult are restricted to certain cell types most notably neurons that express mRNAs with long 3′UTRs (Harrison et al., 2012). Differential expression of 3′UTR isoforms by APA is associated with diverse biological functions, for example long 3′UTR isoforms upregulated during neuroplasticity (An et al., 2008; Miura et al., 2013) and during responses to stress (Graber et al., 2013; Zheng et al., 2018), etc. APA is dysregulated during human disease. For example, shortened 3′UTR

sequences contain fewer interactions sites for micro-RNAs, facilitating oncogene activation and tumor growth (Mayr & Bartel, 2009).

RBP-3'UTR interactions are especially important for remote cellular compartments that rely on a supply of mRNA for local protein synthesis. Sensory neurons have extremely long axons projecting from peripheral targets through their cell bodies (soma) housed in dorsal root ganglia (DRG) terminating in the spinal cord dorsal horn. These termini require a supply of raw materials and machinery that is transported along neuronal processes. RBPs are required for mRNA transport to and translation in these remote domains for maintenance of synapses and for plasticity (Hornberg & Holt, 2013; Wagnon et al., 2012). Numerous "neuron-specific" RBPs have been identified that are highly enriched in neurons and are required for mRNA transport and neuronal excitability (Darnell, 2013; Hornberg & Holt, 2013).

Primary somatosensory neurons are classified according to their responsiveness to specific stimuli (light touch, temperature, stretch, etc.), neurochemistry and neurophysiology (Le Pichon & Chesler, 2014), and can be identified by population-specific histological markers. Sensory neurons with unmyelinated axons (C-fibers) tend to have small-diameter cell bodies, and the great majority of these neurons are considered to be nociceptors (reviewed in Light, 1992: A.R Light, The Initial Processing of Pain and Its Descending Control: Spinal and Trigeminal Systems, Karger, Basel), grossly divided into two subpopulations: peptidergic neurons, which express CGRP and Ntrk1 and are particularly important for the transduction of noxious heat stimuli, and non-peptidergic neurons, which bind the plant lectin IB4 and express the G protein-coupled receptor MrgprD and are particularly important for mechanical nociception (Dong et al.,

2001; Silverman & Kruger, 1988). Most sensory neurons that express the heat- and acid-gated channel TRPV1 in adulthood are peptidergic neurons (Zwick et al., 2002). Peptidergic neurons also include a subset of thinly myelinated (A-delta) nociceptors and a minority of A-beta afferents (Lawson et al., 1993, 1996). A third population of C-fibers is comprised of neurons that express tyrosine hydroxylase (TH) and are not nociceptors; instead they are sensitive mechanoreceptors that have been implicated in the transduction of pleasant "social touch" (Lallemend & Ernfors, 2012; Li et al., 2011). Although markers of the peptidergic/non-peptidergic nociceptor populations are highly segregated in the mouse, the selectivity of these markers varies considerably across species; in mouse, IB4-binding is seen in less than 10% of neurons expressing CGRP or TRPV1, whereas in the mouse this overlap is 35% (Price & Flores, 2007; Zwick et al., 2002), and recent evidence shows that in humans CGRP and TRPV1 are expressed in most nociceptors (IB4 does not bind sensory neurons in human) (Shiers et al., n.d.). Recent advances in single-cell RNA-sequencing (scRNA-Seq) allow sensory neurons to be characterized by comprehensive mRNA expression profiles, rather than a handful of neurochemical markers. These transcriptomics studies align closely with histological data and are a powerful resource for identifying the genomic signatures of functionally-distinct neuronal populations (Usoskin et al., 2015).

Previously, we reported that functionally distinct populations of DRG neurons express mRNA isoforms with divergent 3'UTR sequences (3'UTR isoforms), and that these 3'UTR variants contain different RBP interaction motifs (Harrison et al., 2019). This led us to theorize that sensory neuron diversity is controlled by differential RBP-3'UTR interactions. In the current study we examine the expression of RBPs in DRG, providing further evidence suggesting divergence of

3'UTR-protein interactions in distinct populations of sensory neurons.

METHODS

Animals

All animal procedures were approved by the Institutional Animal Care and Use Committee of the University of New England consistent with federal regulations and guidelines. All animals used for this study were 6 week old male C57/BL6 mice purchased from Jackson Labs (Bar Harbour, Maine).

Dorsal Root Ganglion Single Cell RNA-seq Analysis

Trimming, aligning, and quantification: Previously published single-cell RNA-seq data from adult mouse dorsal root ganglion neurons produced by Usoskin et al. was accessed from GEO using accession code GSE59739 (Usoskin et al., 2015). Fastq files were downloaded and reads were trimmed using Trimmomatic version 0.39 (Bolger et al., 2014) in single-end mode with the following parameters: ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Trimmed reads were aligned to the mouse genome (GRCm38, primary assembly) and transcripts mapping to genes were quantified (Gencode vM25, primary assembly) using STAR version 2.7.5a (Dobin et al., 2013) with default parameters. Transcript quantifications were summed by gene when one cell contained reads from multiple fastq files, and gene counts were tabulated into a gene-by-cell expression matrix.

Clustering, Visualization, and Annotation: Downstream processing was performed using Seurat

version 3.2.0 (Stuart et al., 2019). Cells that expressed fewer than 500 genes, expressed more than 10,000 genes, contained fewer than 1,000 transcripts, or had more than 10% of transcripts mapping to mitochondrial genes were removed, resulting in 811 cells passing all metrics. Genes counts were normalized using "LogNormalize" and counts per 10,000 normalization. 3,000 highly variable genes were selected using the "vst" method. Normalized transcript counts were scaled to 0 mean and unit variance, and principal component analysis with 100 principal components was performed. Cells were Louvain clustered with a resolution of 1 and k = 20 nearest neighbors resulting in six clusters. Cells were embedded using t-SNE and UMAP with default parameters. A grid search of clustering parameters for k nearest neighbors in {10, 15, 20, 30} and resolution in {0.5, 1, 1.5, and 2} was performed with visual inspection of concordance between UMAP embedding and clustering to verify that k = 20 and resolution = 1 were reasonable. Normalized gene expressions for 12 marker genes (Th, Sst, Mrgprd, P2rx3, Tac1, Ntrk1, Calca, Nefh, Pvalb, B2m, Vim, Col6a2) were plotted and cluster cell types were manually annotated as Non-neuronal, Peptidergic (Pep), non-peptidergic (NP), tyrosine hydroxylase-expressing (TH) and neurofilament heavy chain (NEFH)-expressing (NF) (Fig 1). To identify cell subpopulations, each cell type was separately processed using 2000 variable features, 30 principal components, k = 10 nearest neighbors, and a louvain resolution of 1.

Differential Expression Analysis: Differentially expressed genes were identified using a Wilcoxon rank sum test for each cell type compared to all other cell types where the log2 fold change was greater than 0.25 and the gene was expressed in at least 10% of cells. A list of all known 3-UTR-binding proteins was obtained from the Gene Ontology consortium website (Ashburner et al., 2000; Gene Ontology Consortium, 2021) to highlight differentially expressed cell type specific

3'UTR-binding proteins.

mRNA co-expression analysis: Pearson's correlation was calculated by comparison of CELF4 mRNA expression values vs expression values for all other detected mRNAs. mRNAs with highest Pearson values are shown (Fig 5A).

Antibody characterization

Polyclonal rabbit anti-CELF4 antibody (Prestige Antibodies, Sigma-Aldrich, HPA037986): Developed using CUGBP, Elav-like family member 4 recombinant protein epitope signature tag (PrEST). This antibody has been previously characterized for use with immunostaining, where this antibody detected CELF4 protein in the wildtype mouse hippocampal and cortical neurons, but not in the CELF4 knockout samples (Wagnon et al., 2011, 2012).

Polyclonal guinea pig anti-TRPV1 (Neuromics, GP14100): The antigen was residues of the carboxy-terminus of rat TRPV1, YTGSLKPEDAEVFKDSMVPGEK. This antibody has been assessed with wild-type tissue, as well as TRPV1-deficient mice. No apparent immunoreactivity was observed in TRPV1-KO mouse DRG, although the antibody robustly stained the expected population of wild-type neurons (Sand et al., 2015).

Polyclonal chicken anti NF-H (EnCor, CPCA-NF-H): NF-H antibody robustly stained the expected population of medium and large diameter DRG neurons and large caliber axons. No signal was detected in NF-H negative small diameter neurons, small caliber fibers or non-neuronal cells (Fig 4), thus confirming the specificity of this reagent.

Polyclonal guinea pig anti-CGRP (20R-CP001): CGRP antibody was raised in guinea pig using

calcitonin gene-related peptide conjugated to BSA as the immunogen. This antibody has been extensively used for immunostaining DRG tissue slices to identify peptidergic neurons, and specifically stains this population (Harrison et al., 2014).

Immunofluorescence Imaging and Analysis

Immunofluorescent staining was used to characterize populations of DRG neurons that express CELF4 using reagents detailed in Table 1: CGRP for peptidergic, TRPV1 for heat responsive nociceptors, and NF-H for myelinated neurons. Additionally, the plant lectin IB4 conjugated to fluorophore was used to stain the non-peptidergic population (e.g.,(Burnstock, 2000))

Sample Collection and Processing: Animals were exsanguinated under deep pentobarbital anesthesia, before perfusion with cold 4% paraformaldehyde in PBS. DRG were immediately dissected, fixed in 4% paraformaldehyde (PFA) in PBS for 10 minutes before cryoprotection by incubating overnight in 30% sucrose. 15 µm sections were cut using a cryostat, mounted onto adhesion microscope slides, and left to dry, uncovered, at room temperature for at least 1 hour.

Immunostaining: Non-specific binding sites were blocked by incubating sections with 5% donkey serum in 0.4% Triton X-100 PBS for 1 hour, covered. CELF4 antibody was diluted 1:800, TRPV1 1:500, CGRP 1:1000 and/or NF-H 1:40000 in blocking buffer and left on sections overnight (i.e. 18-24 hours) at room temperature. The sections were then washed with 0.4% Triton X-100 PBS; the washing solution was left on for 5 minutes for the first wash, then 10 minutes for the following three washes. Secondary antibodies and counterstains were prepared 1:200 in blocking buffer, added to sections and for 2 hours at room temperature. Alexa 488-conjugated IB4 was diluted at 1:50. The sections were washed a second time using the same

procedure as the first wash. Slides were coverslipped with an antifade mountant and left overnight at room temperature before imaging.

Imaging: Multichannel fluorescent micrographs of entire longitudinal DRG sections were obtained using a Leica DMi8 epifluorescent inverted microscope with 20X objective and automated field stitching. Image analysis was performed as previously published(Harrison et al., 2014): Two sections per animal containing more than 200 neurons with visible nuclei were selected from four animals. Sections were spaced more than 50 µm apart to ensure that no single neuron could be included in the analysis twice. Using Image J software (Schneider et al., 2012), individual neurons with visible nuclei were then manually circled to generate separate regions of interest (ROIs). Mean fluorescent intensities were recorded along with the cross-sectional area from each ROI. All ROI data from both sections from individual animals were then collated into a single dataset per animal, yielding data from over 400 neurons per animal. The median staining intensity for each channel was then calculated for each of the four datasets (i.e., the median intensity of each DRG) to allow for normalization of fluorescence intensities to the mean median value (division by central tendency [median] method). First, to observe the distribution of CELF4 expression according to neuron soma size, normalized signals from all animals were collated into a single dataset and subsequently binned according to soma cross-sectional area (50µm bin sizes) (Fig. 3B). Neurons were assigned to either small (<300µm), medium (300-700µm) or large (>700µm) populations. These size divisions were chosen due to the well-established somatosensory characteristics that vary according to soma size. To identify peptidergic, non-peptidergic and myelinated populations, sections were co-stained with CGRP antibody, IB4 lectin or neurofilament heavy chain (NF-H) antibody respectively (Ruscheweyh et

al., 2007).

Gene ontology analysis of mRNAs with CELF4-binding motifs expressed in TRPV1 nociceptors

All mRNAs detected by scRNA-Seq (see above) in TRPV1+ neurons were assessed for the presence of CELF4-binding motifs using the transite package (Krismer et al., 2020a) with bioconductor (Gentleman et al., 2004; Huber et al., 2015) in R statistical software. This list of genes with CELF4-binding motifs was inputted into the PANTHER gene ontology classification system (Mi et al., 2021) for statistical overrepresentation analysis. Significant (FDR>0.05) biological process (BP) and molecular function (MF) terms are shown (Fig. 6).

RESULTS

To identify RNA-binding proteins (RBPs) expressed in populations of sensory neurons, we obtained publicly available single-cell RNA-Sequencing (scRNA-Seq) data generated from adult mouse DRG (GEO GSE59739) (Usoskin et al., 2015). Since release of those data , analysis tools have evolved significantly. In addition, gene annotations, mapping coordinates and transcript variant sequences are under constant revision. We therefore reanalyzed the raw sequencing data using current software (SEURAT 3.2.0) and annotations (Gencode vM25). Cells clustered into functionally defined populations as expected (Fig. 1): Peptidergic (Pep) that are predominantly nociceptors that express neuropeptides such as CGRP, non-peptidergic (NP) that are predominantly nociceptors that do not express these neuropeptides, tyrosine hydroxylase-

expressing (TH) small-diameter sensory neurons implicated in the pleasurable, affective component of touch and neurofilament heavy chain (NF-H)-expressing that are myelinated and respond to low threshold stimuli such as light touch. Within those clusters, neurons grouped into expected sub populations, for example TRPV1-expressing neurons (heat responsive nociceptors) subclustered within the PEP cluster, and parvalbumin-expressing neurons (chiefly proprioceptors) subclustered within the NF (myelinated) cluster (Fig 1) (Usoskin et al., 2015).

Gene expression profiles from each population (Pep, NP, TH, NF) were compared by differential gene expression analysis using Wilcoxon rank sum test with SEURAT software. Of approximately 1100 known RBPs annotated in the mouse genome, 100 are assigned the gene ontology (GO) term "mRNA 3'-UTR binding" (Table 2). Of those, 84 3'UTR-binding proteins were detectable in DRG neurons (expression > 1 CFM) and 17 were enriched in at least 1 population (Table 2). Therefore, the majority (80%) of 3'UTR-binding proteins detected in DRG neurons were ubiquitously expressed throughout all populations, as exemplified by FXR1 (Fig 2 b). Interestingly, 4 members of the CUGBP Elav-Like Family (CELF) were differentially expressed: CELFs 2, 3, 4 and 6 (Table 3, Fig 2). CELF2 was enriched in the PEP population, CELF3 in the TH population, CELF4 in the PEP population and CELF6 was enriched in both NP and PEP populations.

CELF4 had the highest mean expression value of all family members (Fig 2e) and the largest differential expression value (28.2 fold) over the mean (table 3). This family member negatively regulates neuron excitability in excitatory CNS neurons, likely by limiting translation of mRNAs encoding synaptic proteins and ion channels (Wagnon et al., 2011, 2012). However, expression

and function of CELF4 in sensory neurons has not been described. RNA concentration does not necessarily correlate with protein concentration (e.g., (Edfors et al., 2016)). Therefore, the expression of CELF4 protein was examined using quantitative immunofluorescence microscopy. By measurement of fluorescent signals from individual somata from 3 animals (approx. 200 somata/animal), we observed that 58% of small diameter (<300 µm) and 38% of medium diameter (300-700 µm) DRG neurons were CELF4 positive (Fig 3). Using established histological markers for non-peptidergic (plant lectin IB4), peptidergic (CGRP antiserum) and myelinated (NFH antiserum) neurons, we observed that CELF4 protein distribution mirrors that of CELF4 mRNA. CELF4 was detected in 53% of CGRP+, 34% of IB4-binding and 54% NF-H+ neurons (Fig. 4d). However, CELF4 protein is expressed to significantly lower levels in IB4-binding ($p<0.05$, $n=3$) and NF-H+ ($p<0.01$, $n=3$ mice) neurons.

By comparing CELF4 mRNA expression to that of all other mRNAs detected in the scRNA-Seq dataset, we determined that CELF4 expression is most highly correlated with GFRA3, TRPV1 and CALCA (gene name for CGRP) expression in adult mouse DRG neurons (Fig 5, a). By co-staining DRG sections with TRPV1 and CELF4 antisera, we confirmed that the majority CELF4 immunofluorescence signal intensity is correlated with TRPV1 signal intensity (Fig 5b,c).

Finally, we examined the function of mRNAs with CELF4-binding motifs expressed in DRG neurons. Using gene ontology enrichment analysis with PatherGO (Mi et al., 2021), we determined that transcripts expressed in CELF4+ neurons with CELF4 binding motifs are predominantly RNA-binding proteins (Fig. 6). These putative CELF4-binding mRNAs have known roles in RNA metabolism broadly, including RNA processing, sub-cellular targeting and

translation. Therefore, we predict that CELF4 regulates the fate of mRNA's in CELF4+ sensory neurons.

DISCUSSION

In the current study, we investigated the expression of 3'UTR-binding proteins in DRG, leading to the discovery that members of the CUGBP-binding ELAV-like family of RBPs are expressed in distinct populations of sensory neurons. By analysis of publicly available scRNA-Seq data, we revealed that CELF2 and CELF4 are restricted to the peptidergic (PEP) population, whereas CELF6 is present in both peptidergic and non-peptidergic (NP) populations. These populations predominantly represent different classes of C-fiber nociceptors, and in the case of PEP also a subset of A-delta fibers. Most CGRP-expressing (PEP) C-fibers also express the pro-inflammatory neuropeptide substance P (Jurának & Lembeck, 1997). CGRP and SP are pro-inflammatory peptides that play key roles in neurogenic inflammation; CGRP is a potent vasodilator and substance P induces mast cell degranulation (Lim et al., 2017). In contrast, NP neurons have been reported to suppress mast cell reactivity (Zhang et al., 2021). Notably, several studies have demonstrated that cutaneous IB4-binding neurons transduce mechanical stimuli in the non-noxious range, and optogenetic stimulation of these neurons fails to evoke nocifensive behavior unless they are sensitized by tissue injury, calling into question whether these neurons act as nociceptors in healthy tissue (Warwick et al., 2021). CELF3 is restricted to the TH population of non-peptidergic, innocuous low threshold C-fiber mechanoreceptors, which have been implicated in the pleasurable, affective component of touch (Lallemend & Ernfors, 2012; Li et al.,

2011). Immunostaining revealed that the family member with the highest expression, CELF4 is enriched in neurons that express the capsaicin receptor TRPV1 and are therefore responsive to noxious heat. CELF4 is a negative regulator of protein translation and neural excitability (Wagnon et al., 2011). We therefore propose that further study of CELF4 could uncover mechanisms that regulate translation in sensory neurons that convey pain and are important for neurogenic inflammation.

Previous studies examining the expression profile and function of CELF members have determined that these proteins are expressed in the central nervous system during development and in the adult (Table 4 (Giudice et al., 2016; Krismer et al., 2020b; Samaras et al., 2020; Schmidt et al., 2018)). The RNA-binding motifs for all members are similar, with UG repeats (Table 4), suggesting that these proteins have the ability to bind similar transcripts. However, given that expression of CELF members is restricted in a cell-type dependent manner, this could limit functional redundancy between members. Focusing on CELF4, in the CNS this member is expressed primarily in excitatory neurons, including large pyramidal cells of the cerebral cortex and hippocampus, where it tonically limits excitatory neurotransmission (Wagnon et al., 2012). In these neurons, RNA-protein interaction profiling (RIP-Seq) revealed that CELF4 predominantly binds to the 3'UTR of transcripts, and CELF4 knockout reduced mRNA stability (Wagnon et al., 2012), and resulted in increases in expression of sodium channel Na(v)1.6 (Sun et al., 2013). Therefore, CELF4 is a negative regulator of excitability of excitatory CNS neurons most likely though limiting translation in these neurons. In accordance, electrophysiological experiments using brain slices revealed that loss of CELF4 function lowered the action potential (AP) initiation threshold and increased AP gain (Sun et al., 2013).

This is the first report characterizing expression of CELF4 in populations of neurons in the peripheral nervous system, and its function in those neurons remains unknown. Extrapolating from studies performed with CNS neurons where CELF4 primarily binds to sites within the 3'UTR of mRNAs, indicated molecular functions of this protein could include alternate polyadenylation (APA) thereby coordinating the expression of 3'UTR isoforms. This would be an enticing possibility due to our previous observations of differential 3'UTR isoform expression in DRG neuron populations (Harrison et al., 2014), and mechanisms responsible for cell-type specific expression of 3'UTR variants are currently unknown. However, in the current study we did not detect CELF4 in the nucleus of DRG neurons, making this possibility less likely. Also, in the CNS CELF4 mRNA co-sediments with polysomes purified from neuropil, and CELF4 knockout increases the concentration of key excitability proteins in this compartment (Wagnon et al., 2012). Therefore, CELF4 limits local translation in CNS axons. Together with our findings, this suggests that CELF4 could also limit protein expression in populations of sensory neurons, and experiments to characterize CELF4 expression in central and peripheral endings are currently underway.

Our analyses of CELF4 binding sites on mRNAs expressed in CELF4+ neurons predict that CELF4 interacts with mRNAs encoding RNA-binding proteins required for translation. This suggests that CELF4 could be a master regulator of translation in these neurons. However, this analysis was restricted to consensus transcripts, and could not account for differential expression of 3'UTR isoforms. Longer isoforms harbor additional CELF4-binding sites. We have therefore developed an approach for detection of 3'UTR variants in RNA-Seq data (CSI-UTR) (Harrison et al., 2019), and we are currently optimizing this algorithm for analysis of single-cell sequencing

data.

CELF4 tonically limits protein translation in and excitability of neurons. Therefore, pharmacological targeting of CELF4-3'UTR interactions could lead to the development of tools to study mechanisms of protein translation in neurons. Targeting CELF4 could have wide-ranging implications for diseases of the nervous system. For example in the CNS, CELF4 knockout mice develop a seizure phenotype closely resembling an epilepsy syndrome caused by CELF4 locus (18q12) deletion in Humans (Halgren et al., 2012). Therefore, enhancing CELF4 function has been proposed as a strategy to treat epilepsy. We have determined that in the PNS CELF4 is enriched in C/Aδ neurons, a population of neurons critical for the etiology of pathological pain. Therefore, enhancing CELF4 function in nociceptors could be a novel approach to preventing/alleviating chronic pain.

FIGURE LEGENDS

Figure 1: Utilization of publicly available single cell RNA sequencing (scRNA-Seq) data, derived from adult mouse sensory neurons, to generate population-specific gene expression profiles.

To generate gene expression profiles for specific populations of mouse adult sensory neurons, publicly available scRNA-seq data (GSE59739) were downloaded and processed with the latest

release of SEURAT software using up to date gene annotations. Graphs are Uniform Manifold Approximation and Projections (UMAPs) of gene expression profiles. Each data point is a cell in UMAP space. Yellow shading = high expression, Blue shading = low expression. Expression profiles distributed into clusters as expected: **A-D** Peptidergic neurons were identified by CALCA (CGRP) expression (PEP, purple cluster), non-peptidergic by MRGPRD expression (NP, green cluster), tyrosine hydroxylase–expressing neurons (TH, red cluster) and myelinated neurons by neurofilament heavy chain expression (NF, blue cluster). **E** – Sub-cluster of TRPV1-expressing nociceptors. **F** – Sub-cluster of parvalbumin (PVALB) -expressing proprioceptors.

Figure 2: The CELF family of RNA-binding proteins are expressed in distinct populations of sensory neurons.

A Mean vs variance plot of mRNA expression values for 3'UTR-binding RNA-binding protein genes in adult mouse sensory neurons. X's are individual genes. FXR1 and CELF4 are highlighted for reference. **B** Expression of a representative RBP, FXR1 is broadly expressed in all clusters of DRG neurons. Yellow circles are individual neurons that express FXR1 to high levels, blue circles = cells with low FXR1 expression **C** Heatmap of expression values of 3'UTR-binding RBPs significantly differentially expressed (FDR<0.01, top-10) in distinct populations of sensory neurons. Each line shows expression values in a single cell. Yellow = high expression , Purple = low expression . **D** Expression of significantly differentially expressed CELF-family RBPs in DRG neuron clusters. Yellow = high expression , blue = low expression. **E** Violin plots showing the distribution of expression values for significantly differentially expressed CELF family proteins.

Yellow triangles denote significant enrichment, purple triangles denote significant exclusion. TH = tyrosine hydroxylase–expressing, NP = non-peptidergic, NF = neurofilament heavy chain-expressing and PEP = peptidergic neurons.

Figure 3: **CELF4 protein is expressed in small and medium diameter sensory neurons.**

A Representative immunofluorescent micrograph of a 14µm DRG section stained with CELF4 antibody. **B** Soma size distribution of CELF4+ neurons. Data are collated from 3 sections, from 3 DRG from 3 replicate animals (approx. 100 soma per section). Somata were categorized according to cross-sectional area - small (<300µm), medium (300-700µm) or large (>700µm). **C i** Proportion of small, medium or large cross-sectional area DRG neurons that express CELF4. **ii** Proportion of CELF4-expressing neurons with small, medium or large cross-sectional areas.

Figure 4: CELF4 protein expression in peptidergic (CGRP+), non-peptidergic (IB4-binding) and myelinated (NF-H+) neurons.

A-C 14µm DRG sections stained with CELF4 antibody, co-stained with established histological markers. Peptidergic neurons were identified with CGRP antibody, non-peptidergic with isolectin IB4 and myelinated with NFH. Scatter plots show distribution of fluorescence intensities of somata collated from DRG from 3 replicate animals, 2 non-overlapping cross sections per animal (approx. 100 soma per section). Shaded areas highlight double positive soma. **D** Pie charts showing the proportion of double positive somata from each of the 3 stains. **E** Mean CELF4

fluorescence values. Stats = ANOVA with post hoc t-tests, n=3 animals, *=p<.05, **=p<.01.

Figure 5: CELF4 is co-expressed with the TRPV1 capsaicin receptor in DRG neurons.

A Pearson's r correlation coefficients were calculated for CELF4 mRNA concentration vs all other transcripts. The 3 most positively and negatively correlated genes are shown. **B-C** 14µm DRG sections were stained with CELF4 and TRPV1 antibody, and fluorescence intensity measured in soma of 2 sections from 3 animals. Representative micrographs are shown.

Figure 6: Function of mRNAs with CELF-binding motifs expressed in peptidergic neurons.

Bar graphs showing the number of genes with CELF motifs (observed – black bars), compared to the number of genes expected by random chance (white bars). Shown are functional categories significantly enriched (FDR<0.05), >2-fold above expected. Enriched "Biological Process" (**A**) and "Molecular Function" (**B**) ontology terms are shown.

References

An, J. J., Gharami, K., Liao, G.-Y., Woo, N. H., Lau, A. G., Vanevski, F., Torre, E. R., Jones, K. R., Feng, Y., Lu, B., & Xu, B. (2008). Distinct role of long 3′ UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell*, *134*(1), 175–187. PubMed. https://doi.org/10.1016/j.cell.2008.05.045

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Berkovits, B. D., & Mayr, C. (2015). Alternative 3′UTRs act as scaffolds to regulate membrane protein localization. *Nature*, *522*(7556), 363–367. https://doi.org/10.1038/nature14321

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Burnstock, G. (2000). P2X receptors in sensory neurones. *British Journal of Anaesthesia*, *84*(4), 476–488. https://doi.org/10.1093/oxfordjournals.bja.a013473

Darnell, R. B. (2013). RNA Protein Interaction in Neurons. *Annual Review of Neuroscience*, *36*, 243–270. https://doi.org/10.1146/annurev-neuro-062912-114322

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dong, X., Han, S., Zylka, M. J., Simon, M. I., & Anderson, D. J. (2001). A diverse family of GPCRs expressed in specific subsets of nociceptive sensory neurons. *Cell*, *106*(5), 619–632. https://doi.org/10.1016/s0092-8674(01)00483-4

Edfors, F., Danielsson, F., Hallström, B. M., Käll, L., Lundberg, E., Pontén, F., Forsström, B., & Uhlén, M. (2016). Gene-specific correlation of RNA and protein levels in human cells and tissues. *Molecular Systems Biology*, *12*(10). https://doi.org/10.15252/msb.20167144

Gene Ontology Consortium. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., … Zhang, J. (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, *5*(10), R80. https://doi.org/10.1186/gb-2004-5-10-r80

Giudice, G., Sánchez-Cabo, F., Torroja, C., & Lara-Pezzi, E. (2016). ATtRACT-a database of RNA-binding proteins and associated motifs. *Database: The Journal of Biological Databases and Curation*, *2016*, baw035. https://doi.org/10.1093/database/baw035

Graber, J. H., Nazeer, F. I., Yeh, P., Kuehner, J. N., Borikar, S., Hoskinson, D., & Moore, C. L. (2013). DNA damage induces targeted, genome-wide variation of poly(A) sites in budding yeast. *Genome Research*, *23*(10), 1690–1703. https://doi.org/10.1101/gr.144964.112

Grimson, A., Farh, K. K.-H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Molecular Cell*, *27*(1), 91–105. https://doi.org/10.1016/j.molcel.2007.06.017

Halgren, C., Bache, I., Bak, M., Myatt, M. W., Anderson, C. M., Brøndum-Nielsen, K., & Tommerup, N. (2012). Haploinsufficiency of CELF4 at 18q12.2 is associated with developmental and behavioral disorders, seizures, eye manifestations, and obesity. *European Journal of Human Genetics*, *20*(12), 1315–1319. https://doi.org/10.1038/ejhg.2012.92

Harrison, B. J., Flight, R. M., Eteleeb, A., Rouchka, E. C., & Petruska, J. C. (2012). RNASeq profiling of UTR expression during neuronal plasticity. *BMC Bioinformatics*, *13*(12), A4. https://doi.org/10.1186/1471-2105-13-S12-A4

Harrison, B. J., Flight, R. M., Gomes, C., Venkat, G., Ellis, S. R., Sankar, U., Twiss, J. L., Rouchka, E. C., & Petruska, J. C. (2014). IB4-binding sensory neurons in the adult rat express a novel 3′ UTR-extended isoform of CaMK4 that is associated with its localization to axons. *Journal of Comparative Neurology*, *522*(2), 308–336. https://doi.org/10.1002/cne.23398

Harrison, B. J., Park, J. W., Gomes, C., Petruska, J. C., Sapio, M. R., Iadarola, M. J., Chariker, J. H., & Rouchka, E. C. (2019). Detection of Differentially Expressed Cleavage Site Intervals Within 3′ Untranslated Regions Using CSI-UTR Reveals Regulated Interaction Motifs. *Frontiers in Genetics*, *10*, 182. https://doi.org/10.3389/fgene.2019.00182

Hornberg, H. K., & Holt, C. (2013). RNA-binding proteins and translational regulation in axons and growth cones. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00081

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M.,

Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Ji, Z., & Tian, B. (2009). Reprogramming of 3′ Untranslated Regions of mRNAs by Alternative Polyadenylation in Generation of Pluripotent Stem Cells from Different Cell Types. *PLoS ONE*, *4*(12). https://doi.org/10.1371/journal.pone.0008419

Juránek, I., & Lembeck, F. (1997). Afferent C-fibres release substance P and glutamate. *Canadian Journal of Physiology and Pharmacology*, *75*(6), 661–664.

Kislauskis, E. H., & Singer, R. H. (1992). Determinants of mRNA localization. *Current Opinion in Cell Biology*, *4*(6), 975–978. https://doi.org/10.1016/0955-0674(92)90128-y

Kislauskis, E. H., Zhu, X., & Singer, R. H. (1994). Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. *The Journal of Cell Biology*, *127*(2), 441–451. https://doi.org/10.1083/jcb.127.2.441

Krismer, K., Bird, M. A., Varmeh, S., Handly, E. D., Gattinger, A., Bernwinkler, T., Anderson, D. A., Heinzel, A., Joughin, B. A., Kong, Y. W., Cannell, I. G., & Yaffe, M. B. (2020a). Transite: A Computational Motif-Based Analysis Platform That Identifies RNA-Binding Proteins Modulating Changes in Gene Expression. *Cell Reports*, *32*(8). https://doi.org/10.1016/j.celrep.2020.108064

Krismer, K., Bird, M. A., Varmeh, S., Handly, E. D., Gattinger, A., Bernwinkler, T., Anderson, D. A., Heinzel, A., Joughin, B. A., Kong, Y. W., Cannell, I. G., & Yaffe, M. B. (2020b). Transite: A Computational Motif-Based Analysis Platform That Identifies RNA-Binding Proteins

Modulating Changes in Gene Expression. *Cell Reports*, *32*(8), 108064. https://doi.org/10.1016/j.celrep.2020.108064

Lallemend, F., & Ernfors, P. (2012). Molecular interactions underlying the specification of sensory neurons. *Trends in Neurosciences*, *35*(6), 373–381. https://doi.org/10.1016/j.tins.2012.03.006

Lawson, S. N., McCarthy, P. W., & Prabhakar, E. (1996). Electrophysiological properties of neurones with CGRP-like immunoreactivity in rat dorsal root ganglia. *The Journal of Comparative Neurology*, *365*(3), 355–366. https://doi.org/10.1002/(SICI)1096-9861(19960212)365:3<355::AID-CNE2>3.0.CO;2-3

Lawson, S. N., Perry, M. J., Prabhakar, E., & McCarthy, P. W. (1993). Primary sensory neurones: Neurofilament, neuropeptides, and conduction velocity. *Brain Research Bulletin*, *30*(3–4), 239–243. https://doi.org/10.1016/0361-9230(93)90250-f

Le Pichon, C. E., & Chesler, A. T. (2014). The functional and anatomical dissection of somatosensory subpopulations using mouse genetics. *Frontiers in Neuroanatomy*, *8*, 21–21. PubMed. https://doi.org/10.3389/fnana.2014.00021

Li, L., Rutlin, M., Abraira, V. E., Cassidy, C., Kus, L., Gong, S., Jankowski, M. P., Luo, W., Heintz, N., Koerber, H. R., Woodbury, C. J., & Ginty, D. D. (2011). The functional organization of cutaneous low-threshold mechanosensory neurons. *Cell*, *147*(7), 1615–1627. https://doi.org/10.1016/j.cell.2011.11.027

Lim, J. E., Chung, E., & Son, Y. (2017). A neuropeptide, Substance-P, directly induces tissue-repairing M2 like macrophages by activating the PI3K/Akt/mTOR pathway even in the

presence of IFNγ. *Scientific Reports*, *7*(1), 9417. https://doi.org/10.1038/s41598-017-09639-7

Majoros, W. H., & Ohler, U. (2007). Spatial preferences of microRNA targets in 3′ untranslated regions. *BMC Genomics*, *8*(1), 152. https://doi.org/10.1186/1471-2164-8-152

Mayr, C., & Bartel, D. P. (2009). Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*, *138*(4), 673–684. https://doi.org/10.1016/j.cell.2009.06.016

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T., & Thomas, P. D. (2021). PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research*, *49*(D1), D394–D403. https://doi.org/10.1093/nar/gkaa1106

Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O., & Lai, E. C. (2013). Widespread and extensive lengthening of 3′ UTRs in the mammalian brain. *Genome Research*, *23*(5), 812–825. https://doi.org/10.1101/gr.146886.112

Price, T. J., & Flores, C. M. (2007). Critical evaluation of the colocalization between calcitonin gene-related peptide, substance P, transient receptor potential vanilloid subfamily type 1 immunoreactivities, and isolectin B4 binding in primary afferent neurons of the rat and mouse. *The Journal of Pain: Official Journal of the American Pain Society*, *8*(3), 263–272. https://doi.org/10.1016/j.jpain.2006.09.005

Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L., & Singer, R. H. (1997). Characterization of a beta-actin mRNA zipcode-binding protein. *Molecular and Cellular Biology*, *17*(4), 2158–2165.

Ruscheweyh, R., Forsthuber, L., Schoffnegger, D., & Sandkühler, J. (2007). Modification of classical neurochemical markers in identified primary afferent neurons with Aβ-, Aδ-, and C-fibers after chronic constriction injury in mice. *Journal of Comparative Neurology*, *502*(2), 325–336. https://doi.org/10.1002/cne.21311

Samaras, P., Schmidt, T., Frejno, M., Gessulat, S., Reinecke, M., Jarzab, A., Zecha, J., Mergner, J., Giansanti, P., Ehrlich, H.-C., Aiche, S., Rank, J., Kienegger, H., Krcmar, H., Kuster, B., & Wilhelm, M. (2020). ProteomicsDB: A multi-omics and multi-organism resource for life science research. *Nucleic Acids Research*, *48*(D1), D1153–D1163. https://doi.org/10.1093/nar/gkz974

Sand, C. A., Grant, A. D., & Nandi, M. (2015). Vascular Expression of Transient Receptor Potential Vanilloid 1 (TRPV1). *Journal of Histochemistry and Cytochemistry*, *63*(6), 449–453. https://doi.org/10.1369/0022155415581014

Schmidt, T., Samaras, P., Frejno, M., Gessulat, S., Barnert, M., Kienegger, H., Krcmar, H., Schlegl, J., Ehrlich, H.-C., Aiche, S., Kuster, B., & Wilhelm, M. (2018). ProteomicsDB. *Nucleic Acids Research*, *46*(D1), D1271–D1281. https://doi.org/10.1093/nar/gkx1029

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Shiers, S. I., Sankaranarayanan, I., Jeevakumar, V., Cervantes, A., Reese, J. C., & Price, T. J. (n.d.). Convergence of peptidergic and non-peptidergic protein markers in the human dorsal root ganglion and spinal dorsal horn. *Journal of Comparative Neurology*, *n/a*(n/a). https://doi.org/10.1002/cne.25122

Silverman, J. D., & Kruger, L. (1988). Lectin and neuropeptide labeling of separate populations of dorsal root ganglion neurons and associated "nociceptor" thin axons in rat testis and cornea whole-mount preparations. *Somatosensory Research*, *5*(3), 259–267. https://doi.org/10.3109/07367228809144630

Srikantan, S., Tominaga, K., & Gorospe, M. (2012). Functional interplay between RNA-binding protein HuR and microRNAs. *Current Protein & Peptide Science*, *13*(4), 372–379.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Sun, W., Wagnon, J. L., Mahaffey, C. L., Briese, M., Ule, J., & Frankel, W. N. (2013). Aberrant sodium channel activity in the complex seizure disorder of Celf4 mutant mice. *The Journal of Physiology*, *591*(1), 241–255. https://doi.org/10.1113/jphysiol.2012.240168

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lönnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggström, J., Kharchenko, O., Kharchenko, P. V., Linnarsson, S., & Ernfors, P. (2015a). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nature Neuroscience*, *18*(1), 145–153. https://doi.org/10.1038/nn.3881

Wagnon, J. L., Briese, M., Sun, W., Mahaffey, C. L., Curk, T., Rot, G., Ule, J., & Frankel, W. N. (2012). CELF4 Regulates Translation and Local Abundance of a Vast Set of mRNAs, Including Genes Associated with Regulation of Synaptic Function. *PLoS Genetics*, *8*(11). https://doi.org/10.1371/journal.pgen.1003067

Wagnon, J. L., Mahaffey, C. L., Sun, W., Yang, Y., Chao, H.-T., & Frankel, W. N. (2011). Etiology of a genetically complex seizure disorder in Celf4 mutant mice. *Genes, Brain, and Behavior*, *10*(7), 765–777. https://doi.org/10.1111/j.1601-183X.2011.00717.x

Warwick, C., Cassidy, C., Hachisuka, J., Wright, M. C., Baumbauer, K. M., Adelman, P. C., Lee, K. H., Smith, K. M., Sheahan, T. D., Ross, S. E., & Koerber, H. R. (2021). MrgprdCre lineage neurons mediate optogenetic allodynia through an emergent polysynaptic circuit. *Pain*, *162*(7), 2120–2131. https://doi.org/10.1097/j.pain.0000000000002227

Zhang, S., Edwards, T. N., Chaudhri, V. K., Wu, J., Cohen, J. A., Hirai, T., Rittenhouse, N., Schmitz, E. G., Zhou, P. Y., McNeil, B. D., Yang, Y., Koerber, H. R., Sumpter, T. L., Poholek, A. C., Davis, B. M., Albers, K. M., Singh, H., & Kaplan, D. H. (2021). Nonpeptidergic neurons suppress mast cells via glutamate to maintain skin homeostasis. *Cell*, *184*(8), 2151-2166.e16. https://doi.org/10.1016/j.cell.2021.03.002

Zheng, D., Wang, R., Ding, Q., Wang, T., Xie, B., Wei, L., Zhong, Z., & Tian, B. (2018). Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. *Nature Communications*, *9*(1), 2268. https://doi.org/10.1038/s41467-018-04730-7

Zwick, M., Davis, B. M., Woodbury, C. J., Burkett, J. N., Koerber, H. R., Simpson, J. F., & Albers, K. M. (2002). Glial cell line-derived neurotrophic factor is a survival factor for isolectin B4-positive, but not vanilloid receptor 1-positive, neurons in the mouse. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*(10), 4057–4065. https://doi.org/20026394

Target	*Immunogen*	*RRID*	*Host*	*Supplier*	*Cat #*	*Dilution Factor*
CELF4	MYIKMATLANGQA DNASLSTNGLGSSPG SAGHMNGLSHSPG NPSTIPMKDH	AB_10673035	Rabbit	Millipore Sigma	HPA037986	1:800
TRPV1	YTGSLKPEDAEVFKD SMVPGEK	AB_1624142	Guinea Pig	Neuromics	GP14100	1:500
NF-H	NF-H purified from bovine spinal cord	AB_2149761	Chicken	EnCor	CPCA-NF-H	1:40000
CGRP	ACDTATCVTHRLAGL LSRSGGVVKNNFVPT NVGSKAF	AB_1282813	Guinea Pig	Fitzgerald Industries International	20R-CP007	1:1500

Table 1: Primary antisera used for this study

171mm
113mm
84mm
234mm

Gene Ontology (GO) Term	*3'UTR-binding proteins (100)*	*Detected in DRG neurons (84)*	*Enriched in DRG neuron population(s) (17)*
posttranscriptional regulation of gene expression (GO:0010608)	8	8	1
regulation of mRNA metabolic process (GO:1903311)	33	32	8
regulation of translation (GO:0006417)	43	35	4
regulation of mRNA catabolic process (GO:0061013)	33	11	
regulation of mRNA stability (GO:0043488)	39	33	5
mRNA processing (GO:0006397)	45	43	10
RNA splicing (GO:0008380)	32	31	10
RNA localization (GO:0006403)	5	4	
regulation of gene silencing by miRNA (GO:0060964)	10	9	
RNA transport (GO:0050658)	10	7	1
cytoplasmic ribonucleoprotein granule (GO:0036464)	34	28	4
growth cone (GO:0030426)	6	6	2
dendritic spine (GO:0043197)	5	5	1
ribosome (GO:0005840)	6	6	2
distal axon (GO:0150034)	4	4	
nucleolus (GO:0005730)	16	16	1
dendrite (GO:0030425)	13	12	2
postsynaptic density (GO:0014069)	7	7	1
nuclear body (GO:0016604)	13	13	2
neuronal cell body (GO:0043025)	10	10	3
postsynapse (GO:0098794)	8	8	2

Table 2: Function of 3'UTR mRNA-binding proteins in populations of adult mouse sensory (DRG) neurons. 100 mouse genes are assigned the "3'UTR-binding" ontology term (Ashburner et al., 2000). mRNA's for 84 of these were detected in publicly available scRNA-Seq data derived from adult mouse DRG neurons (GSE59739, (Usoskin et al., 2015). 17 were significantly differentially expressed (FDR<0.05) between one or more of 4 DRG neuron populations (tyrosine hydroxylase–expressing, peptidergic, non-peptidergic, myelinated).

Gene	Population	Average Enrichment	p
CELF4	PEP	28.11	4.07E-12
CARHSP1	NP	15.76	8.95E-63
RBMS3	TH	5.52	4.68E-33
AUH	NF	3.64	1.67E-35
CELF6	PEP	2.98	2.25E-06
CELF6	NP	2.94	1.86E-20
CELF2	PEP	2.58	0.002704
CELF3	TH	2.53	1.35E-21
HNRNPA2B1	TH	2.21	1.65E-10
FXR2	NF	2.02	1.24E-14
ELAVL4	TH	2.02	5.72E-13
CSDC2	NF	1.93	1.37E-25
FXR2	NP	1.92	2.30E-05
RBMS1	NP	1.88	2.24E-13
RBFOX1	TH	1.78	0.000245
FXR2	PEP	0.56	0.006902
AUH	NP	0.46	0.003118
YBX3	NF	0.45	0.007494
TAF15	NF	0.40	2.32E-15
CELF6	NF	0.36	1.05E-07
CELF4	TH	0.31	1.38E-05
RBMS3	NP	0.29	3.77E-05
HNRNPL	NF	0.23	4.45E-05
RPS7	NF	0.20	0.000137
CARHSP1	NF	0.18	0.000383
CARHSP1	TH	0.17	3.98E-09

Table 3: 3'UTR mRNA-binding proteins significantly enriched (FDR<0.05) in DRG neuron populations, sorted by average fold enrichment. TH = tyrosine hydroxylase–expressing, NP = non-peptidergic, NF = neurofilament heavy chain-expressing and PEP = peptidergic neurons.

171mm
113mm
84mm
234mm

Family member	*Tissue Expression (Adult)*	*Subcellular Location*	*RNA-binding motif*	*Molecular/biological function(s)*
CELF1	Brain (m), spleen (m),liver(m), testis(m) ubiquitous(h)	Nuclear and cytoplasmic	UGUGUG,UGUU, CUGUCUG,UUGUG,	splicing,RNA stability, translation regulation (1-5)
CELF2	Brain (m), spleen (m),heart(m), lung(m),testis(m), ubiquitous(h)	Nucleoplasm, vesicles	UGUGU, UUGUU,UGUU,	RNA binding, mRNA processing,splicing
CELF3	Brain (m), adrenal gland (h),retina (h), pancreas (h)	Nucleus, cytoplasm	UGUUGUG (h)	RNA binding,mRNA splicing,spermatogenesis
CELF4	Brain (m,h), adrenal gland (h), retina (h)	Nucleus, cytoplasm, Axons/Dendrites	UGUGUKK(h), UGUGUGU*	RNA binding, mRNA processing, mRNA splicing
CELF5	Brain (m), N/A(h)	Nucleus, cytoplasm	UGUGUKK (h), UGUGUGU*	RNA binding
CELF6	Brain (m), N/A (h)	Nucleus, cytoplasm, synapse	UGUGKKG (h)	RNA binding, mRNA processing

Table 4: **CELF family RNA-binding protein tissue distribution, subcellular location, RNA recognition motifs and summarized function**. Tissue expression data from ProteomicsDB (Samaras et al., 2020; Schmidt et al., 2018). RNA-binding motifs were obtained from Transite (Krismer et al., 2020) and ATtRACT (Giudice et al., 2016) databases. * reported consensus sequence

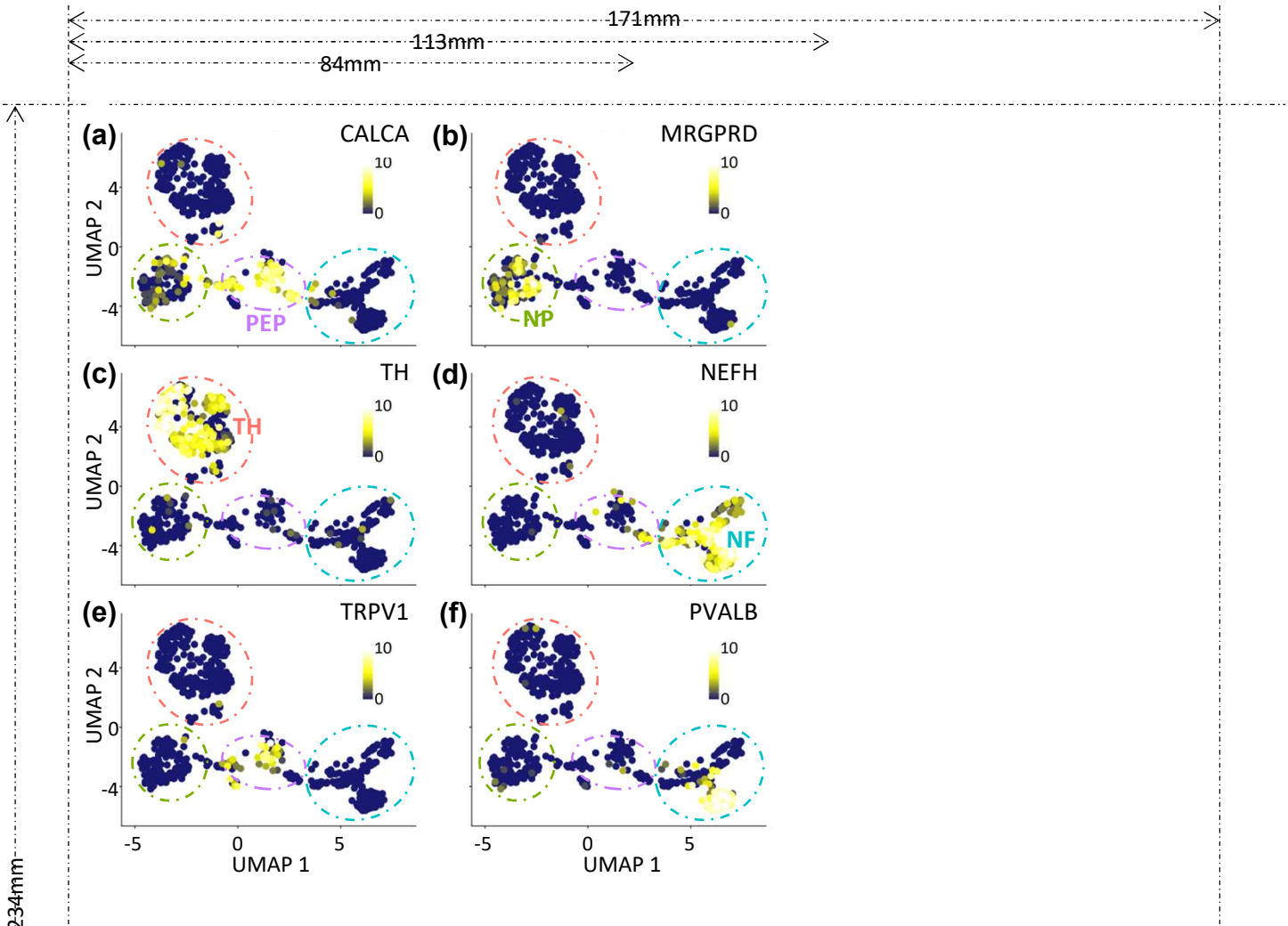


Figure 1 (Methods)

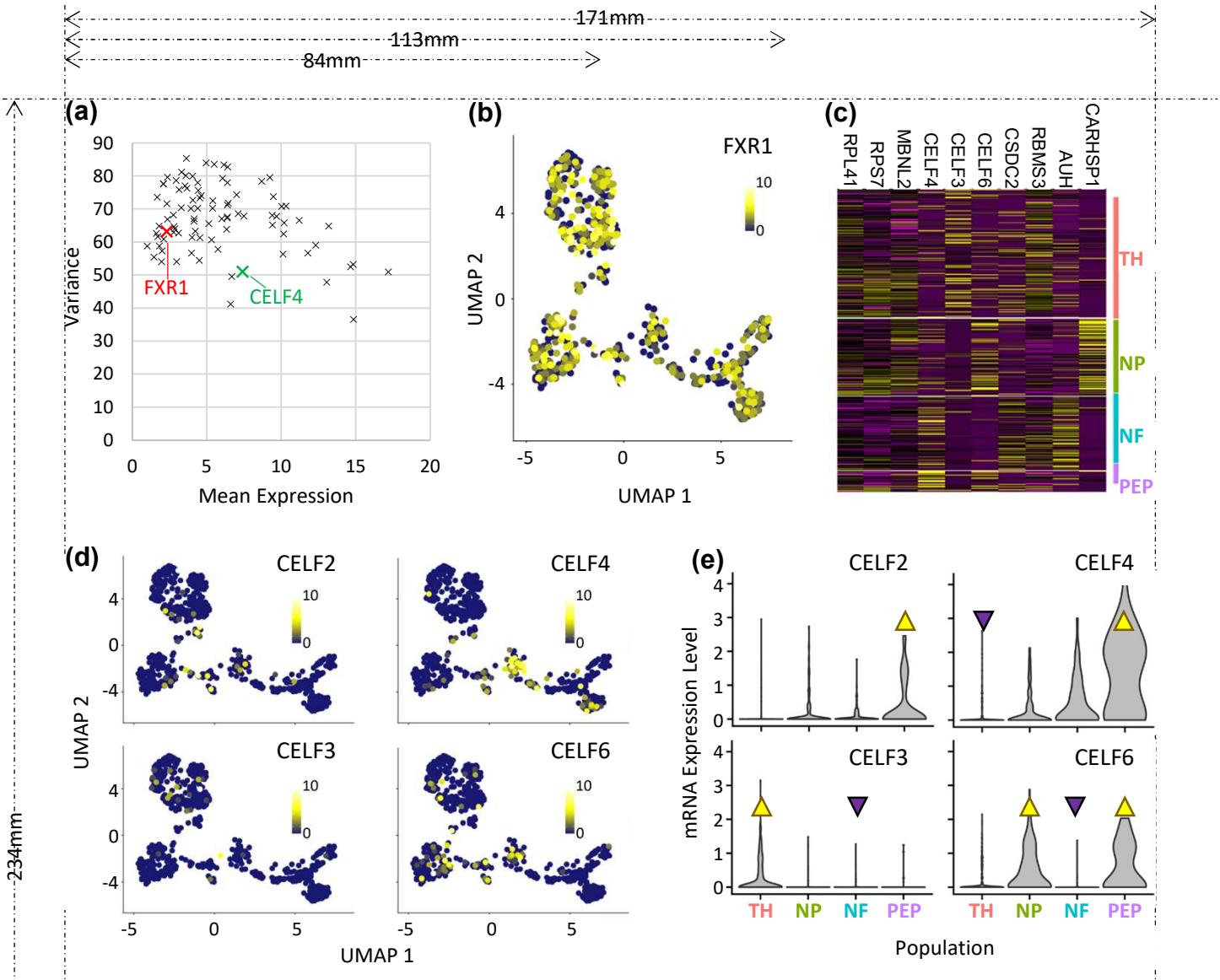
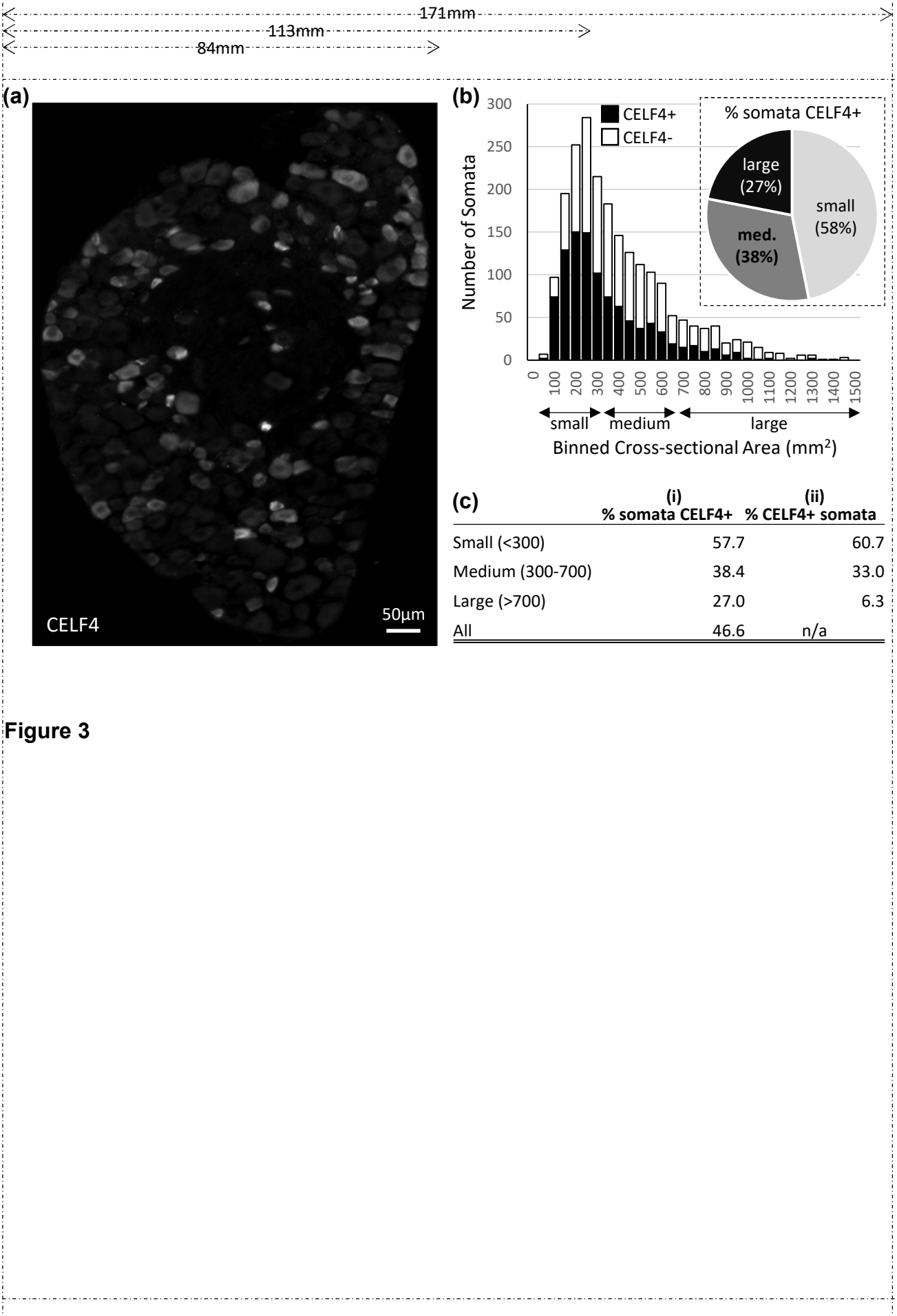


Figure 2



(c)

	(i) % somata CELF4+	(ii) % CELF4+ somata
Small (<300)	57.7	60.7
Medium (300-700)	38.4	33.0
Large (>700)	27.0	6.3
All	46.6	n/a

Figure 3

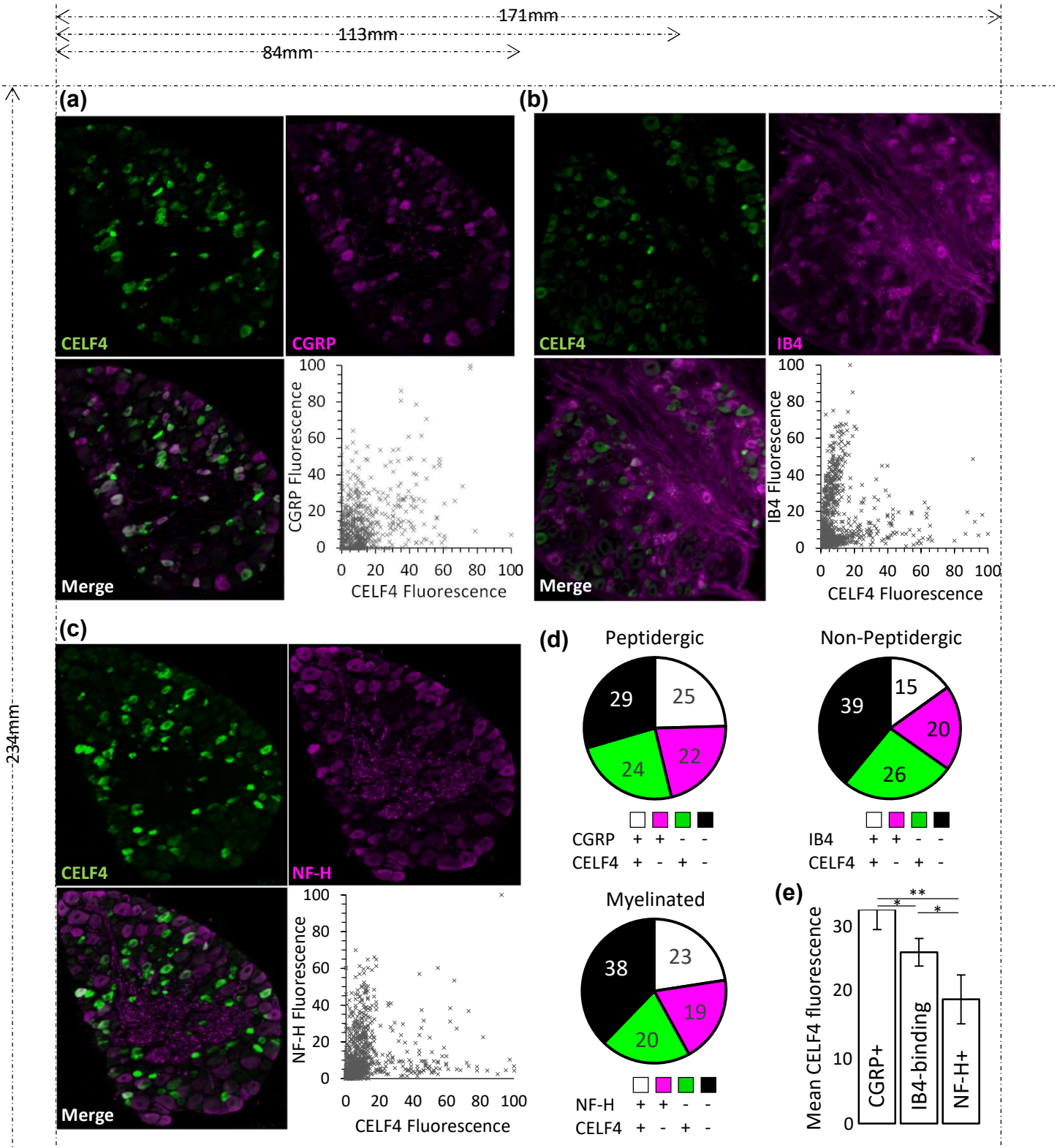


Figure 4

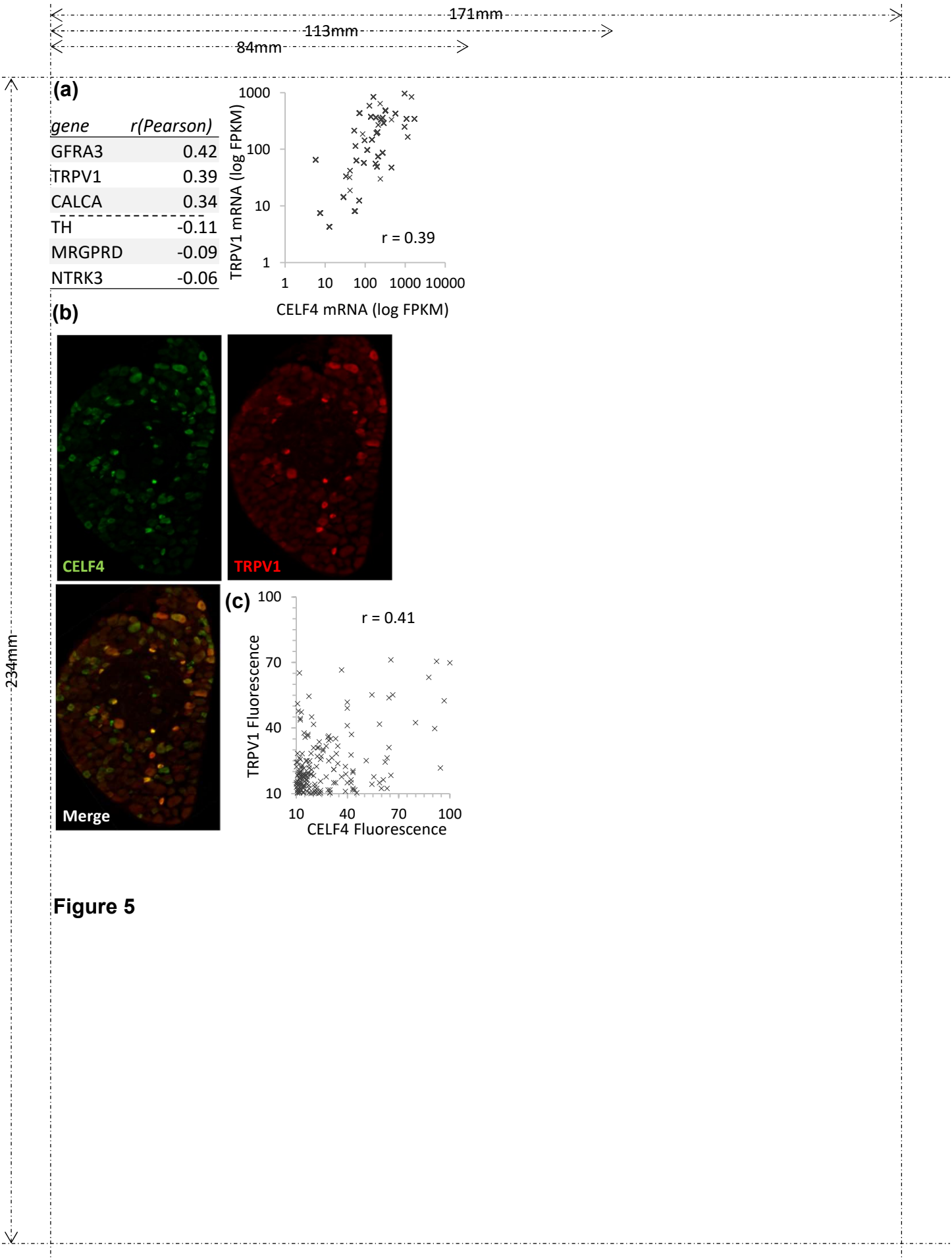


Figure 5

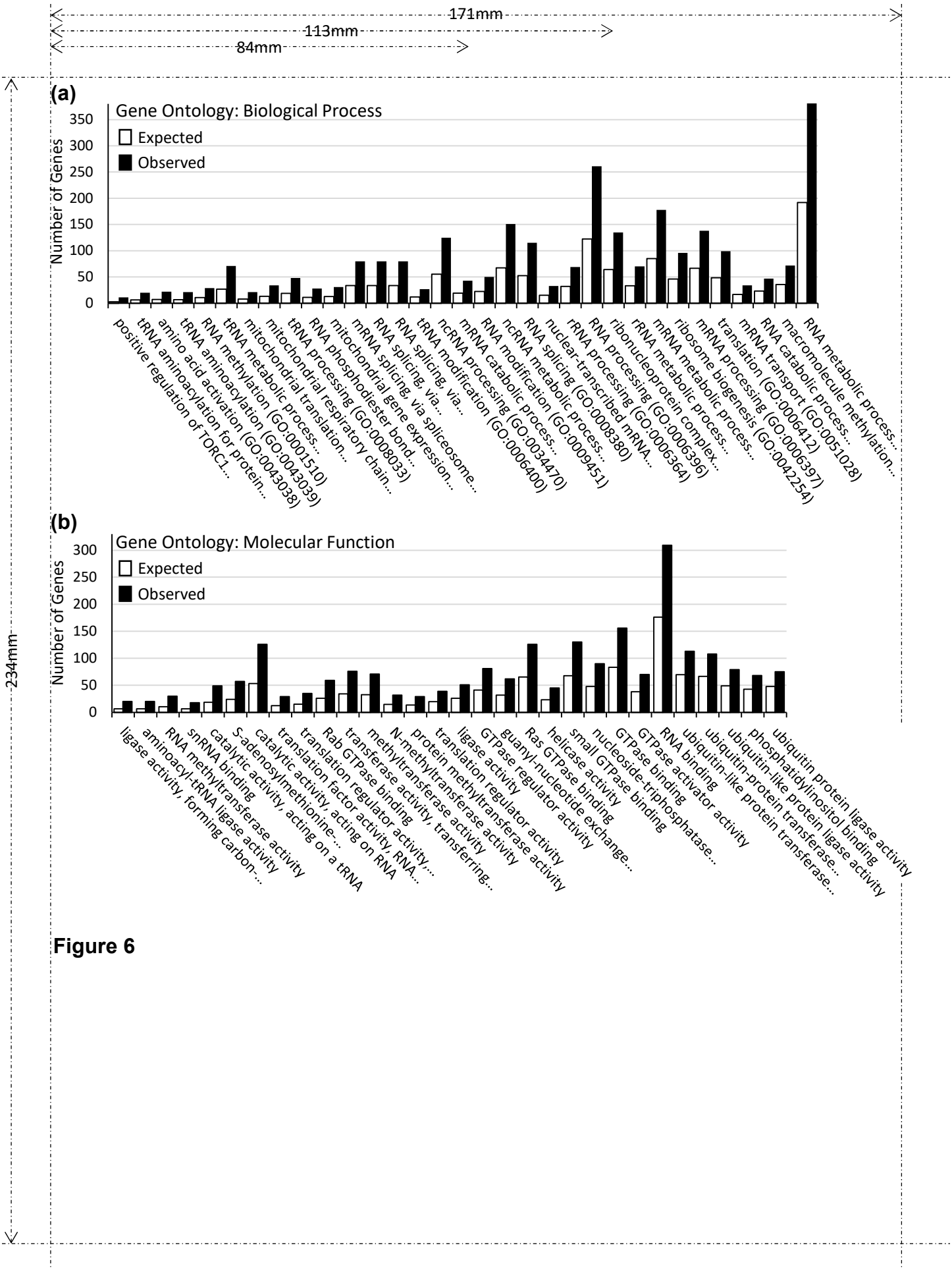


Figure 6