

RESEARCH ARTICLE

Open Access



# Comprehensive catalog of dendritically localized mRNA isoforms from sub-cellular sequencing of single mouse neurons

Sarah A. Middleton<sup>1,4</sup>, James Eberwine<sup>2</sup> and Junhyong Kim<sup>1,3\*</sup> 

## Abstract

**Background:** RNA localization involves cis-motifs that are recognized by RNA-binding proteins (RBP), which then mediate localization to specific sub-cellular compartments. RNA localization is critical for many different cell functions, e.g., in neuronal dendrites, localization is a critical step for long-lasting synaptic potentiation. However, there is little consensus regarding which RNAs are localized and the role of alternative isoforms in localization. A comprehensive catalog of localized RNA can help dissect RBP/RNA interactions and localization motifs. Here, we utilize a single cell sub-cellular RNA sequencing approach to profile differentially localized RNAs from individual cells across multiple single cells to help identify a consistent set of localized RNA in mouse neurons.

**Results:** Using independent RNA sequencing from soma and dendrites of the same neuron, we deeply profiled the sub-cellular transcriptomes to assess the extent and variability of dendritic RNA localization in individual hippocampal neurons, including an assessment of differential localization of alternative 3'UTR isoforms. We identified 2225 dendritic RNAs, including 298 cases of 3'UTR isoform-specific localization. We extensively analyzed the localized RNAs for potential localization motifs, finding that B1 and B2 SINE elements are up to 5.7 times more abundant in localized RNA 3'UTRs than non-localized, and also functionally characterized the localized RNAs using protein structure analysis.

**Conclusion:** We integrate our list of localized RNAs with the literature to provide a comprehensive list of known dendritically localized RNAs as a resource. This catalog of transcripts, including differentially localized isoforms and computationally hypothesized localization motifs, will help investigators further dissect the genome-scale mechanism of RNA localization.

**Keywords:** RNA localization, Single-cell RNA-seq, Sub-cellular RNA-seq, RNA secondary structure, SINE

## Background

RNA localization is critical to many inter-cellular processes. Neurons are an excellent system to study RNA localization because their extreme polar morphology (neurites of the neurons) creates clear spatial differentiation of the localized RNA and makes it relatively easy to isolate the localized RNA. In addition, RNA localization is critical to neuronal function in which neurons require

local protein synthesis within the dendrites to produce long-lasting synaptic potentiation [1–3]. In order for this local synthesis to occur, mRNAs must first be transported to the dendrites. Although RNA localization and local translation have been studied for over 20 years, including initial Sanger sequencing of isolated single dendrite RNA [4, 5], a more detailed and thorough analysis is required to generate a consensus set of dendritically localized RNAs. Surprisingly, the advent of high-throughput sequencing has not greatly improved matters: of three recent RNA-seq studies of dendritically localized RNA [6–8], only 1% of the identified RNAs overlapped between all three studies (44 of 4441). Although these differences can be partly attributed to differences in sample origin, organism, and experimental protocol between each study, these examples

\* Correspondence: [junhyong@sas.upenn.edu](mailto:junhyong@sas.upenn.edu)

<sup>1</sup>Graduate Program in Genomics and Computational Biology, Biomedical Graduate Studies, University of Pennsylvania, 160 BRB II/III - 421 Curie Blvd, Philadelphia, PA 19104-6064, USA

<sup>3</sup>Department of Biology, University of Pennsylvania, 415 S. University Ave, Philadelphia, PA 19104, USA

Full list of author information is available at the end of the article



nonetheless point to a need for further studies to understand the full range and variability of dendritic RNAs.

There are several major challenges in profiling the dendritic transcriptome: (1) cleanly separating the somatic and dendritic compartments so that they can be profiled separately, (2) differentiating transcript variation (e.g., alternative 3'UTRs) in addition to localization, (3) accounting for single cell variation in both somatic expression and dendritic localization, and (4) distinguishing actively translocated RNA from randomly diffused RNA. Here, we approach these challenges by performing simultaneous RNA sequencing of the somatic and dendritic compartments of single neurons from primary cultures to allow for a direct contrast of the dendritic transcriptome with its parent soma and to enable the assessment of heterogeneity of localization across neurons. Approaching the problem by single-neuron-matched sub-cellular sequencing has two advantages. First, by matching the dendrite and soma samples, we are able to more clearly identify differentially expressed, and therefore likely to be actively translocated, RNA. Second, since randomly diffused RNA is likely to be different in different individual cells, by examining the variability and consistency across the individual cells, we are again likely to identify actively translocated RNA. Our approach also allows us to examine individual cell variability in localization as well as effects of isoform usage. Given that substantial gene expression heterogeneity has already been observed on the whole-neuron level [9], it would not be surprising if there is variability of localization across cells, as was found in an early single dendrite Sanger sequencing study [4]. In addition, localization variability in neurons may arise from the use of alternative 3'UTR isoforms. Neurons uniquely express a large number of extended 3'UTR isoforms that are conserved between human and mouse [10], and one possibility is that a subset of these 3'UTRs contain dendritic localization signals. A few specific examples of differentially localized 3'UTR isoforms have already been characterized [11], such as BDNF [12, 13], and this phenomenon was recently surveyed on a genome-wide scale in brain-derived cell lines and cortical neurons [8] and rat hippocampal tissue slices [14] resulting in the identification of hundreds of cases of differential localization of alternative 3'UTR isoforms. Using our single neuron sub-cellular sequencing approach, we identify dendritically enriched RNAs on both the gene and isoform levels, including several of the recently identified neuron-enriched distal 3'UTR extensions [10]. We identify a total of 2225 candidate dendritic RNAs, including 298 that showed differential localization of 3'UTR isoforms that was consistent across the individual cells. Using structure- and sequence-based computational techniques, we extensively annotate these dendritic RNAs to explore their functions and identify

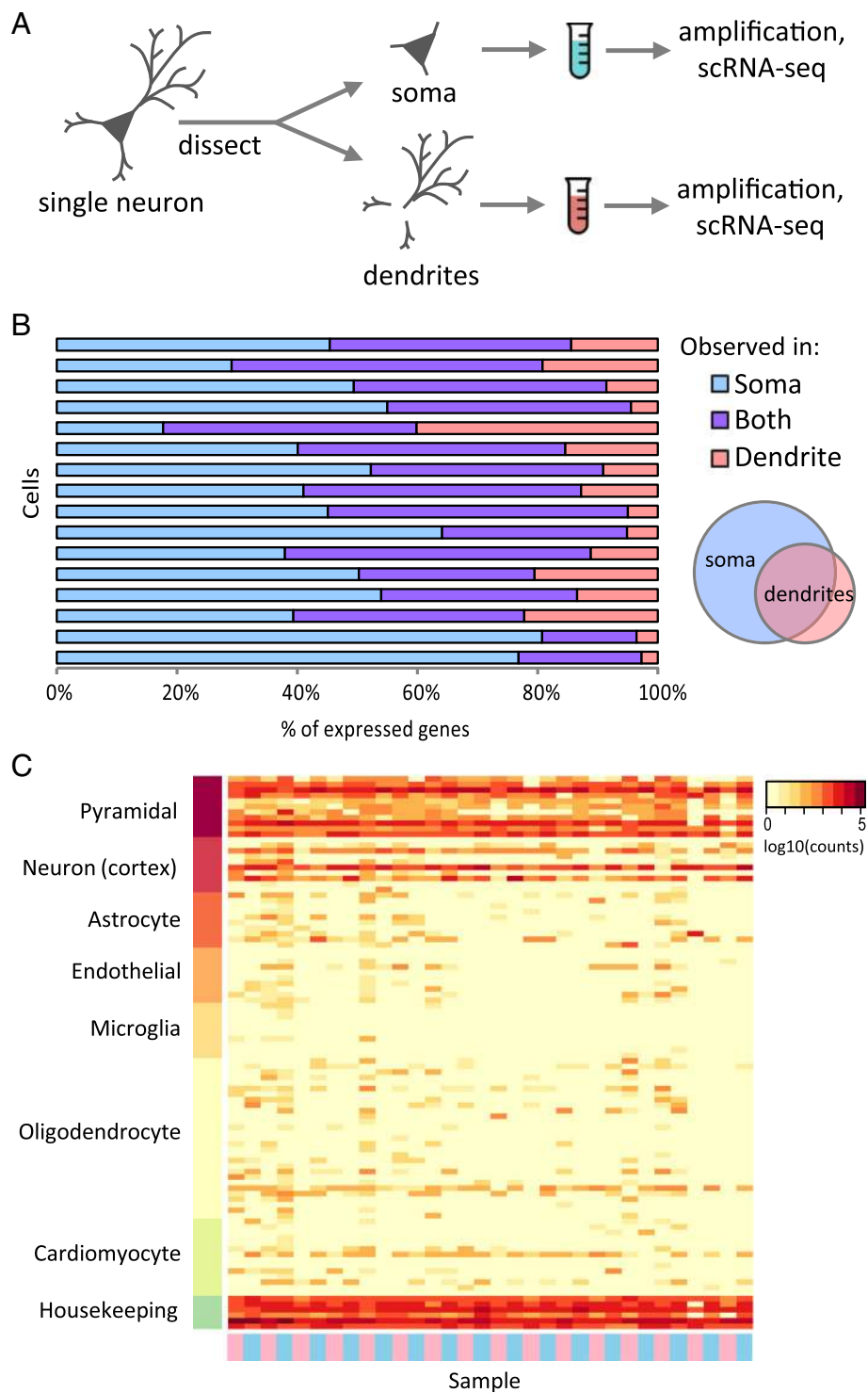
possible motifs involved in dendritic targeting. These new computational models provide a library of testable predictions that will help dissect the molecular mechanism of dendritic localization and dendritic RNA function. Finally, we integrate our list of dendritic genes with the current literature, producing a definitive list of dendritic RNAs that have been observed to date in high-throughput studies.

## Results

### Identification of dendritically localized RNAs

To compare the RNAs present in dendrites and somas of individual neurons, we manually separated the dendrites and soma of primary mouse hippocampal neurons using a micropipette [4] and performed RNA sequencing on each sub-cellular fraction such that we obtained the sub-cellular transcriptomes of the same cell (Fig. 1a). We note that the axon is generally small at this culture stage (~5% the volume of the dendrites) with a thin gauge (< 1µM) and has a flush axon hillock which is easily distinguishable from a dendrite's graded hillock. Thus, we do not expect the axon to be harvested in our procedure, and any axon that was collected would not make up a large fraction of the isolated dendrite samples. A total of 16 individual neurons were collected (32 soma and dendrite samples). Extracted RNA was amplified using the aRNA procedure [15–17] and sequenced to an average depth of 25 million reads per sample. Somas generally contained a wider variety of transcripts than their corresponding dendrites, with an average of 9206 and 5827 genes identified in each compartment respectively. As expected, the genes represented in the dendrites were largely a subset of the soma-expressed genes of the same cell (Fig. 1b). Due to detection limitations of single-cell sequencing, it is possible that some of the genes found only in soma or dendrites are present in the other compartment but simply not captured during sequencing (dropouts). For example, for each individual cell, the genes specific to the dendrites were generally more lowly expressed than genes shared between soma and dendrites (136.2 and 448.8 reads on average, respectively), suggesting that they are more prone to dropout, which may explain the absence in the soma of some dendritically observed genes. Nonetheless, due to the high sequencing depth used in this study as compared to a typical single cell study, we were able to characterize the transcriptomes of each compartment relatively deeply as illustrated by the number of detected genes. All soma and dendrite samples expressed housekeeping genes and neuronal marker genes at high levels, especially pyramidal cell markers such as *Grin1*, *Mtap2*, and *Neurod6*, with little expression of other brain cell type markers (Fig. 1c).

To identify potentially localized RNAs, we used DESeq2 [18] to perform a differential expression analysis using a paired design, where soma and dendrites of the same original cell were directly compared. DESeq2 reported 3811

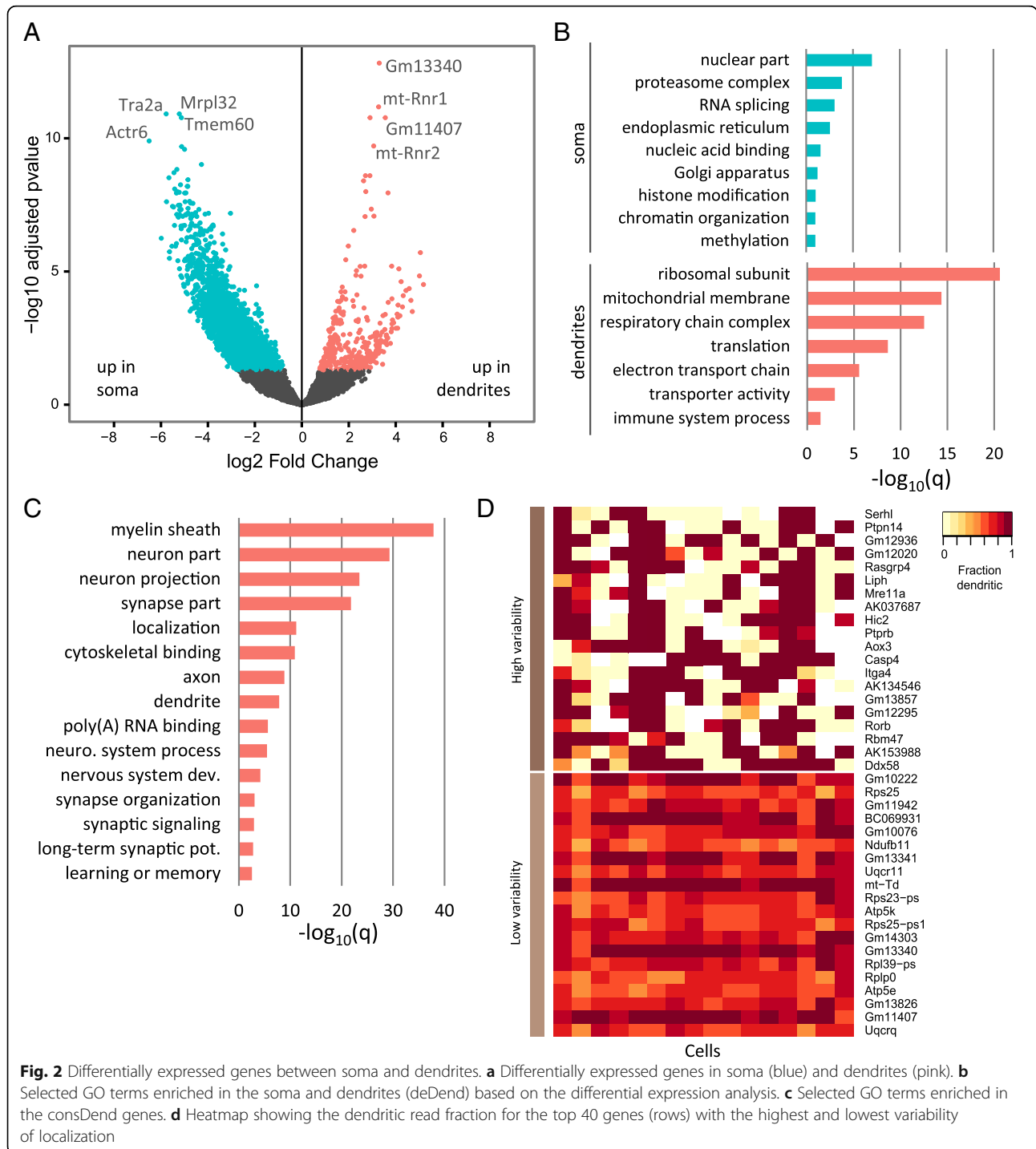


**Fig. 1** Sub-single cell profiling of soma and dendrite RNA. **a** Isolated single neurons were dissected to separate the soma and neurites, which were collected into separate tubes for amplification and RNA sequencing. **b** Overlap of expressed genes ( $\geq 10$  reads) between soma and dendrites from the same original cell. Each horizontal bar shows the results from a single neuron. The Venn diagram depicts the general relationship between the somatic and dendritic transcriptomes observed in the chart, where the dendritic transcriptomes were largely a subset of the somatic transcriptome of the same cell. **c** Marker gene expression for several brain cell types. Samples (columns) are indicated as either dendritic samples (pink) or soma samples (blue). Expression values were normalized by library size using DESeq2. Cardiomyocyte markers are included as a control cell type that is electrically active but unrelated to brain cells

genes significantly more highly expressed in somas and 387 genes significantly higher in dendrites (FDR corrected  $p \leq 0.05$ ) (Fig. 2a). Given their relatively higher expression in dendrites compared to soma, these 387 genes are likely to be actively localized, and we therefore refer to them as localized RNAs. The localized RNAs were strongly enriched for GO terms related to translation and mitochondria, consistent with previous reports [7, 8, 19],

whereas the somatic RNAs were enriched for functions related to the nucleus, including RNA splicing and chromatin organization (Fig. 2b and Additional file 1). Notably, there was no significant enrichment among these localized genes for terms specifically related to plasticity or synaptic function.

Differential expression analysis may not identify all localized RNAs because not all localized RNAs are



**Fig. 2** Differentially expressed genes between soma and dendrites. **a** Differentially expressed genes in soma (blue) and dendrites (pink). **b** Selected GO terms enriched in the soma and dendrites (deDend) based on the differential expression analysis. **c** Selected GO terms enriched in the consDend genes. **d** Heatmap showing the dendritic read fraction for the top 40 genes (rows) with the highest and lowest variability of localization

expected to have higher expression concentration in the dendrites than the soma. This may be particularly relevant when expression is profiled at the single-cell level, since factors such as bursting transcription and variable rates of localization can lead to high variability in the relative amounts of RNA in each compartment at the time of collection. Therefore, we additionally identified RNAs that were consistently present in the dendrites across the profiled cells, since these RNAs are likely to have important dendrite function even if they are not differentially at higher concentration in the dendrites compared to the soma. We found 1863 RNAs in at least 90% of the dendrite samples, which included well-characterized localized RNAs such as *Actb*, *Bdnf*, *Calm1*, *Dlg4*, *Grin1*, and *Map2*. To differentiate from the 387 differentially expressed genes described above, we refer to this set as the constitutive dendritic (consDend) RNAs, and the previous set as the differentially expressed dendritic (deDend) RNAs. The consDend RNAs covered many of the same ontology functions as the deDend RNAs, such as mitochondria and translation, but additionally were strongly enriched for a large number of synaptic and localization-related GO terms (Fig. 2c and Additional file 1). The consDend RNAs also contained a large number of genes with the GO annotation “myelin sheath,” which is unexpected given that this term is normally associated with axons. However, closer examination showed that this term includes genes with a wide variety of other functions (Additional file 1), and the consDend list does not contain myelin basic protein (*Mbp*). Overall, the differences between the deDend and consDend lists suggest that at the single-cell level, RNAs with important dendritic and synaptic functions are often not localized to the point of having higher expression concentration in the dendrites relative to the soma, but are nonetheless consistently present in the dendrites at a lower level.

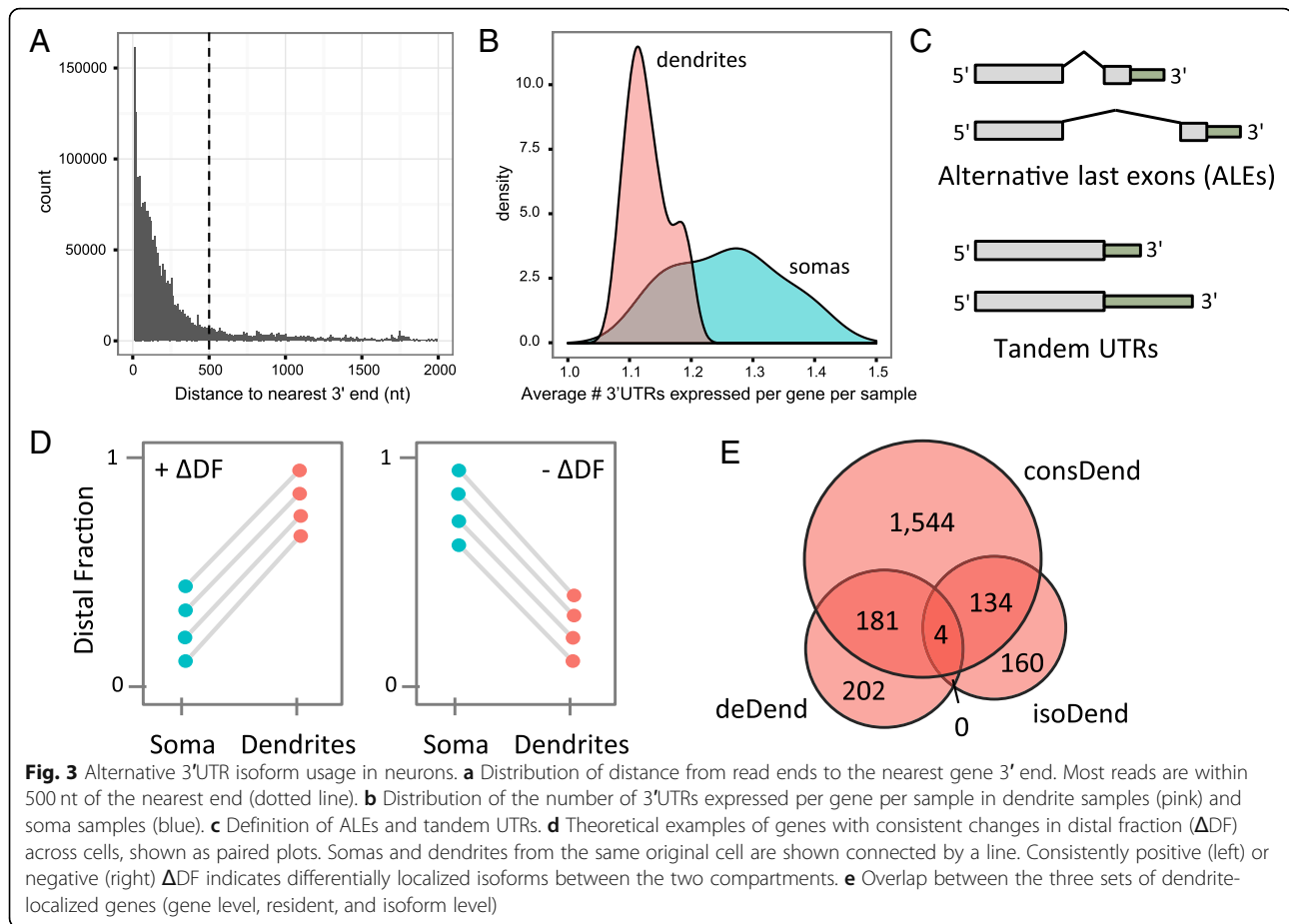
Single-cell analysis also allows us to examine the variability of localization across cells. For each of the 387 deDend RNAs, we calculated the variation of localization across cells based on the variance of the dendritic read fraction (defined as the number of dendritic reads divided by the sum of the dendritic and somatic reads for each cell). The top 40 genes with the highest and lowest localization variability are shown in Fig. 2d (mean variance 0.22 and 0.01 respectively). The high variability genes had lower median total-cell expression (dendritic + somatic reads) than the low variability genes (76.6 and 415.7 reads, respectively), and it should be noted that differences in expression level can potentially contribute to observed variability in single-cell experiments. To examine the effects of read sample size difference on the variability statistic, we created random subsets of soma and dendritic reads for the low variability genes by setting the total number of reads to 10 (minimum read threshold) and sampling

randomly from either compartment in proportion to the original frequencies. We then computed the down-sampled dendritic read fraction and its variation across cells. This procedure was repeated 1000 times to compute a non-parametric confidence interval. While the mean variance of the low variability genes increased 10-fold from the original value (mean variance  $0.1 \pm 1.9e-5$  from 1000 resamplings), it was still significantly lower than the high variability genes. From a biological perspective, low variability of localization suggests a gene is localized by a constitutive mechanism and is needed in constant supply in the dendrites, whereas high variability suggests more dynamic localization mechanisms which may be activated in response to stimuli. The genes with the highest variability of localization included several enzymes (*Serhl*, *Ptpn14*, *Liph*, *Mre11*, *Aox3*, *Casp4*, *Ddx58*), most of which do not currently have a defined dendritic function, although mutations in *Mre11* have been previously associated with Ataxia-telangiectasia-like disorder 1 [20]. These high variability genes also showed more “all-or-nothing” localization than the low variability genes, with most cells having a dendritic read fraction of close to either zero or one (Fig. 2d; see also Additional file 2 for subsampled version). Genes with the least variable localization included components of the ubiquinol-cytochrome c reductase complex (*Uqcrcq*, *Uqcrc11*), ATP synthase complex (*Atp5e*, *Atp5k*), and ribosomal subunits (*Rplp0*, *Rps25*), some of which in humans have been implicated in schizophrenia and schizoaffective disorder [21]. The presence of ribosomal subunits is somewhat perplexing given that ribosomes are assembled elsewhere. One speculative possibility is localized regulation of differential stoichiometry of ribosomal subunits [22, 23]. Overall, these results give further support to the idea that genes involved in respiration and translation are needed in constant supply in the dendrites, and suggest that this might be accomplished by a constitutive localization mechanism that is relatively constant across cells.

### Differential localization of 3'UTR isoforms

Given the potential importance of alternative 3'UTR usage in dendritic localization, we sought to better define genes that have 3'-isoform-specific dendritic localization in primary neurons. As a result of the aRNA single-cell RNA amplification process [15–17], the majority of our sequencing reads map within 500 nt of a 3' end (Fig. 3a), and we thus have high coverage of these regions for identifying expressed 3'UTR isoforms. We quantified the expression of individual 3' isoforms based on the last 500 nt of each isoform, merging any 3' ends that were closer than 500 nt into a single feature due to the potential ambiguity of quantification for closer ends (reducing this merge distance did not change the major conclusions we report here; see Additional file 3). Individual





cells widely expressed multiple 3' isoforms per gene, with somas showing slightly more alternative expression than dendrites on average (1.26 and 1.13 expressed 3'UTR isoforms per gene, respectively; Fig. 3b). When multiple isoforms were expressed, one isoform tended to be dominant, making up ~85% of the gene reads on average in both compartments. Next, we compared differential isoform representation between soma and dendrite. For simplicity, we limited the considered 3'UTR isoforms to only the top two most highly expressed isoforms per gene, which accounted for the vast majority of reads in most genes (82% of the genes expressing three or more 3'UTRs had at least 90% of their reads mapping to the top two UTRs). The top two isoforms were labeled "proximal" (the more 5' isoform) or "distal" (the more 3' isoform), and isoform preference for each gene in each sample was summarized as the fraction of reads mapping to the distal isoform (distal reads divided by distal plus proximal reads), which we refer to as the distal fraction (DF). We focused our analysis only on multi-3'UTR genes that had at least 10 total reads in both the soma and dendrites of at least five cells, which resulted in 3638 considered genes. We note that alternative 3'UTRs can be generated by two distinct mechanisms: alternative

splicing, which generates alternative last exons (ALEs), or alternative cleavage and polyadenylation, which generates tandem UTRs (Fig. 3c). Therefore, we split our set of multi-3'UTR genes into ALE and tandem groups based on the relationship between the designated proximal and distal 3'UTR for that gene. ALEs made up the majority of the considered multi-3'UTR genes (3108 ALE versus 530 tandem).

To identify 3'UTR isoforms that are differentially localized to dendrites, we looked for genes that had consistent patterns of isoform preference across our cells. That is, we looked for cases where the change in distal fraction ( $\Delta DF$ ; defined as  $DF_{\text{dendrite}} - DF_{\text{soma}}$  and calculated separately for each soma-dendrite pair) was in a consistent direction (+/-) across multiple cells (Fig. 3d). Using a Wilcoxon signed-rank test ( $p < 0.1$ ), we identified 298 genes that met this criterion. For clarity, we will refer to these 298 genes as isoform-specific dendritic (isoDend) RNAs. Most of the isoDend RNAs were categorized as ALEs (249 ALE, 49 tandem), but neither type was significantly enriched in this group compared to the full set of multi-3'UTR genes. Unlike the deDend and consDend sets, the isoDend RNAs were not significantly enriched for particular GO functional categories. Only four of the isoDend

RNAs overlapped with the deDend list (*mt-Rnr2*, *Rpl31*, *Rpl21*, and *Map2*), indicating that gene-level and isoform-level localized genes are distinct sets. In contrast, approximately half of each the deDend and isoDend sets overlapped with the consDend set (Fig. 3e).

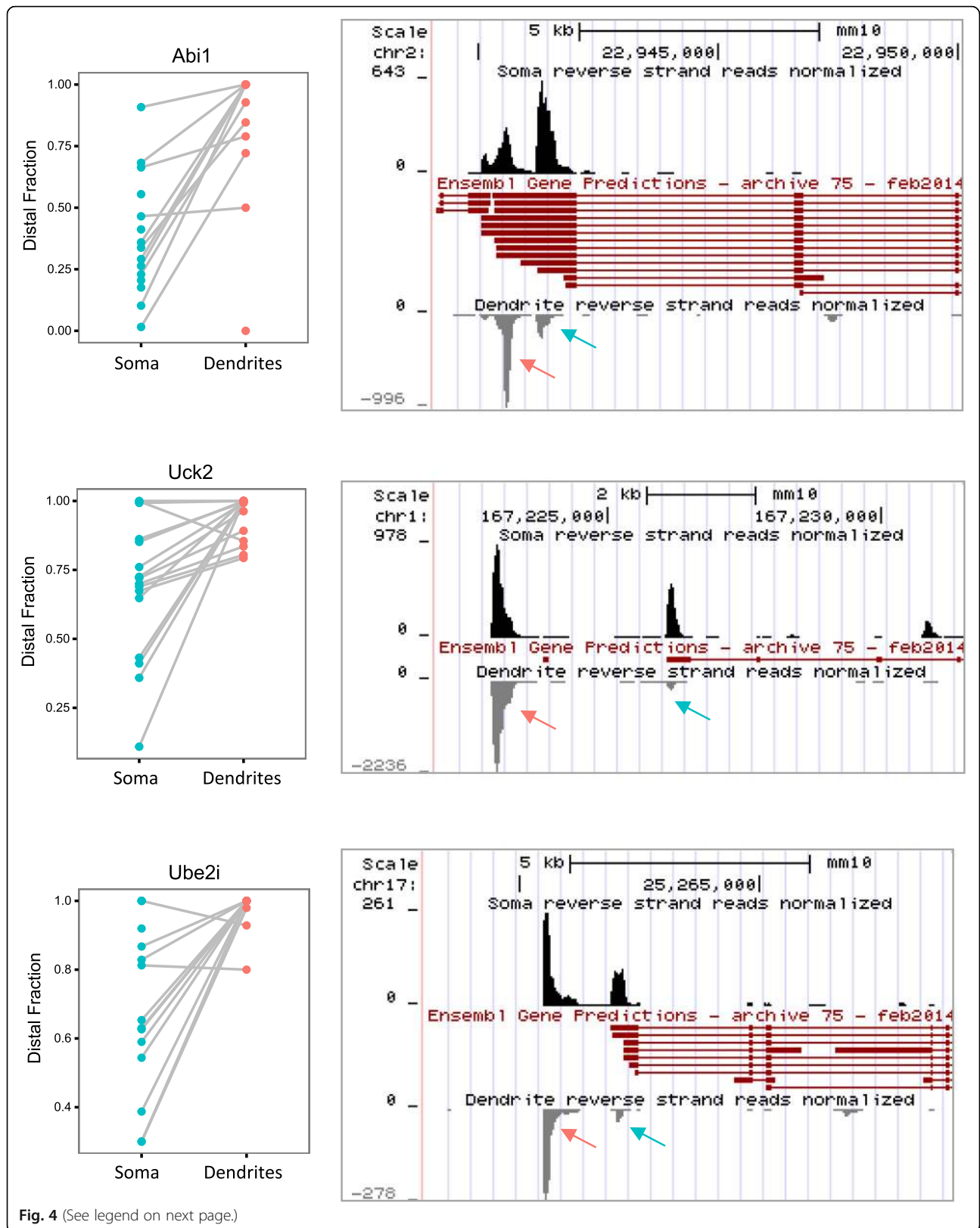
Among the 298 isoDend isoform pairs, we found that the dendrite-preferred isoforms were significantly longer than the soma-preferred isoforms for both ALE and tandem types ( $p < 0.01$ , paired  $t$ -test), which agrees with the findings of a recent study in rat hippocampal slices [14]. In addition, dendrites preferred the distal isoform in 64% of cases, which was independent of ALE/tandem status. This preference diverged significantly from expectation: in the full set of 3638 multi-3'UTR genes, dendrites preferred the distal isoform in only 44% of cases ( $p = 3.7e-13$ ; odds ratio = 2.4; Fisher's exact test). A preference for distal 3' isoforms in dendrites/neurites has also been observed in rat hippocampal slices [14] and brain-derived cell lines and cortical neurons [8]. Next, we examined the cell-to-cell variability of isoform preferences, particularly focusing on the differences in DF variability between somas and dendrites. For each gene, the variance of DF across samples was calculated separately for soma and dendrite samples. We found that 60.1% of the isoDend genes had a more variable DF in the soma than in the dendrites. Again, this observation diverged significantly from expectation based on the full set of multi-3'UTR genes, where only 29.4% of the genes had a more variable DF in the soma ( $p < 2.2e-16$ ; odds ratio = 3.6; Fisher's exact test). The median expression in the somas and dendrites differed (705 and 172 reads respectively). To examine the effect of expression levels, we randomly subsampled the soma reads down to the minimum threshold of 10 reads and recomputed the DF statistic and repeated the random downsampling 1000 times. We observed on average 61% of the isoDend genes with more variable DF in the soma than in the dendrites, consistent with the original analysis. Thus, dendrites showed more specific and consistent isoform preference among the isoDend genes compared to somas, potentially suggesting that certain isoforms are being selectively concentrated in the dendrites due to the presence of *cis* localization signals in the alternative portion of the 3'UTR. Figure 4 provides three representative examples of genes with these isoform patterns, showing the consistent preference for the distal isoform in the dendrites compared to soma for multiple individual cells, and the lower variability of DF in the dendrites compared to the somas. Finally, we looked to see how many of the dendrite-preferred isoforms were among the ~2000 new, distal 3'UTRs annotated recently by Miura et al. in several tissues [10]. Thirty-eight of the dendrite-preferred isoforms overlapped this list (including *Uck2* and *Ube2i* shown in Fig. 4), 12 of which were specific to hippocampal neurons in that study [10].

### Dendritic targeting motifs

We computationally analyzed the 3'UTRs of the deDend, isoDend (localized isoform only), and consDend gene lists to identify potential dendritic targeting elements (DTEs) enriched in each set compared to a length-matched non-localized background (see "Methods"). We first searched for instances of known RBP motifs. The greatest enrichment was seen for SRSF3-binding motif AUCA WCG, which was 2.4 times more common in the deDend RNAs than background and occurred in 59 of the 387 genes in this set. The same SRSF3 motif was also the most enriched motif in the consDend set (1.5 times more common than background) and occurred in 265 of the 1863 genes in this set. SRSF3 is a brain-expressed splicing factor, and although no specific role for this RBP in neurons has been described, it was recently shown in mouse P19 cells to promote 3'UTR lengthening through distal polyadenylation site usage and promote nuclear export through recruitment of NXF1 [24]. Therefore, one hypothesis could be that SRSF3 plays a role in the early steps of dendritic localization by promoting inclusion of alternative 3'UTRs (theoretically containing DTEs) and by facilitating nuclear export. We also performed a de novo motif analysis using HOMER [25] to see if any previously unidentified motifs were enriched in our sequences. The top motif in each set was UUCGAU ( $p = 0.0001$ , odds ratio = 2.9, hypergeometric test), CCGCAA ( $p = 1e-7$ , odds ratio 1.7), and GUGGGU ( $p = 0.01$ , odds ratio = 1.2) in the deDend, consDend, and isoDend sets, respectively. One motif, CGCR, was enriched in all three sets, but was only slightly more common in localizers than background (odds ratio < 1.2).

Since G-quadruplexes have been implicated previously in dendritic localization [26], we also searched our localized sequences for regions that could potentially form this structure. Using a regular expression (see "Methods"), we searched for potential G-quadruplexes in the 3'UTRs of each localized gene or isoform. G-quadruplexes were 2.0 times more common in the deDend RNAs ( $p = 0.003$ , Fisher's exact test), 1.9 times more common in the consDend RNAs ( $p = 5.0e-12$ , Fisher's exact test), and 1.7 times more common in the isoDend RNAs (not significant;  $p = 0.14$ , Fisher's exact test) than the non-localized background. Overall, 448 of the 2225 localized genes had at least one potential G-quadruplex in the localized 3'UTR. These results support a possible role for G-quadruplexes in localization in deDend and consDend RNAs, and possibly to a lesser extent in isoDend, but overall it does not appear that this motif alone is enough to explain the majority of localization.

To examine potential structural localization motifs more widely, we applied the de novo secondary structure motif-finding tool NoFold [27] to the localized 3'UTR sequences. Eighty-five motifs were significantly





(See figure on previous page.)

**Fig. 4** Examples of genes with significantly differentially localized 3' isoforms. Paired plots on the left show the DF for each soma-dendrite pair (connected by gray lines). The genome browser plots on the right show the read pile-ups for somas (top track; black peaks) compared to dendrites (bottom track; gray peaks; reversed orientation) relative to the annotated gene models from Ensembl (middle track; red). The dendrite-preferred 3' isoform is indicated by a pink arrow, and the non-preferred isoform is indicated by a blue arrow. Note that for Uck2 and Ube2i, the dendrite-preferred 3' isoform is a new isoform from [10] and thus is not part of the Ensembl gene models. Genes shown are on the reverse strand and only reverse-strand reads are displayed

enriched compared to non-localized background sequences ( $p < 0.01$ , Fisher's exact test). Two motifs in particular stood out as occurring in a large number of sequences (over 20 unique genes each). Though more conserved on the structure level, the instances of these motifs had enough sequence similarity to suggest a common origin. Using RepeatMasker [28], we identified these motifs as instances of the B1 and B2 SINE families, which are  $\sim 175$  nt retrotransposons that form long hairpin structures. To verify that these SINEs were enriched in the localized sequences, we created covariance models (CMs) for B1 and B2 using their canonical sequences and secondary structures and used these CMs to comprehensively identify structurally conserved matches to these elements in our sequences. Compared to non-localized background sequences, B1 structures were found 2.5 times more often in deDend RNAs ( $p = 0.00047$ , Fisher's exact test), 1.8 times more often in consDend RNAs ( $p = 7.6e-7$ , Fisher's exact test), and 1.9 times more often in isoDend RNAs (not significant;  $p = 0.33$ , Fisher's exact test), and B2 structures were found 2.5, 1.9, and 5.7 times more often in the deDend, consDend, and isoDend RNAs respectively ( $p < 0.001$ , Fisher's exact test). Overall, 255 and 165 localized genes out of the 2225 contained a B1 or B2 match, respectively. These results show that B1 and B2 SINE-related sequences are widespread and over-represented in localized RNAs, suggesting a possible role as DTEs analogous to the role of ID retrotransposon elements in rat dendritic localization [29]. Of note, only three genes contained both a G-quadruplex and a B1 or B2 motif, indicating that these signals likely operate on distinct sets of genes.

#### Functional analysis of the "local proteome" using structure information

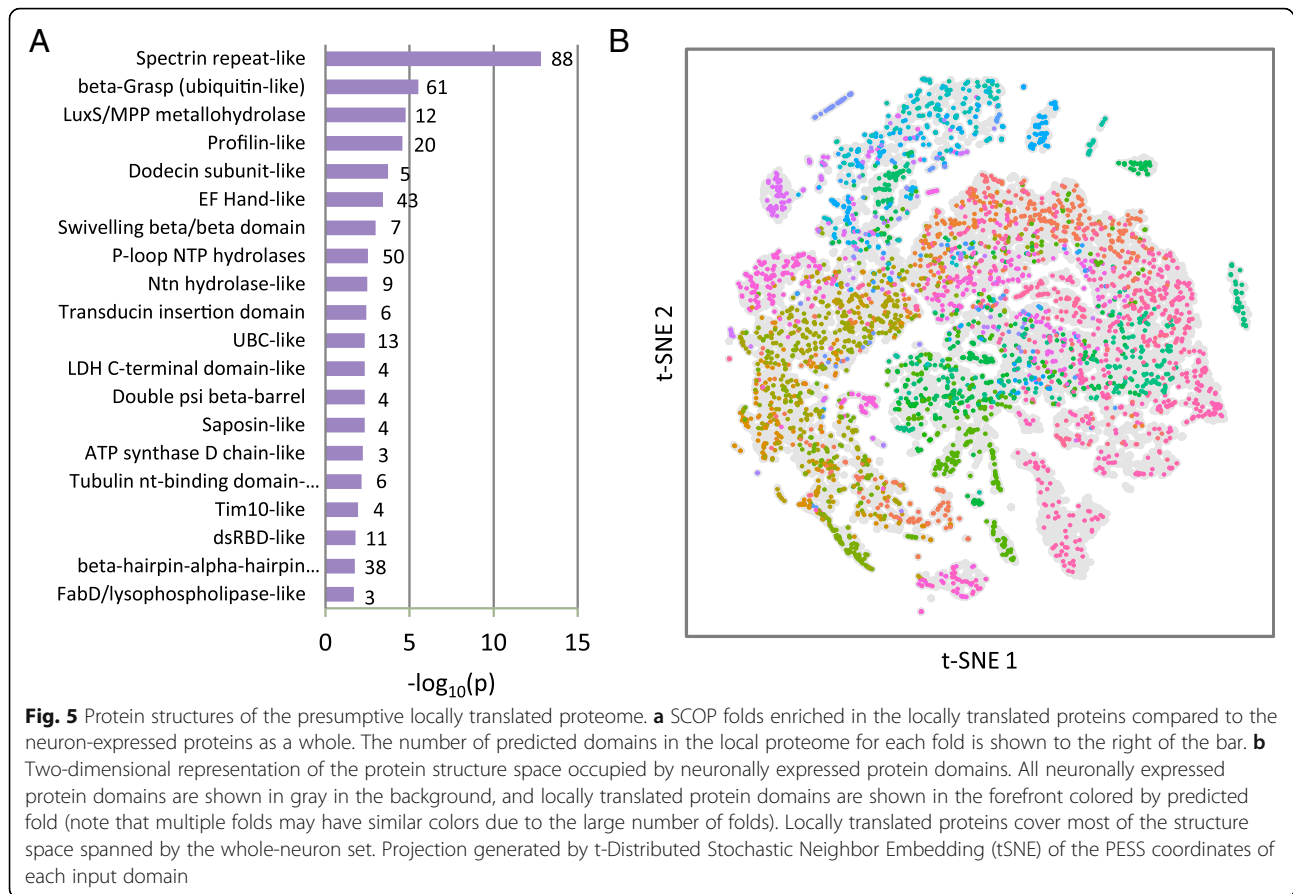
Only some of the dendritic RNAs might be involved in local protein translation. Nevertheless, to gain a better understanding of potential "local proteome," we performed a domain-level tertiary structure prediction on the protein products of 1930 localized mRNAs (combining the deDend, isoDend, and consDend lists and excluding non-coding RNAs). Full-length proteins were split into one or more predicted domains (where "domain" is defined as an amino acid chain that likely folds into a compact, independently stable tertiary structure;

see "Methods"), yielding a total of 6845 domains. Each domain was classified into a SCOP structural fold using our PESS pipeline [30]. Using this approach, we were able to predict the fold of 2005 additional domains beyond previous structural annotation [31]. Using the whole-neuron proteome as a background, we found that the local dendritic proteome was highly enriched for multiple different folds, including several related to cytoskeletal structure such as Spectrin repeats and actin-binding Profilin domains (Fig. 5a). Overall, 503 different folds were represented by at least one domain in the local dendritic proteome, covering almost the entire spectrum of folds expressed in the neuron as a whole (609 folds) (Fig. 5b). This suggests that rather than being highly specialized, the local dendritic RNA has the potential to encode for a diversity of protein functions on par with the whole cell.

To highlight some of the insight that can be gained through structure analysis, we selected several folds with important neuronal functions and assessed their representation within the locally translated set, which is described in Additional file 4: Tables S1-S3. A full catalog of predicted protein folds is provided in Additional file 5.

#### A master list of dendritic RNA

Towards creating a definitive list of dendritic RNAs that have been observed thus far in high-throughput studies, we obtained lists of dendritic genes from seven publications that profiled the dendritic transcriptome using microarray or RNA-seq [6–8, 14, 32–34] and combined those lists with our own. Of a total of 5827 unique genes on this list, only 1547 (27%) were observed in at least two studies, and none were found in all studies. The top 40 most frequently observed dendritic genes are listed in Table 1. Ribosomal proteins dominate the list, underscoring the importance of translation-related machinery in the dendrites. The most frequently observed genes were *Rps29*, *Ppp1r9b* (Neurabin-2, an actin-binding protein involved in synaptic transmission and dendritic spine morphology), and *Tpt1* (a calcium-binding protein involved in microtubule stabilization), which were each observed in six different studies. The full list of dendritic genes is available in Additional file 6: Table S4, and the full lists of deDend, consDend, and isoDend genes from this study can be found in Additional file 6: Table S5-S7.



## Discussion

Neurons have special RNA localization needs compared to other cell types: their unique morphology—long, extended processes that can be many times the length of the soma—combined with an extensive need for local translation means that neurons must transport a wide variety of RNAs long distances from their origination point in the nucleus. Here, we carried out single neuron sub-cellular RNA sequencing to more precisely identify a total of 2225 unique genes present in mouse dendrites, including 298 genes for which only a subset of the expressed transcripts were localized, depending on their 3'UTR isoform. Several of these differentially localized 3' UTR isoforms were among the set of recently identified distal 3'UTRs expressed in neurons [10]. Using de novo RNA structure motif analysis, we identified several secondary structures enriched in the 3'UTRs of the localized RNAs, including two hairpin structures derived from B1 and B2 SINE elements, which may act as localization signals. Finally, we applied a protein fold prediction algorithm to make structural and functional predictions for the set of proteins that are putatively translated locally at the synapse.

Based on our results, there are almost 300 genes with alternative 3' isoforms where one isoform was consistently

more dendritically localized than the other. The use of alternative 3'UTRs is an attractive model for how neurons might regulate localization, especially since 3'UTRs theoretically have the potential to provide an element of tissue-specificity to localization. In light of this, it is somewhat surprising that of the 38 dendrite-targeted isoforms we identified that were also profiled by [10], only 12 were specific to hippocampal neurons according to the Miura data. The other 26 isoforms were found in at least one of the other mouse tissue types profiling in that study, which included the spleen, liver, thymus, lung, and heart, suggesting a general lack of tissue specificity of these dendritically targeted isoforms. Instead, we postulate that tissue-specific localization may be achieved by tissue-restricted expression of *trans* factors (e.g., RBPs) rather than by regulation of DTE-containing isoform expression. In addition, although we observed significant enrichment of several candidate DTEs, including RBP recognition sites, G-quadruplexes, and SINE-mediated hairpin structures, none of the potential regulatory elements were universal or unique to localized RNA sequences. These results suggest that dendritic RNA localization involves multiple pathways and overlapping mechanisms [29, 35] and that “aggregate” localization signals composed of multiple DTEs may be necessary to improve specificity

**Table 1** Top 40 most frequently observed dendritic RNAs

	Gene	# Obs	Refs
1	Rps29	6	1,2,5,6,7,8
2	Ppp1r9b	6	1,2,3,6,7,8
3	Tpt1	6	1,2,4,5,7,8
4	Rpl37	5	1,2,5,7,8
5	Rpl17	5	1,2,5,6,8
6	Dlg4	5	1,2,3,6,8
7	Rpl4	5	1,2,4,7,8
8	Arl3	5	2,5,6,7,8
9	Map2	5	1,2,3,6,8
10	Rplp0	5	1,4,5,7,8
11	Psd	5	1,2,3,6,8
12	Rpl28	5	2,5,6,7,8
13	Rpl21	5	1,2,5,7,8
14	Arpc1b	5	1,4,5,7,8
15	Ftl1	5	1,2,4,5,8
16	Rps12	5	1,2,5,7,8
17	Eif3f	5	1,2,5,6,8
18	Uba52	5	1,2,5,7,8
19	Ube2m	5	1,2,5,6,8
20	Rpl32	4	1,5,7,8
21	Rpl31	4	1,2,5,8
22	Abhd17a	4	1,5,6,8
23	Fgf13	4	2,6,7,8
24	Rpl15	4	1,2,5,8
25	Rpl13	4	1,2,5,8
26	Rpl19	4	1,2,5,8
27	Ids	4	1,2,3,8
28	Serbp1	4	1,2,5,8
29	Sptbn2	4	1,2,6,8
30	Pacsin1	4	1,2,6,8
31	Hpcal4	4	1,2,3,6
32	Rtn2	4	1,5,6,8
33	Rnf165	4	1,2,6,8
34	Hint1	4	1,2,5,8
35	Eef2	4	1,2,4,8
36	Rpsa	4	1,2,5,8
37	Rps2	4	1,2,5,8
38	Rps8	4	2,5,7,8
39	Selenow	4	2,3,5,8
40	Camk2a	4	1,2,3,6

<sup>1</sup>Ainsley et al. 2014 [7]<sup>2</sup>Cajigas et al. 2012 [6]<sup>3</sup>Lein et al. 2007 [32]<sup>4</sup>Poon et al. 2006 [33]<sup>5</sup>Taliaferro et al. 2016 [8]<sup>6</sup>Tushev et al. 2018 [14]<sup>7</sup>Zhong et al. 2006 [34]<sup>8</sup>This study

and possibly also refine the destination of dendritically targeted transcripts.

An intriguing finding was that the composition of the deDend set was skewed towards RNAs that encode proteins that modulate RNA translation and mitochondrial function, as compared to the larger consDend set which covered many more dendrite- and synapse-specific functions. This leads us to speculate that translational regulation of dendritic protein synthesis might be dynamically modulated through stimulated transient local production of proteins that enhance the capacity to make ATP thereby facilitating translation. This would suggest a generalized but specific regulatory mechanism that could act on whatever RNAs are present at the site, without the need for individualized translation regulation of each dendritic RNA. Such a mechanism would allow the standard cellular translation mechanism to be specific without requiring the existence of new RNA transport proteins or transcript-specific translation. Regulation of local protein synthesis by the global mechanism of spatial translational control as opposed to individual RNA translational enhancement is different from current models of how dendritic protein synthesis is regulated, suggesting avenues for future experiments.

A crucial remaining question is what role individual locally translated proteins play in long-lasting synaptic potentiation. The post-synaptic density and surrounding dendritic spine are highly structured formations that depend on a scaffold of interacting proteins [36–38], which in turn usually require a specific three-dimensional fold in order to function properly. Here, we provide a fold-level structure-function annotation of 1930 proteins that we predict to be locally translated at the synapse based on our RNA localization analysis. Given that mutations linked to neuropsychiatric diseases have been found to be enriched in synaptic proteins in human and mouse, and several of these mutations appear to disrupt important structures [39, 40], structural knowledge of these proteins is important for understanding these disorders. A more complete picture of the structures of locally translated proteins will help both in functional understanding and mutation-impact analysis.

One limitation of our study is that neurons were only surveyed at the basal state, rather than after synaptic stimulation. Several studies have shown that RNA localization changes after stimulation [2, 41–43]; therefore, the set of dendrite RNAs identified here may still be only a subset of the RNAs needed for LTP. There also may be important differences between neurons in culture and in vivo that would be missed in our analysis. We observed significant overlap between our localized set and a set of localized RNAs derived partly from tissue-based studies conducted after fear conditioning [7], suggesting a reasonable amount of concordance between basal

primary cultures and post-stimulation tissue samples. Nonetheless, an important future direction will be to repeat the sub-cellular sequencing described here after stimulation. It will be particularly interesting to see if groups of RNAs that share a DTE undergo coordinated changes in localization post-activation, and conversely, if coordinated RNAs share any new DTEs.

## Conclusions

In sum, our study represents a comprehensive resource for RNA localization in mouse neurons consisting of our new sub-cellular RNA sequencing dataset, a compilation of previous dendritic RNA studies, as well as computational annotation of motifs and structures. The resource generated here may have broad utility for continued study of mechanisms of dendritic RNA localization and the role of localized RNA in neuronal function and dysfunction.

## Methods

To approach this project, we cultured neurons from embryonic mice and manually dissected dendrites and soma, individually from each cell, collecting the material from each compartment separately. These sub-cellular fractions from single cells were amplified and sequenced. We used within-cell differential expression analysis as well as between-cell consistency analysis to identify localized RNA and possible isoform variants that differentially localize. We then used computational analyses to identify possible structural motifs mediating the localization and the proteomic functions of the localized RNA. We collated our data with existing studies to create a resource for the community.

### Neuron culture and collection

Hippocampal neurons from embryonic day 18 (E18) mice (C57BL/6) were cultured as described in [44] for 15 days. Isolated single neurons were selected for collection. A micropipette with a closed, tapered end was used to sever dendrites from the cell body. Another micropipette was used to aspirate the soma, which was deposited into a tube containing a first-strand synthesis buffer and RNase inhibitor and placed on ice. A separate micropipette was used to aspirate the dendrites, which were deposited into a separate tube as above. Samples were transferred to  $-80^{\circ}\text{C}$  within 30 min and stored there until first-strand synthesis. Sixteen neurons (32 total samples) were collected from multiple cultures across multiple days.

### Single-cell RNA amplification and sequencing

ERCC spike-in control RNA was diluted 1:4,000,000 and 0.9  $\mu\text{L}$  was added to each tube. Poly-adenylated RNA was amplified using two or three rounds of the aRNA in vitro transcription-based amplification method, as described in [15]. The quality and quantity of the amplified RNA was verified using a Bioanalyzer RNA assay. Strand-specific

sequencing libraries were prepared using the Illumina TruSeq Stranded kit according to the manufacturer's instructions, except that the initial poly-A capture step was skipped because the aRNA amplification procedure already selects for poly-adenylated RNA. Samples were sequenced on a HiSeq (100 bp paired-end) or NextSeq (75 bp paired-end) to an average depth of 25 million reads. Reads were trimmed for adapter and poly-A sequence using in-house software and then mapped to the mouse genome (mm10) using STAR [45]. Uniquely mapped reads were used for feature quantification using VERSE [46]. The features used for each analysis are described below.

### Gene-level expression and localization

Three sources of gene annotations were combined to obtain a comprehensive definition of known 3' ends: Ensembl genes (downloaded from UCSC, Dec. 2015), UCSC genes (downloaded from UCSC, Dec. 2015), and the set of  $\sim 2000$  new 3'UTRs determined by Miura *et al.* [10]. The 3'UTR regions of these annotations were used for quantification of reads. A single 3'UTR feature was created for each gene by taking the union of all 3'UTR regions for that gene. Read counts were calculated for each gene based on how many reads mapped to this 3'UTR region. Quantification was done using VERSE with options `"-s 1 -z 3 --nonemptyModified"`. For differential expression analysis, we used only the genes that had at least one read in at least half (16) of the samples. Read counts were normalized for library size using the size factor method of DESeq2 and differentially expressed genes between the dendrites and soma were identified using DESeq2 with a paired experimental design. A FDR corrected  $p \leq 0.05$  was used to identify significantly differentially expressed genes. The consDend genes were identified separately based on having at least one read in at least 90% (i.e., 15 out of 16) of the dendrite samples.

GO functional enrichment of deDend and consDend genes was calculated using the GOrilla webserver [47]. For deDend genes, the background set for GO analysis was all genes with at least one read in half the samples; for the consDend genes, the background was all genes with at least one read in at least 15 samples (i.e., the input sets for each analysis).

Gene markers of pyramidal neurons and cardiomyocytes, as well as housekeeping genes, were obtained from [9]. Markers of other mouse brain cell types were obtained from [48].

### Isoform-level expression and localization

An overview of these methods is shown in Additional file 7. To quantify individual 3' isoforms of genes, we used the last 500 nt of each 3' end for that gene as the isoform quantification feature. This was done to normalize length differences between 3'UTRs and because the



vast majority of reads were mapped within 500 nt of a 3' end (Fig. 3a). Any 3' ends that were less than 500 nt apart were merged together into a single quantification feature. Thus, the final set of 3' isoform quantification features is non-overlapping. Isoform read counts were calculated by VERSE using the same parameters as above. Genes with only one expressed 3' isoform were removed from further analysis to focus on alternative expression of 3' isoforms.

To identify the top two 3' isoforms for each gene, the following procedure was used (Additional file 7). For each gene in each sample, the fraction of reads mapping to each isoform was calculated (that is, the number of reads mapping to that isoform divided by the total reads for all isoforms of the gene). The fractions for each isoform were then summed up across samples (unless a sample had fewer than 10 reads total for that gene, in which case it was skipped), and the two isoforms with the highest total per gene were considered the top two isoforms for that gene. The purpose of this process was to give each sample equal weight in the final decision of the top 3'UTR, while also excluding samples with too few reads to give a reliable estimate of the isoform fractions. This process was repeated for each gene with at least two expressed isoforms in the dataset. Then for each gene, whichever of the top two isoforms was more 5' (as defined by the locations of their 500-nt quantification features) was designated the "proximal" isoform and whichever was more 3' was designated the "distal" isoform. Finally, for each gene in each sample, we calculated the distal fraction (DF) as the fraction of reads mapping to the distal isoform divided by the total reads mapping to the distal and proximal isoforms.

We defined the proximal and distal isoforms as being, relative to each other, generated by alternative splicing (ALEs) or alternative cleavage and polyadenylation (tandem UTRs) by the following criterion: if the full-length 3'UTRs of a pair of isoforms were directly adjacent or overlapping, they were called tandem; otherwise, they were called ALEs.

The differential localization of isoforms was determined based on the change in distal fraction between soma and dendrites of the same original neuron. A non-parametric paired test of differences (Wilcoxon signed-rank test) was used to identify genes with consistent changes in distal fraction across samples. Only genes with at least five pairs of samples (where a "pair" means the soma and dendrites from the same original neuron) where each member of the pair had at least 10 combined reads for the two isoforms were tested (3638 genes), to ensure there was enough read and sample support to reliably identify these events.

GO enrichment was done on the dendrite-enriched isoforms as described in the previous section, using the input set of 3638 genes as background.

### Background datasets for motif enrichment

We generated a pool of "non-localized" background 3' UTR sequences based on the list of genes that were significantly higher expressed in the soma from the gene-level DESeq2 analysis described above (3811 genes). We filtered this set to remove any overlap with one of the other localized lists (i.e., the consDend list and the isoDend list) and any overlap with previously annotated dendritically localized genes in order to make this list as specific to non-localized genes as possible, which resulted in removal of 471 and 531 genes respectively leading to a final pool of 2809 genes from which to draw 3'UTR sequences to make up a background. Since motif frequency in a sequence can be related to sequence length, we created a length-matched background set for each of the three localized gene lists as follows: (1) for each localized gene in the set, scan the pool of non-localized genes in order of their somatic specificity (starting with the most soma-specific, as indicated by its DESeq2 test statistic); (2) select the first non-localized gene encountered with a 3' UTR length within 100 nt of the localized gene's 3' UTR length; (3) add the selected non-localized gene to the background set and remove it from the pool; (4) if no background gene can be found that meets the 100-nt criteria, select whichever gene in the pool that has the most similar 3'UTR length to the localized gene's 3'UTR. Using this protocol resulted in background sets with highly similar length characteristics to the foreground set.

### RNA motif analysis

Linear motifs were identified using the HOMER motif-finding suite [25]. De novo-enriched motif searches were done using the script "findMotifs.pl" and set to look for either short motifs (4 or 6 nt) or long motifs (8, 10, or 12 nt). Enrichment of known RBP-binding motifs was analyzed using the same script with option "-known" in combination with a custom set of positional weight matrices specifying binding preferences that was downloaded from CISBP-RNA (version 0.6) [49]. A log-odds threshold for RBP motif matching was set for each motif separately based on the number of informative positions in the motif such that longer, more specific motifs had a higher log-odds threshold for calling a match. The background sets used for enrichment testing were the length-matched non-localized sets described above.

G-quadruplexes were identified by regular expression search using the "re" module in Python. The search pattern was '([gG]{3,}\w{1,7}){3,}[gG]{3,}', which requires three consecutive matches to the pattern "three or more G's followed by 1–7 of any nucleotide" and then ending with a fourth set of three or more G's. The background set was the same as described in the previous section.



De novo identification of enriched RNA secondary structures was performed using NoFold [27]. Sliding windows of 100 nt (slide = 75 nt) across the localized sequences were used for input. Background datasets were the same as described in the previous section and also converted to sliding windows with the same parameters. Additional matches to the B1 and B2 elements were found by creating a CM for each element based on its canonical sequence(s) downloaded from RepeatMasker [28] and its predicted MFE structure from RNAfold [50]. The sequences and structures used to create the CM are as follows:

B1 sequence:

```
GAGGCAGGCGGATTTCTGAGTTCGAGGCC
AGCCTGGTCTACAGAGTGAGTTCAGGACA
GCCAGGGCTACACAGAGAAACCCTGTCTC
```

B1 structure:

```
(((((.....(((((((((((((.....((.....)))))).....)))))).....)))))).....))))))
```

B2 sequence:

```
GCTGGTGAGATGGCTCAGTGGGTAAGAGC
ACCCGACTGCTCTCCGAAGGTCAGGAGTT
CAAATCCCAGC
```

B2 structure:

```
(((((.....(((((((((((((.....((.....)))))).....)))))).....)))))).....))))))
```

Bitscore cutoffs for high-quality matches were set to 50 for B1 and 35 for B2 based on the length of the model. Enrichment was computed using Fisher's exact test based on the number of high-quality matches in the localized set compared to the non-localized background (same background as above). Only one match was counted per gene for the purposes of enrichment testing.

### Protein structure analysis

For each predicted dendritic RNA, we obtained the canonical protein sequence, if any, from UniProt [51]. The canonical isoform is defined by UniProt to usually be the one that is most inclusive of exons/domains. We refer to this protein set as the "local proteome". We also obtained the canonical protein sequences for the full set of expressed genes in soma and dendrite samples (at least 1 read in at least 15 samples) to use as a background for comparison with the local proteome.

Each protein was split into domains based on Domain-Finder Gene3D predictions [31, 52]. If there were regions between, before, or after predicted domains that

were longer than 30 amino acids (aa) but did not have a Gene3D prediction, we also included these. If a "filled in" region such as this was longer than 450 aa, we used a sliding window of 300 aa (slide = 150 aa) to break it into smaller pieces, since domains are rarely larger than this. The fold of each domain was predicted using the method described in [30]. A nearest neighbor distance threshold of  $\leq 17.5$  was used to designate "high confidence" predictions, and a more lenient threshold of  $\leq 30$  was used to designate "medium confidence" predictions.

### Additional files

**Additional file 1:** GO annotations. Full list of enriched GO terms for differentially expressed (deDend and deSoma) and consistent dendrite (consDend) lists for biological processes (BP), molecular functions (MF), and cellular components (CC). Related to Fig. 2b and c. (XLSX 1129 kb)

**Additional file 2:** Heatmap of subsampled variability of localization, related to Fig. 2d. The low variability genes were subsampled to 10 reads, the high variability genes are displayed as their original values. (PDF 222 kb)

**Additional file 3:** Re-analysis of differential localization of 3'UTR isoforms using 250 nt merge distance. (PDF 78 kb)

**Additional file 4:** Expanded structure analysis of potential locally translated proteins. **Table S1.** Predicted transmembrane structures. **Table S2.** Predicted RNA-binding structures. **Table S3.** Predicted structures commonly found in synaptic proteins. (PDF 189 kb)

**Additional file 5:** Full list of predicted protein structural folds for dendritic genes found in this study. (XLSX 506 kb)

**Additional file 6:** Catalog of dendritic genes. **Table S4.** Full list of dendritic genes from current study and seven previous publications.

**Table S5.** Full deDend gene list. **Table S6.** Full consDend gene list.

**Table S7.** Full isoDend gene list. (XLSX 346 kb)

**Additional file 7:** Overview of 3'UTR definition, quantification, selection of top two isoforms, and calculation of distal fraction. (PDF 275 kb)

### Acknowledgements

Not applicable.

### Funding

This work was funded in part by NIMH U01MH098953 to JK and JE, NIGMS R01 GM110005 to JE and JK, and Health Research Formula Fund from the Pennsylvania Commonwealth to JK. SAM was supported by a DOE CSGF fellowship (DE-FG02-97ER25308). The funding agencies played no direct role in design, analyses, and conclusions presented in this work.

### Availability of data and materials

Annotated data is included as additional files. The datasets supporting the conclusions of this article are available in the GEO and SRA repositories under accessions GSE115480 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115480>] and SRP150011 [<https://www.ncbi.nlm.nih.gov/sra/?term=SRP150011>].

### Authors' contributions

SAM carried out all laboratory experiments and primary data analysis and drafted the manuscript. JK advised and helped design the experiments, designed the analysis, and refined the manuscript. JE advised and helped design the experiments and refined the manuscript. All authors read and approved the final manuscript.

### Ethics approval

All animal protocols were executed under by-products protocol, which is exempt under the guidelines of IACUC of University of Pennsylvania.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Graduate Program in Genomics and Computational Biology, Biomedical Graduate Studies, University of Pennsylvania, 160 BRB II/III - 421 Curie Blvd, Philadelphia, PA 19104-6064, USA. <sup>2</sup>Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, 829 BRB II/III, 421 Curie Blvd, Philadelphia, PA 19104, USA. <sup>3</sup>Department of Biology, University of Pennsylvania, 415 S. University Ave, Philadelphia, PA 19104, USA. <sup>4</sup>Present Address: Computational Biology, Target Sciences, GlaxoSmithKline R&D, 1250 S. Collegeville Road, Collegeville, PA 19426, USA.

Received: 18 October 2018 Accepted: 16 January 2019

Published online: 24 January 2019

**References**

- Aakalu G, Smith WB, Nguyen N, Jiang C, Schuman EM. Dynamic visualization of local protein synthesis in hippocampal neurons. *Neuron*. 2001;30:489–502.
- Eberwine J, Miyashiro K, Kacharmina JE, Job C. Local translation of classes of mRNAs that are targeted to neuronal dendrites. *Proc Natl Acad Sci*. 2001;98:7080–5.
- Job C, Eberwine J. Identification of sites for exponential translation in living dendrites. *Proc Natl Acad Sci*. 2001;98:13037–42.
- Miyashiro K, Dichter M, Eberwine J. On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. *Proc Natl Acad Sci U S A*. 1994;91:10800–4.
- Crino PB, Eberwine J. Molecular characterization of the dendritic growth cone: regulated mRNA transport and local protein synthesis. *Neuron*. 1996;17:1173–87.
- Cajigas IJ, Tushev G, Will TJ, tom Dieck S, Fuerst N, Schuman EM. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron*. 2012;74:453–66.
- Ainsley JA, Drane L, Jacobs J, Kittelberger KA, Reijmers LG. Functionally diverse dendritic mRNAs rapidly associate with ribosomes following a novel experience. *Nat Commun*. 2014;5:4510.
- Taliaferro JM, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, et al. Distal alternative last exons localize mRNAs to neural projections. *Mol Cell*. 2016;61:821–33.
- Dueck H, Khaladkar M, Kim T, Spaethling J, Francis C, Suresh S, et al. Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biol*. 2015;16:122.
- Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res*. 2013;23:812–25.
- Miura P, Sanfilippo P, Shenker S, Lai EC. Alternative polyadenylation in the nervous system: to what lengths will 3' UTR extensions take us? *Bioessays*. 2014;36:766–77.
- An JJ, Gharami K, Liao GY, Woo NH, Lau AG, Vanevski F, et al. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell*. 2008;134:175–87.
- Liao G-Y, An JJ, Gharami K, Waterhouse EG, Vanevski F, Jones KR, et al. Dendritically targeted Bdnf mRNA is essential for energy balance and response to leptin. *Nat Med*. 2012;18:564–71.
- Tushev G, Glock C, Heumüller M, Biever A, Jovanovic M, Schuman EM. Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron*. 2018;98:495–511.
- Morris J, Singh JM, Eberwine JH. Transcriptome analysis of single cells. *J Vis Exp*. 2011;50:e2634.
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A*. 1990;87:1663–7.
- Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, et al. Analysis of gene expression in single live neurons. *Proc Natl Acad Sci U S A*. 1992;89:3010–4.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Francis C, Natarajan S, Lee MT, Khaladkar M, Buckley PT, Sul J-Y, et al. Divergence of RNA localization between rat and mouse neurons reveals the potential for rapid brain evolution. *BMC Genomics*. 2014;15:883.
- Stewart G, Maser R, Stankovic T, Bressan D, Kaplan M, Jaspers N, et al. The DNA double strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia like disorder. *Cell*. 1999;99:577–87.
- Arion D, Corradi JP, Tang S, Datta D, Boothe F, He A, et al. Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder. *Mol Psychiatry*. 2015;20:1397–405.
- Slavov N, Semrau S, Airoldi E, Budnik B, van Oudenaarden A. Differential stoichiometry among core ribosomal proteins. *Cell Rep*. 2015;13:865–73.
- Shi Z, Fujii K, Kovary KM, Genuth NR, Röst HL, Teruel MN, et al. Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol Cell*. 2017;67:71–83.
- Müller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, et al. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev*. 2016;30:553–66.
- Brenner C. HOMER: Software for motif discovery and next generation sequencing analysis. 2010; Available from: <http://homer.ucsd.edu>
- Subramanian M, Rage F, Tabet R, Flatter E, Mandel J-L, Moine H. G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep*. 2011;12:697–704.
- Middleton SA, Kim J. NoFold: RNA structure clustering without folding or alignment. *RNA*. 2014;20:1671–83.
- Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013; Available from: <http://www.repeatmasker.org>
- Buckley PT, Lee MT, Sul J-Y, Miyashiro KY, Bell TJ, Fisher SA, et al. Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. *Neuron*. 2011;69:877–84.
- Middleton SA, Illuminati J, Kim J. Complete fold annotation of the human proteome using a novel structural feature space. *Sci Rep*. 2017;7:1–10.
- Lees J, Yeats C, Perkins J, Sillitoe I, Rentzsch R, Dessailly BH, et al. Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis. *Nucleic Acids Res*. 2012;40:465–71.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*. 2007;445:168–76.
- Poon MM, Choi S-H, Jamieson CAM, Geschwind DH, Martin KC. Identification of process-localized mRNAs from cultured rodent hippocampal neurons. *J Neurosci*. 2006;26:13390–9.
- Zhong J, Zhang T, Bloch LM. Dendritic mRNAs encode diversified functionalities in hippocampal pyramidal neurons. *BMC Neurosci*. 2006;7:17.
- Holt CE, Schuman EM. The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron*. 2013;80:648–57.
- Kim E, Sheng M. PDZ domain proteins of synapses. *Nat Rev Neurosci*. 2004;5:771–81.
- Dalva MB, McClelland AC, Kayser MS. Cell adhesion molecules: signalling functions at the synapse. *Nat Rev Neurosci*. 2007;8:206–20.
- Zheng C-Y, Seabold GK, Horak M, Petralia RS. MAGUKs, synaptic development, and synaptic plasticity. *Neurosci*. 2011;17:493–512.
- Liu-Yesucevitz L, Bassell GJ, Gitler AD, Hart AC, Klann E, Richter JD, et al. Local RNA translation at the synapse and in disease. *J Neurosci*. 2011;31:16086–93.
- Grant SG. Synaptopathies: diseases of the synaptome. *Curr Opin Neurobiol*. 2012;22:522–9.
- Tongiorgi E, Righi M, Cattaneo A. Activity-dependent dendritic targeting of BDNF and TrkB mRNAs in hippocampal neurons. *J Neurosci*. 1997;17:9492–505.
- Steward O, Wallace CS, Lyford GL, Worley PF. Synaptic activation causes the mRNA for the IEG Arc to localize selectively near activated postsynaptic sites on dendrites. *Neuron*. 1998;21:741–51.
- Yoon YJ, Wu B, Buxbaum AR, Das S, Tsai A, English BP, et al. Glutamate-induced RNA localization and translation in neurons. *Proc Natl Acad Sci*. 2016;113:E6877–86.

44. Buchhalter JR, Dichter MA. Electrophysiological comparison of pyramidal and stellate nonpyramidal neurons in dissociated cell culture of rat hippocampus. *Brain Res Bull.* 1991;26:333–8.
45. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
46. Zhu Q, Fisher SA, Shallcross J, Kim J. VERSE: a versatile and efficient RNA-Seq read counting tool. *bioRxiv.* 2016. <https://doi.org/10.1101/053306>.
47. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 2009;10:48.
48. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keefe S, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci.* 2014;34:11929–47.
49. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature.* 2013;499:172–7.
50. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. The Vienna RNA Websuite. *Nucleic Acids Res.* 2008;36:W70–4.
51. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2017;45:D158–69.
52. Yeats C, Redfern OC, Orengo C. A fast and automated solution for accurately resolving protein domain architectures. *Bioinformatics.* 2010;26:745–51.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

