Published in final edited form as:

Nat Struct Mol Biol. 2014 September ; 21(9): 833–839. doi:10.1038/nsmb.2876.

Asymmetric mRNA localization contributes to fidelity and sensitivity of spatially localized systems

Robert J. Weatheritt^{1,*}, Toby J Gibson², and M. Madan Babu^{1,*}

¹MRC Laboratory of Molecular Biology, Cambridge, United Kingdom

²Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany.

Abstract

While many proteins are localized after translation, asymmetric-protein distribution is also achieved by translation after mRNA localization. Why are certain mRNA transported to a distal location and translated on-site? Here we undertake a systematic, genome-scale study of asymmetrically distributed protein and mRNA in mammalian cells. Our findings suggest that asymmetric-protein distribution by mRNA localization enhances interaction fidelity and signaling sensitivity. Proteins synthesized at distal locations frequently contain intrinsically disordered segments. These regions are generally rich in assembly-promoting modules and are often regulated by post-translational modification sites. Such proteins are tightly regulated but display distinct temporal dynamics upon stimulation with growth factors. Thus proteins synthesized on-site may rapidly alter proteome composition and act as dynamically regulated scaffolds to promote the formation of reversible cellular assemblies. Our observations are consistent across multiple mammalian species, cell types, and developmental stages suggesting that localized translation is a recurring feature of cell signaling and regulation.

Spatial localization of cellular components is crucial for functional specialization and versatility. This organization allows biomolecules to come together when required to control regulatory processes such as signal transduction, asymmetric cell division and changes to cell morphology^{1,2}. Indeed, the mislocalization of a several proteins has been documented to have dramatic effects on development³ and cell morphology⁴, and has also been linked to a number of neurodevelopmental and neurodegenerative diseases⁴⁻⁶. It is well established that asymmetric localization of proteins can be achieved by transporting proteins after mRNA translation^{7,8}. In recent years, it has also become increasingly clear that mRNA localization coupled to protein synthesis at a distal site is another prevalent mechanism to asymmetrically localize proteins^{6,9}. The presence of these two mechanisms, co-existing in the same cell type, raises the question of why certain proteins are translated at their site of action in distal locations (Fig. 1a).

*rweather@mrc-lmb.cam.ac.uk and madanm@mrc-lmb.cam.ac.uk.

Author Contributions: RJW collected the datasets and performed the analysis; RJW and MMB designed the study; RJW, TJG and MMB conceived the idea and wrote the paper.

Individual studies have shown that localization of mRNAs is widespread, evolutionarily conserved^{5,10} and functionally important^{3,4,11,12}. However, it is unclear whether the proteins localized by the subcellular targeting of their mRNAs differ in their properties from those localized by protein transport. We therefore set out to answer the question: are there differences between proteins that are transported after translation and those that are translated after mRNA localization? In this study, we systematically analyzed genome-scale data on asymmetric localization of proteins⁷ and transcripts¹³ (Fig. 1b and Fig. 1c) within the neurites of N1E-115 mouse neuroblastoma cells. Both studies, though undertaken in different labs, employ the same experimental set-up using microporous filters^{7,13}. This provides a framework for the isolation and analysis of asymmetrically localized proteins or transcripts in the neurites compared to the cell body or soma. The data from these independent studies allowed us to directly compare the characteristics of gene products enriched in distal regions of neurites compared to the cell body. In particular, we compared the properties of distal site synthesis (DSS) proteins⁷ with the transport after synthesis (TAS) proteins¹³ (see Supplementary Fig. 1 and Fig. 1d). We defined the DSS group of proteins as the proteins whose mRNAs were detected as asymmetrically distributed in the neurites in microarray studies (see Supplementary note). The TAS group of proteins is defined as those proteins that were detected as asymmetrically distributed in proteomic studies of the neurites, and whose transcripts were not identified in the DSS group (Online Methods and Fig. 1b).

An initial functional and phenotype analysis of the two groups of proteins revealed that distal site synthesis proteins are likely to be functionally distinct from proteins that are transported after synthesis (Supplementary note). This observation prompted us to carry out a more comprehensive analysis. Our genome- and proteome-scale analyses that integrate multiple large-scale datasets (Fig. 1c) revealed that distal site synthesis proteins often contain intrinsically disordered regions (IDRs) and play a central role in promoting reversible multivalent protein complex assembly, which may provide an additional layer of regulation and spatial organization to signaling networks. In particular, our findings suggest that localized translation of asymmetrically distributed transcripts can locally change protein abundance rapidly, indicating a partial decoupling of the response of signaling networks from transcriptional regulation. Notably, our observations are conserved across tissue types, developmental stages and organisms and highlight the role of mRNA localization and localized translation in cell regulation.

Results

IDRs and interaction promiscuity are features of neurite DSS proteins

An analysis of large-scale protein interaction networks revealed that distal site synthesis proteins tend to have a larger number of interaction partners than the transport after synthesis proteins (Fig. 2a and Online Methods) though we found no preference for either group to be associated with large and stable protein complexes ($P < 0.30$, Wilcoxon-test) (Supplementary note). In fact, we observed that different members of the same protein complex are seen in the DSS and TAS protein groups (see Discussion and Supplementary note). Further analysis also revealed a significant difference in the structural properties

between the two groups (Fig. 2a). While the transport after synthesis proteins are enriched in structured globular domains, the distal site synthesis proteins have a significant enrichment in intrinsically disordered regions, which are polypeptide segments that lack stable tertiary structure¹⁴ (Fig 2a). The enrichment of IDRs in distal site synthesis proteins likely instills several advantages including conformational flexibility and a greater surface area for biomolecular interactions that may facilitate interactions with diverse partners^{14,15}, as demonstrated by Shank1, a member of the postsynaptic density complex, which has extensive regions of intrinsic disorder and whose localization to axons is controlled by mRNA localization¹⁶.

Binding motifs within DSS proteins may aid scaffolding

A key functional module enriched within IDRs is the linear motif, which often consists of 3-5 residues that are essential for mediating physical interactions¹⁷. However, due to the limited number of amino acids involved in binding, the interaction affinity is often weak¹⁷. Using the Anchor program¹⁸, we observed that putative linear motifs are significantly enriched within IDRs in the distal site synthesis proteins (Fig. 3a). Regions that are compositionally biased (i.e. repeating amino acids), multiple occurrence of linear motifs and/or their binding domains are a key feature of many signaling proteins¹⁹. Such segments have recently been shown to mediate molecular phase transitions from small, soluble entities to large macromolecular assemblies²⁰. Such assemblies manifest as granules (also referred as cellular bodies), concentrating protein components in spatially restricted regions, thereby increasing the rate of bio-molecular interactions^{14,21,22}. Compared to the TAS group of proteins, we observed that the DSS proteins show a significant enrichment in repeating proline-rich linear motif binding sites and other phase transition promoting low complexity polypeptide segments such as FG-rich repeating regions and potential amyloid forming Q/N rich regions (Online Methods, Supplementary note, Fig. 3a and Supplementary Fig. 2). This observation is statistically significant despite the small sample size of the datasets and is in line with recent evidence demonstrating the importance of repeating motifs for the dynamic control of important aspects of cell regulation; for instance, as seen in the Wasf3 protein where repeating linear motifs that are recognized by SH3 domains regulate actin polymerization^{20,23}. Taken together, our findings suggest that translation of asymmetrically localized mRNA at distal sites may dynamically and reversibly influence higher-order assembly formation and facilitate the spatial organization of components through interactions mediated by linear motifs within intrinsically disordered regions.

DSS proteins are regulated by PTMs to create molecular switches

A common mechanism for regulating protein interactions is by post-translational modifications (PTMs). Indeed, PTM sites in or near the vicinity of linear motifs have frequently been shown to conditionally switch motif-mediated interactions between “on” and “off” states¹⁹. We find that despite there being no significant difference in the number of PTM sites in the distal site synthesis group of proteins, as compared to the transport after synthesis set ($P < 0.26$; Wilcoxon-test with correction for multiple testing), distal site synthesis proteins are significantly enriched for PTMs within 5 residues of the putative linear motifs (Fig. 3a and Online Methods). The enrichment of PTMs around and within linear motifs suggests the presence of “on-demand” molecular switches²⁴. The presence of

such switches in proteins that are prone to form reversible assemblies suggest that PTM switches may control the flow of information by regulating the interactions mediated by motifs within disordered regions involving distal site synthesis proteins^{19,20}. This view is supported by the occurrence of assembly promoting regions within proteins such as Apc, whose mRNA is asymmetrically localized and whose protein interactions are regulated by phosphorylation²⁵.

The observed trends are consistent across diverse systems

To test the generality of our observations, we also analyzed datasets from different cell-types, developmental stages and another organism (Fig. 2b, Fig. 3b and Online Methods). Firstly we investigated the asymmetrically distributed proteins⁸ and transcripts²⁶ in the pseudopodia of the fibroblast-like cells of the mouse 3T3 fibroblasts²⁶ and primate COS-7 fibroblasts⁸ cells (see Online Methods, Supplementary note, Fig. 1b and Supplementary Fig. 1 for details). Secondly, we undertook an analysis of the proteomic^{27,28} and transcriptomic²⁹ data of the sensory neurons of adult (3-5 month old)²⁸ and embryonic (E16-P1) rats²⁷. The observations in a different tissue of the same organism, as well as both a different tissue and a different developmental stage of another organism are all qualitatively consistent with our findings in mouse neuroblastoma cells reported above (Fig. 2b, Fig. 3b and Supplementary note). The same trends are therefore identified in different cell types that are polarized for functionally distinct reasons (e.g., pseudopodia of fibroblasts and axons of neurons) suggesting that the observed trends are not simply explained due to the specific properties of proteins involved in forming the neurites within neuroblastoma cells. Furthermore, the consistency of results across datasets from multiple studies suggests that the observed trends are unlikely to be an artifact. It should be noted that while our observations reported here are consistent with a number of well-known distal site synthesis proteins such as Shank1¹⁶ (Table 1), all the trends reported here do not need to, and are unlikely to apply to every distal site synthesized protein (Supplementary note).

DSS proteins and their transcripts are tightly regulated

To analyze the genome-wide studies on gene expression in mouse fibroblast cells³⁰, we used the aforementioned dataset on asymmetric localization in fibroblast-like cells. This dataset allowed us to investigate the different stages of gene expression in the asymmetrically localized gene products within fibroblast-like cells. This analysis revealed that the transcripts of distal site synthesis proteins tend to display a significantly lower transcription rate, a shorter half-life and a lower abundance than transcripts of the transport after synthesis proteins in line with the properties of proteins with high IDR content³¹ (Figure 4a). Furthermore, analysis of the 3' UTR of transcripts encoding distal site synthesis proteins revealed enrichment for mRNA elements that either repress the translation of these transcripts or cause their rapid degradation (Supplementary Fig. 3, 4 and Supplementary note). This suggests that these proteins are under tight post-transcriptional control and supports the notion that cis-elements within localized mRNA are key for regulating their translation at distal sites³². In contrast, we found no significant difference in transcript variability by alternative splicing or alternative promoter usage between the distal site synthesis and transport after synthesis genes (Supplementary note). At the protein level, the distal site synthesis proteins also have a significantly lower abundance (Fig. 4a), and a

shorter half-life than transport after synthesis proteins, despite no discernable difference in translation rate, which is also in line with their high disordered content³³ (Fig. 4a). Given the distinct structural, biophysical and interaction properties of distal site synthesis proteins, it appears that DSS proteins are more highly regulated (compared to TAS proteins) to ensure they are made only when needed and are not present longer than required.

DSS proteins display distinct temporal and PTM dynamics

Proteins need to be present in the right abundance to mediate function, as disruption of balanced gene dosage is usually detrimental to normal cellular behavior³⁴. Given the relatively lower abundance of distal site synthesis proteins, we investigated how such proteins can contribute to cell regulation and signaling. Due to limited availability of proteome-wide time-series datasets, we used data for the mouse orthologs from non-mouse cell lines for this analysis (see Supplementary note). First we investigated a proteome-wide study focusing on temporal changes in protein abundance after the stimulation of the extracellular signal-regulated kinase (ERK) pathway³⁵ in a PC12 rat cell line. We found that distal site synthesis proteins display a significant increase in relative abundance at the 30 minutes post-stimulation time-point, as compared to the 5-minute time point (possibly due to the decoupling of translation from transcription and transport), suggesting there are rapid and dynamic fluctuations in their abundance after stimulation with growth factor (Fig. 4b). We next evaluated changes in phosphorylation state in sites surrounding linear motifs upon stimulation with angiotensin in HEK-293 cells³⁶ (Fig. 4c). We found that distal site synthesis proteins maintain a more robust and long-lasting phosphorylation state upon stimulation compared to TAS proteins (see also Supplementary Fig 6). Furthermore, we found similar trends when analyzing time-series phosphorylation data from two other cell lines, SCC-9³⁷ and HeLa³⁸ cells (see Supplementary Fig. 6). These observations suggest that the abundance and phosphorylation of DSS proteins display distinct and rapid temporal dynamics upon activation of signaling and that the reported trends are independent of cell type and signaling state of the cell.

Together these results suggest that the prior localization of certain transcripts, and their local translation on-demand^{39,40} leads to rapid and dynamic changes in local protein abundance of distal site synthesis proteins by decentralizing gene expression (i.e. decoupling translation from mRNA synthesis in the nucleus and transport to the cytosol). In the synaptic regions, on-site translation of asymmetrically localized transcripts would enable a dynamic turnover of the proteome, which could otherwise take hours or even days if proteins need to be localized after synthesis by protein transport alone⁴¹. The rapid change in abundance and modification state may also permit non-linear input-output responses of spatially localized signaling proteins. Furthermore, the tight post-transcriptional and post-translational control of such assembly-promoting proteins may allow them to act as lynchpins to rapidly and dynamically bring together the signaling and regulatory proteins in space and time (see Supplementary note).

Discussion

Spatial control of biomolecular interactions by localized translation may increase interaction fidelity and signaling sensitivity. Our observations suggest potential benefits for the cell of undertaking protein synthesis after asymmetric subcellular mRNA localization to a distal cellular compartment. These benefits can be broadly categorized into two classes: increasing fidelity of interactions and enhancing sensitivity of spatially restricted systems (Fig. 5 and Supplementary note).

Our systematic analysis indicates that proteins encoded by mRNA that are asymmetrically localized and translated at distal locations tend to have a large number of low-specificity linear motifs within intrinsically disordered regions. The misregulation³¹ or over-expression of proteins containing linear motifs has been suggested to result in toxicity⁴² due to promiscuous interactions^{43,44}. Synthesizing such proteins at distal locations (on-site) would reduce the distance required for these proteins to travel to their site of action. Furthermore, the observed tight temporal regulation might ensure that such proteins (compared to those transported after synthesis) are synthesized only when required and present for only as long as they are needed. In this way, the volume and time available for encountering off-target interaction partners would be limited, minimizing the likelihood of noisy, off-target interactions and thereby increasing interaction fidelity (Fig. 5).

Our results further indicate that the localized translation of such proteins may permit dynamic changes in proteome composition by altering their concentration in spatially restricted locations in a rapid manner. This may be achieved by regulating translational de-repression of the localized transcripts¹¹, by ribosomal activation through distinct signaling pathways⁴⁵ or through specialized ribosomes⁴⁶. This type of regulation would rapidly increase the abundance of specific proteins upon stimulation in distinct sub-cellular locations, thereby facilitating the spatially compartmentalized information processing of spatially restricted signals received by a cell. Whereas relying on changes in transcriptional response alone would create a more general response and could take many hours due to the distance required for the gene products to travel between different locations such as the cell body and synaptic terminals in neurons^{4,41}.

Our results support the emerging view that by decentralizing gene expression, on-site protein synthesis may contribute to interaction fidelity by ensuring that such proteins are rapidly available at the right place, and only when required (Fig. 5). Decoupling of transcription and translation also facilitates rapid response during retrograde signaling⁴ wherein protein products of asymmetrically localized transcripts (e.g. transcription factors; see Supplementary Table 2a) can be rapidly synthesized at a distal location and transported back to regulate processes in a different locations (e.g. transcription in the nucleus). The importance of mRNA localization in maintaining the fidelity of cell regulation is further supported by observations identifying mRNAs localized to their target organelles^{12,40,47}. Such highly specific directed transport of mRNA to sub-cellular organelles, as well as preliminary data from our assessment of mRNAs localized to the mitochondria in yeast (see Supplementary note and Supplementary Fig. 7), supports the notion that an important

purpose of mRNA localization to specific organelles is to maintain fidelity by ensuring that the relevant proteins are made and present in the right location.

Localization in general increases the concentration of components in a restricted environment. In this manner, it facilitates the precise regulation of signal propagation⁴⁸ and enables many regulatory proteins to be present in low copy numbers³⁰. We propose that localized translation of the distal site synthesis proteins can result in rapid but very localized changes in their relative abundance. This would promote the formation of reversible assemblies, which may act as localized scaffolding and reaction centers⁴⁹ that can be dynamically regulated by the local environment (Fig. 5). Localized translation of such proteins will also ensure that the reversible protein assemblies will not form until the scaffolding protein is synthesized locally even though the interaction partners may be widely expressed in the cell. This view is supported by our observation that different members of the same protein complex are seen in the DSS and TAS protein groups (see Supplementary note), as well as the emerging evidence that the asymmetrically localized transcript may itself act as nucleators of cellular assemblies¹⁰. In addition, we suggest that by producing proteins that are prone to mediating off-target, promiscuous interactions at a specific location in a temporally regulated manner^{11,48}, localized translation might (i) facilitate ultra-sensitive (non-linear) input-output behavior, and (ii) reduce the background signaling noise resulting from transient spurious interactions^{22,50}, possibly further sharpening the sensitivity of signaling networks (Fig. 5).

Our findings suggest that the mislocalization of mRNA or misregulation of localized translation may alter the localization and availability of proteins synthesized at a distal site. This may result in off-target and potentially ectopic signaling events, and might be the molecular basis for phenotypic outcomes such as disease as has been demonstrated in the case of Fragile X Syndrome⁵¹. We propose that together with temporal cues, such as signal integration via post-translational modification⁵², the spatial control of proteins by localized translation of asymmetrically distributed transcripts is an important aspect of cell signaling.

Online Methods

Asymmetrically Localized Protein and Transcript Datasets

Procedure to generate the Distal Site Synthesis (DSS) and Transport After Synthesis (TAS) datasets for neurites and fibroblasts—Asymmetrically localized proteins may have multiple subcellular locations and therefore many proteins, though translated on-site, may not appear in our datasets. For this reason, we did not do a comparison between the DSS proteins and the whole proteome because we cannot assume that all other proteins are not asymmetrically localized outside the polarised regions that were investigated. Furthermore, while RNA molecules themselves may be functional when asymmetrically localized⁵⁸ we consider here that protein-coding genes are translated at some point during their lifetime at the distal site. Finally, we did not consider long non-coding RNA molecules.

All mRNAs identified by microarray analysis are assumed to be translated locally at some point in the lifetime of the cell (pM1). The DSS group consists of those genes identified by

microarray analysis in distal locations (i.e. pseudopodia or axon). The TAS group consists of those genes identified by proteomic analysis in the distal process that are not in the DSS group. More precisely the pM1 genes are subtracted from the asymmetrically localized protein dataset (P1) to obtain the TAS group of proteins (P2). The transcripts that are asymmetrically localized (M1) are subtracted from the transcripts whose proteins are asymmetrically localized (mP1) to obtain the transcripts whose protein products are transported after synthesis (mP2) (see Fig. 1b).

Mouse neuroblastoma—The mouse neuroblastoma group of proteins was extracted from two independent studies investigating the asymmetric distribution of proteins⁷ and transcripts¹³ in the neurite of neuroblastoma cells. The raw microarray data from the mouse neuroblastoma transcriptome dataset was extracted from the GEO (Gene Expression Omnibus) database and mapped to the latest version of the Ensembl mouse genome (version 68; <http://www.ensembl.org/>) using the Bioconductor package in R (distal site synthesis (DSS) dataset only). The original processed proteomic data from the paper was used. Only those proteins or mRNAs enriched within the neurite dataset were investigated. In total, there were 79 genes in the DSS dataset and 611 genes in the TAS dataset (see Supplementary Data Table 4).

Fibroblast-like cells—The mouse pseudopodia transcriptome datasets (M1 in Fig. 1b) were extracted from microarray analysis experiments of fibroblast-like cells upon stimulation by either fibronectin or LPA²⁶ from the GEO database and mapped to the latest version of the Ensembl mouse genome (version 68; <http://www.ensembl.org/>) using the Bioconductor package in R (distal site synthesis (dOSS) dataset only). The proteome of the pseudopodia (P1 in Fig. 1b) was extracted from a proteomic study in primate fibroblasts-like cells stimulated by either fibronectin or LPA⁸ using those proteins with multiple peptides identified in pseudopodia sample, as compared to cell body sample. The primate proteins were mapped to 1:1 mouse orthologs using the InParanoid orthologs database⁵⁹ (see Supplementary Data Table 3 and Supplementary note). Both sets of genes identified in these studies were filtered for those whose expression dynamics were globally quantified in mouse fibroblast cells by Schwanhauser et al (2011)^{30,60}. In total, 289 genes were in the DSS protein dataset and 1292 genes in the TAS protein dataset.

Rat sensory neurons—The DSS sensory neuron dataset was extracted from a microarray transcriptome profiling of the growth cones for both adult and embryo rats²⁹. In the same manner as the mouse neuroblastoma database and mouse pseudopodia dataset this was extracted from GEO and mapped to the latest version of the Ensembl mouse genome using the Bioconductor package in R. The TAS protein dataset for adult rat was extracted from a proteome survey of the plasma membrane of DRG sensory neurons by CapLC-MS/MS. All cell body only proteins were removed from dataset²⁸. The embryonic protein dataset was extracted from a proteome study undertaken in the sensory neurons extracted from rat forebrain growth cones at 1-2 days postnatal²⁷. The same procedure as described in fibroblast-like cells was applied to obtain the final rat DRG datasets. In total, 1644 genes were in the embryo DSS dataset, 946 in the adult DSS dataset, 894 genes in the embryonic TAS dataset and 494 genes in the adult TAS dataset.

Yeast mitochondrial dataset—The data for both the DSS and TAS dataset was extracted from a dataset generated by using DNA microarrays probed with mRNA populations associated with free and mitochondrion-bound polysomes⁴⁷. Those mRNA identified with a mitochondrial localization of mRNA (MLR) score⁴⁷ of greater than 80 were considered as distal site synthesis proteins (DSS). Those mRNA confidently identified by the MLR as not associated with mitochondrion-bound polysomes (score less than 40⁴⁷) and that exhibited a mitochondrial localization signal (as annotated by UniProt) were considered as transport after synthesis (TAS) proteins.

Analysis

Sequence Data and Scripts—All sequence data was extracted from Ensembl (version 68; <http://www.ensembl.org/>) BioMart⁶¹ and in-house scripts were written in Python with statistical analysis done using the R statistical platform.

Functional Analysis—Gene Ontology (GO) Term analysis was done using DAVID⁶² (<http://david.abcc.ncifcrf.gov/>) and phenotypic term analysis was performed using MouseMine⁶³ (<http://www.mousemine.org>). The functional analysis uses a combination of the DSS and TAS datasets as the background population of asymmetric proteins (see Supplementary Table 2a and 2c). Further functional analysis combines all mouse datasets to assess general enrichment of asymmetrically distributed proteins against the entire mouse proteome (Supplementary Table 2b).

Analysis of Protein-Protein Interactions—The number of interactions a protein has (i.e. its degree) was calculated using the Python module NetworkX (<http://networkx.github.io/>). Protein-protein interaction data was extracted from the STRING database⁶⁴ (version 8.2; <http://string-db.org/>) with only experimentally validated data included in the analysis. Also, a STRING cut-off score of 0.6 was used to obtain high-confidence interactions. We also analyzed protein complex membership using annotation from the CORUM database⁶⁵ (<http://mips.helmholtz-muenchen.de/genre/proj/corum>).

Analysis of Structural and Functional Protein Segments—The sequence and structural analysis of intrinsically disordered regions (IDRs) and putative linear motifs was undertaken using the IUPred⁶⁶ (long mode; <http://iupred.enzim.hu/>) and ANCHOR¹⁸ (<http://anchor.enzim.hu>) algorithms, respectively. A cut-off value of 0.4 was used as specified in numerous previous studies^{17,67}. To qualify as a candidate linear motif, at least 5 consecutive residues must have an ANCHOR score above 0.4, in agreement with the average length of an annotated motif from the Eukaryotic Linear Motif (ELM) database¹⁷. Structured domains were extracted for all datasets from the Pfam proteomic-wide surveys⁶⁸ (version 27; <http://pfam.sanger.ac.uk/>).

Analysis of Phase Transition Promoting Segments—The role of higher order assembly promoting segments (i.e. phase transition modules) were investigated in three ways:

- (i) First in terms of the valency of the interacting species, as measured by the number of repeating putative linear motifs and/or their binding domains. This incorporates the

presence of repeats (multivalency) of the same known linear motif binding domains (as defined in iELM⁶⁷) or a polyproline region (identified using regular expression: “P+.?P2,?P+”¹⁸).

(ii) The second type of phase transition module is the Q/N-rich regions or polyQ regions that were calculated based on an algorithm described in the studies referenced here^{31,69-72} and defined as a region containing more than 20 Q- or N-enriched amino acids in an 80 residue stretch and containing less than 10 K or E residues in the same region. PolyQ regions are defined as having at least 8 Qs in a 10 residue stretch.

(iii) The third type of phase transition promoting segment are the repeating instances of phenylalanine amino acids occurring next to a glycine residue within an IDR⁷³⁻⁷⁵. These FG repeat domains were initially identified within nuclear pore proteins⁷³ but have recently also been identified as being important for P-body formation and phase transitions in other proteins⁷⁶. It should be noted that it is likely that a combinatorial synergy of all these repeating interaction surfaces may be exploited to promote phase transition.

Analysis of Post-Translational Modifications and Molecular Switches—Post-Translational Modifications (PTMs) were extracted from the PhosphoSite database⁷⁷ (<http://www.phosphosite.org>) and only PTMs experimentally validated by multiple independent studies were included in the analysis. Molecular switches are defined as a PTM site within a putative motif or ± 5 amino acids from the site of a putative linear motif as defined by ANCHOR¹⁹ or by a phase transition module (defined above), with only proteins annotated with the PhosphoSite database included in the analysis. This definition of ± 5 amino acids comes from the previous analyses of annotated examples of molecular switches¹⁸. Transmembrane regions and external PTMs (e.g. glycosylation sites and disulfide bridges) were extracted from annotations within the UniProt database⁷⁸ (version November 2012; <http://www.uniprot.org/>).

Analysis of Functional mRNA Modules—The mRNA sequence data was extracted from Ensembl BioMart⁶¹. miRNA target sites were extracted from the miRANDA website (<http://www.microrna.org/>)⁷⁹ for rat and TargetScan⁸⁰ for mouse (release 6.2; <http://www.targetscan.org/>). AU-rich elements (ARE) were downloaded from the ARED organism database for both rat and mouse⁸¹ (<http://brp.kfshrc.edu.sa/ARED/>). Cytoplasmic polyadenylation elements (CPE) sites were downloaded from a genome-wide analysis performed by Pique et al. (2008)⁸².

Analysis of Proteomic Data—Both the temporal abundance data, as well as the phosphoproteomic data was collected from studies that used SILAC mass spectrometry. SILAC mass spectrometry contrasts two cell populations cultivated in a growth medium either with amino acid labeled with a heavy isotope (condition A) or amino acids labeled with a light/normal isotope (condition B). The proteins from both cell populations are combined and analyzed together by mass spectrometry with pairs of chemically identical peptides differentiated owing to their mass difference. The ratio of peak intensities in the mass spectrum for such peptides pairs reflects the abundance ratio for the two peptides. The

amount of genome-wide proteomic datasets available for temporal abundance data is relatively limited. Therefore data was extracted from a study by von Kriegsheim et al. (2009)³⁵ in PC12 rat cells. The asymmetrically localized proteins of mouse were mapped using 1:1 orthologs using InParanoid⁵⁹. The changes in abundance at a given time point were compared to the protein ratio abundance at the previous time point to obtain the relative change of abundance. This was calculated by assessing the difference in the number of identical peptides retrieved for each protein detected. Peptides were quantified with MSQuant (<http://msquant.sourceforge.net/>) and subsequently normalized³⁵.

Similarly, the temporal phosphoproteomic data³⁸ were obtained from HEK-293 cells³⁶, SCC-9 cells³⁷, and HeLa cells³⁸ and were mapped to obtain the mouse orthologs using InParanoid⁵⁹. Both the SCC-9 and HEK-293 cells samples were analyzed using SILAC mass spectrometry. For each phosphopeptide, Maxquant determines the ratio difference for each of the intensity/time elution profiles for all the isotope distributions present in the different SILAC states³⁶. For the HEK-293 cells, the two conditions are before and after stimulation (3 and 15 minute time points; samples merged in Fig. 3c and unmerged in Supplementary Fig. 6) with angiotensin³⁶. The phosphopeptide ratios in the original paper in SCC-9 cells at 3, 10, and 30 minutes after stimulation with lysophosphatidic acid were compared to the control sample (0 minutes)³⁷. These ratios from the publication were then compared to each previous time point to assess changes in phosphorylation state over time. Only samples that showed evidence of regulation (2 fold change) were included in our calculation. For the HeLa sample, SILAC was also used but data analysis was slightly different. The relative change in phosphorylation state, as measured by extracted ion current (XIC) values were compared to the initial phosphorylation state (i.e. time point 0) for those instances when a single phosphorylation site could be assigned to a peptide in the vicinity of around putative motif. The XIC value is a standard measurement used in quantitative mass spectrometry studies and represents the sum of the total signal observed during the period of peptide elution (see original paper for details³⁸).

Analysis of Transcript Variation—Exon data and splice variant information was extracted from the Ensembl database using the BioMart resource. The alternative promoter annotation was extracted from the Mammalian Promoter Database (MPromDB) (version 2.1; <http://mpromdb.wistar.upenn.edu/>)⁸³.

Statistical Significance—Statistical analysis was done using the R statistical package. Statistical significance was assessed using a one-tailed Wilcoxon-sum rank test when comparing distributions and a one-tailed Fisher's exact test when comparing enrichments. For distributions assessed by the Wilcoxon-sum rank test, the homoscedasticity of the data was first assessed using boxplots. Homogeneity of variance was also formally tested when appropriate. Correction for multiple testing were undertaken using the FDR method designed by Benjamini, Hochberg, and Yekutieli to control for the false discovery rate. The effect size was measured using R effect size for the Wilcoxon-sum rank test or using the odds ratio for the Fisher's exact test. The "Common Language Effect Size (CLES) statistic suggested by McGraw and Wong (1992)⁸⁴ and Cohen's U3 statistic was calculated using the compute.es R package. The common language metric describes the "probability that a data

point sampled at random from distribution A will be greater than a data point sampled from distribution B". Cohen's U3 statistic describes 'the percentage of the A population which the upper half of the cases of the B population exceeds'. Since Wilcoxon is a non-parametric test, based on recommendations by Grissom and Kim (2012)⁸⁵, the following equations were used to convert the U statistic to Cohen's d effect size, which was then converted to

CLES and U3 by the compute.es R package: $U = W - \frac{n_s(n_s+1)}{2}$, where n_s is the smaller of n_a and n_b and W is the Wilcoxon test statistic $\hat{p}_{a.b} = \frac{U}{n_a n_b}$, where n_a and n_b are the sample size of each dataset and P is the probability that a score randomly drawn from population a will be greater than a score randomly drawn from population b

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank S. Bullock, S Balaji, C Chothia, M Buljan, B Lang, M Hegde, G Chalancon, K Van Roey, T Flock, N Latysheva, AJ Venkatakrishnan, and G Toedt for stimulating discussions and their comments on this work. This work was supported by the Medical Research Council (MC_U105185859; RJW;MMB), Human Frontiers Science Program (HFSP) (RGY0073/2010; MMB), the European Molecular Biology Organization (EMBO) Young Investigator Program (MMB), ERASysBio+ (GRAPPLE; RJW;MMB), the European Molecular Biology Laboratory (EMBL) International PhD Program (RJW) and a Canadian Institute of Health Research (CIHR) Postdoctoral Fellowship (RJW).

References

1. Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. *Cell*. 2009; 136:719–730. [PubMed: 19239891]
2. Scott JD, Pawson T. Cell signaling in space and time: where proteins come together and when they're apart. *Science*. 2009; 326:1220–1224. [PubMed: 19965465]
3. Ephrussi A, Dickinson LK, Lehmann R. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell*. 1991; 66:37–50. [PubMed: 2070417]
4. Holt CE, Bullock SL. Subcellular mRNA localization in animal cells and why it matters. *Science*. 2009; 326:1212–1216. [PubMed: 19965463]
5. Jung H, Gkogkas CG, Sonenberg N, Holt CE. Remote Control of Gene Function by Local Translation. *Cell*. 2014; 157(1):26–40. [PubMed: 24679524]
6. Lecuyer E, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell*. 2007; 131:174–187. [PubMed: 17923096]
7. Pertz OC, et al. Spatial mapping of the neurite and soma proteomes reveals a functional Cdc42/Rac regulatory network. *Proc Natl Acad Sci U S A*. 2008; 105:1931–1936. [PubMed: 18245386]
8. Wang Y, et al. Profiling signaling polarity in chemotactic cells. *Proc Natl Acad Sci U S A*. 2007; 104:8328–8333. [PubMed: 17494752]
9. Cajigas IJ, et al. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron*. 2012; 74:453–466. [PubMed: 22578497]
10. Nevo-Dinur K, Nussbaum-Shochat A, Ben-Yehuda S, Amster-Choder O. Translation-independent localization of mRNA in *E. coli*. *Science*. 2011; 331:1081–1084. [PubMed: 21350180]
11. Huttelmaier S, et al. Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature*. 2005; 438:512–515. [PubMed: 16306994]
12. Weis BL, Schleiff E, Zerges W. Protein targeting to subcellular organelles via mRNA localization. *Biochim Biophys Acta*. 2013; 1833:260–273. [PubMed: 23457718]

13. Feltrin D, et al. Growth cone MKK7 mRNA targeting regulates MAP1b-dependent microtubule bundling to control neurite elongation. *PLoS Biol.* 2012; 10:e1001439. [PubMed: 23226105]
14. van der Lee R, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev.* 2014; 114:6589–6631. [PubMed: 24773235]
15. Van Roey K, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* 2014; 114:6733–6778. [PubMed: 24926813]
16. Bockers TM, et al. Differential expression and dendritic transcript localization of Shank family members: identification of a dendritic targeting element in the 3' untranslated region of Shank1 mRNA. *Mol Cell Neurosci.* 2004; 26:182–190. [PubMed: 15121189]
17. Davey NE, et al. Attributes of short linear motifs. *Mol Biosyst.* 2012; 8:268–281. [PubMed: 21909575]
18. Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics.* 2009; 25:2745–2746. [PubMed: 19717576]
19. Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE. The switches.ELM Resource: A Compendium of Conditional Regulatory Interaction Interfaces. *Sci Signal.* 2013; 6:rs7. [PubMed: 23550212]
20. Li P, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature.* 2012; 483:336–340. [PubMed: 22398450]
21. Cumberworth A, Lamour G, Babu MM, Gsponer J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem J.* 2013; 454:361–369. [PubMed: 23988124]
22. Wu H. Higher-order assemblies in a new paradigm of signal transduction. *Cell.* 2013; 153:287–292. [PubMed: 23582320]
23. Takenawa T, Suetsugu S. The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. *Nat Rev Mol Cell Biol.* 2007; 8:37–48. [PubMed: 17183359]
24. Van Roey K, Gibson TJ, Davey NE. Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol.* 2012; 22:378–385. [PubMed: 22480932]
25. Honnappa S, et al. An EB1-binding motif acts as a microtubule tip localization signal. *Cell.* 2009; 138:366–376. [PubMed: 19632184]
26. Mili S, Moissoglu K, Macara IG. Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature.* 2008; 453:115–119. [PubMed: 18451862]
27. Nozumi M, et al. Identification of functional marker proteins in the mammalian growth cone. *Proc Natl Acad Sci U S A.* 2009; 106:17211–17216. [PubMed: 19805073]
28. Xiong X, et al. Enrichment and proteomic analysis of plasma membrane from rat dorsal root ganglions. *Proteome Sci.* 2009; 7:41. [PubMed: 19889238]
29. Gumy LF, et al. Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization. *RNA.* 2011; 17:85–98. [PubMed: 21098654]
30. Schwanhaussner B, et al. Global quantification of mammalian gene expression control. *Nature.* 2011; 473:337–342. [PubMed: 21593866]
31. Gsponer J, Babu MM. Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Rep.* 2012; 2:1425–1437. [PubMed: 23168257]
32. Sasaki Y, Gross C, Xing L, Goshima Y, Bassell GJ. Identification of axon-enriched MicroRNAs localized to growth cones of cortical neurons. *Dev Neurobiol.* 2014; 74:397–406. [PubMed: 23897634]
33. Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science.* 2008; 322:1365–1368. [PubMed: 19039133]
34. Gibson TJ, Seiler M, Veitia RA. The transience of transient overexpression. *Nat Methods.* 2013; 10:715–721. [PubMed: 23900254]
35. von Kriegsheim A, et al. Cell fate decisions are specified by the dynamic ERK interactome. *Nat Cell Biol.* 2009; 11:1458–1464. [PubMed: 19935650]

36. Christensen GL, et al. Quantitative phosphoproteomics dissection of seven-transmembrane receptor signaling using full and biased agonists. *Mol Cell Proteomics*. 2010; 9:1540–1553. [PubMed: 20363803]
37. Mausbacher N, Schreiber TB, Machatti M, Schaab C, Daub H. Proteome-wide analysis of temporal phosphorylation dynamics in lysophosphatidic acid-induced signaling. *Proteomics*. 2012; 12:3485–3498. [PubMed: 23090842]
38. Olsen JV, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*. 2006; 127:635–648. [PubMed: 17081983]
39. Buxbaum AR, Wu B, Singer RH. Single beta-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science*. 2014; 343:419–422. [PubMed: 24458642]
40. Yoon BC, et al. Local translation of extranuclear lamin B promotes axon maintenance. *Cell*. 2012; 148:752–764. [PubMed: 22341447]
41. Scott DA, Das U, Tang Y, Roy S. Mechanistic logic underlying the axonal transport of cytosolic proteins. *Neuron*. 2011; 70:441–454. [PubMed: 21555071]
42. Vavouri T, Semple JJ, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell*. 2009; 138:198–208. [PubMed: 19596244]
43. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol*. 2011; 21:432–440. [PubMed: 21514144]
44. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature*. 2006; 439:168–174. [PubMed: 16273093]
45. Roux PP, Topisirovic I. Regulation of mRNA translation by signaling pathways. *Cold Spring Harb Perspect Biol*. 2012; 4
46. Xue S, Barna M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol*. 2012; 13:355–369. [PubMed: 22617470]
47. Marc P, et al. Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep*. 2002; 3:159–164. [PubMed: 11818335]
48. Gibson TJ. Cell regulation: determined to signal discrete cooperation. *Trends Biochem Sci*. 2009; 34:471–482. [PubMed: 19744855]
49. Hyman AA, Brangwynne CP. Beyond stereospecificity: liquids and mesoscale organization of cytoplasm. *Dev Cell*. 2011; 21:14–16. [PubMed: 21763600]
50. Ladbury JE, Arold ST. Noise in cellular signaling pathways: causes and effects. *Trends Biochem Sci*. 2012; 37:173–178. [PubMed: 22341496]
51. Darnell JC, et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*. 2011; 146:247–261. [PubMed: 21784246]
52. Zheng Y, et al. Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature*. 2013; 499:166–171. [PubMed: 23846654]
53. Ji SJ, Jaffrey SR. Intra-axonal translation of SMAD1/5/8 mediates retrograde regulation of trigeminal ganglia subtype specification. *Neuron*. 2012; 74:95–107. [PubMed: 22500633]
54. Kundel M, Jones KJ, Shin CY, Wells DG. Cytoplasmic polyadenylation element-binding protein regulates neurotrophin-3-dependent beta-catenin mRNA translation in developing hippocampal neurons. *J Neurosci*. 2009; 29:13630–13639. [PubMed: 19864575]
55. Yudin D, et al. Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve. *Neuron*. 2008; 59:241–252. [PubMed: 18667152]
56. Hengst U, Deglincerti A, Kim HJ, Jeon NL, Jaffrey SR. Axonal elongation triggered by stimulus-induced local translation of a polarity complex protein. *Nat Cell Biol*. 2009; 11:1024–1030. [PubMed: 19620967]
57. Gavis ER, Lehmann R. Localization of nanos RNA controls embryonic polarity. *Cell*. 1992; 71:301–313. [PubMed: 1423595]

Reference for Online Methods

58. Han TW, et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell*. 2012; 149:768–779. [PubMed: 22579282]
59. Ostlund G, et al. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res*. 2010; 38:D196–D203. [PubMed: 19892828]
60. Schwanhauss B, et al. Corrigendum: Global quantification of mammalian gene expression control. *Nature*. 2013; 495:126–127. [PubMed: 23407496]
61. Kinsella RJ, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*. 2011; 2011:bar030. [PubMed: 21785142]
62. Dennis GJ, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*. 2003; 4:P3. [PubMed: 12734009]
63. Bult CJ, Eppig JT, Blake JA, Kadin JA, Richardson JE. The mouse genome database: genotypes, phenotypes, and models of human disease. *Nucleic Acids Res*. 2013; 41:D885–D891. [PubMed: 23175610]
64. Franceschini A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41:D808–D815. [PubMed: 23203871]
65. Ruepp A, et al. CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res*. 2010; 38:D497–D501. [PubMed: 19884131]
66. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol*. 2005; 347:827–839. [PubMed: 15769473]
67. Weatheritt RJ, Luck K, Petsalaki E, Davey NE, Gibson TJ. The identification of short linear motif-mediated interfaces within the human interactome. *Bioinformatics*. 2012; 28:976–982. [PubMed: 22328783]
68. Punta M, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012; 40:D290–D301. [PubMed: 22127870]
69. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell*. 2009; 137:146–158. [PubMed: 19345193]
70. Halfmann R, et al. Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol Cell*. 2011; 43:72–84. [PubMed: 21726811]
71. Michelitsch MD, Weissman JS. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A*. 2000; 97:11910–11915. [PubMed: 11050225]
72. Toombs JA, McCarty BR, Ross ED. Compositional determinants of prion formation in yeast. *Mol Cell Biol*. 2010; 30:319–332. [PubMed: 19884345]
73. Ader C, et al. Amyloid-like interactions within nucleoporin FG hydrogels. *Proc Natl Acad Sci U S A*. 2010; 107:6281–6285. [PubMed: 20304795]
74. Kato M, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. 2012; 149:753–767. [PubMed: 22579281]
75. Updike DL, Hachey SJ, Kreher J, Strome S. P granules extend the nuclear pore complex environment in the *C. elegans* germ line. *J Cell Biol*. 2011; 192:939–948. [PubMed: 21402789]
76. Malinovska L, Kroschwald S, Alberti S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim Biophys Acta*. 2013; 1834:918–931. [PubMed: 23328411]
77. Hornbeck PV, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res*. 2012; 40:D261–D270. [PubMed: 22135298]
78. Consortium U. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. 2012; 40:D71–D75. [PubMed: 22102590]
79. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008; 36:D154–D158. [PubMed: 17991681]

80. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120:15–20. [PubMed: 15652477]
81. Halees AS, El-Badrawi R, Khabar KS. ARED Organism: expansion of ARED reveals AU-rich element cluster variations between human and mouse. *Nucleic Acids Res*. 2008; 36:D137–D140. [PubMed: 17984078]
82. Pique M, Lopez JM, Foissac S, Guigo R, Mendez R. A combinatorial code for CPE-mediated translational control. *Cell*. 2008; 132:434–448. [PubMed: 18267074]
83. Gupta R, Bhattacharyya A, Agosto-Perez FJ, Wickramasinghe P, Davuluri RV. MPromDb update 2010: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-seq experimental data. *Nucleic Acids Res*. 2011; 39:D92–D97. [PubMed: 21097880]
84. McGraw KO, Wong SP. A common language effect size statistic. *Psychological bulletin*. 1992
85. Grissom, RJ.; Kim, JJ. Effect sizes for research: Univariate and Multivariate Applications. Routledge; 2012.

Figure 1. Classification and characterization of the Transport After Synthesis (TAS) and Distal Site Synthesis (DSS) group of proteins

a, The two major mechanisms for localizing proteins to distal sites in the cell. **b**, Datasets used to identify groups of DSS and TAS transcripts (M1 and mP2, respectively), as well as DSS and TAS proteins (pM1 and P2, respectively) in mouse neuroblastoma cells (N1E-115 neuronal-like cells), fibroblast-like pseudopodia (COS-7 and NIH3T3 cells, respectively) and rat sensory neurons. For the fibroblast-like cells dataset, mouse genes that are one-to-one orthologs to the primate genes identified in the COS-7 cell line were used in the study

(please see Supplementary note about the validity of this approach). All mRNA identified by microarray analysis are assumed to be translated locally at some point in the lifetime of the cell (pM1). This list of proteins is subtracted from the asymmetrically localized protein dataset (P1) to obtain the transport after synthesis (TAS) group of proteins (P2). The transcripts that are asymmetrically localized (M1) are subtracted from the transcripts whose proteins are asymmetrically localized (mP1) to obtain the transcripts whose protein products are transported after synthesis (mP2) (see Online Methods). **c.** The genome-scale datasets used to investigate the differences between the DSS and TAS groups of proteins (see Supplementary Data Table 1). **d.** An illustration of the concept of asymmetric localization of proteins and mRNA. It is important to note that in neurons, protein transport can take hours or even days to transport proteins between locations^{4,41}.

Figure 2. Structural analysis of Distal Site Synthesis Proteins reveals an enrichment in disordered regions

(a, b) Graphs and boxplots of the distribution of the various structural properties of the distal site synthesis (DSS) (red plots) and transport after synthesis (TAS) (grey plots) proteins of the mouse neuroblastoma datasets (a), the mouse pseudopodia, the rat embryonic sensory neuron dataset and the adult sensory neuron dataset (b). Statistical significance was assessed using the one-tailed Wilcoxon-rank sum test for comparing distributions and the one-tailed Fishers-exact test for comparing enrichments with a false-discovery rate correction for

multiple testing. See individual plots for p-values and each row for respective samples size. The effect size is displayed for each boxplot with a common language metric and Cohen's U3-, D- and Odd-ratio (OR)-statistic (see Online Methods). For example, the common language metric describes "the probability that a score sampled at random from distribution A will be greater than a score sampled from distribution B". The median value for each group of proteins is shown with a horizontal black line. Boxes enclose values between the first and third quartile. Interquartile range is calculated by subtracting the first quartile from the third quartile. All values outside this range are considered to be outliers and were removed from the graphs to improve visualization. The smallest and highest values that are not outliers are connected with the dashed line. The notches correspond to ~95% confidence interval for the median.

Figure 3. Analysis of Distal Site Synthesis Proteins reveals an enrichment for linear motifs, phase-transition (i.e. higher order assembly) promoting segments and PTM sites that act as molecular switches

(a, b) Graphs and boxplots of the distribution of the various regulatory and structural properties of the distal site synthesis (DSS) (red plots) and transport after synthesis (TAS) (grey plots) proteins of the mouse neuroblastoma datasets (a), the mouse pseudopodia, the rat embryonic sensory neuron dataset and the adult sensory neuron dataset (b). The effect size is displayed for each boxplot with a common language metric and Cohen's U3 statistic. See Figure 2 and Online Methods for description of boxplots and statistical tests used.

Figure 4. Dynamic regulation of Distal Site Synthesis transcripts and proteins

a, Boxplot comparing the genome-wide quantitative measurements of gene expression of DSS (red) and TAS (grey) proteins in mouse fibroblast cells. DSS transcripts and proteins have a lower abundance and shorter half-lives suggesting tighter temporal regulation of distal site synthesis transcripts and proteins. **b**, Changes in the abundance of TAS and DSS proteins at 5 and 30 minutes compared to 0 minutes and 5 minutes, respectively, after activation of the extracellular signal-related kinase (ERK) pathway (data from von Kriegsheim *et al*³⁵). DSS proteins have a significant increase in abundance between 5th and

30th minute after stimulation possibly associated with rapid protein synthesis due to decentralized gene expression. **c**, The HEK-293 cells at 3 and 15 minutes (samples merged) after stimulation with angiotensin were compared to the control sample (0 minutes), as calculated in the original paper. A relative increase in the abundance of regulated phosphopeptides in the DSS group of proteins is observed between the control sample and the stimulated samples, as compared to TAS proteins (see Online Methods for details). Only regulated phosphopeptides samples (2 fold change) were included in the analysis. The effect size is displayed for each boxplot with a common language metric and Cohen's U3 statistic. See Figure 2 and Online Methods for description of boxplots and statistical tests used.

Figure 5. An overview of the potential advantages conferred by distal site protein synthesis

An overview of the key inferences from our analysis. Turquoise and red filled circle represents off-target and correct interaction partners, respectively. Wavy lines represent a disordered region within a distal site synthesis protein. Grey and red line in graphs represents profiles of the transport after synthesis and distal site synthesis group of proteins, respectively.

Table 1
Some examples of experimentally validated signaling proteins synthesized at a distal site that are consistent with our observations with functional processes linked to their localized translation.

DSS signaling proteins	Function of local translation	Reference
SMAD 1 / 5 / 8	tissue patterning	Ji & Jaffrey (2012) ⁵³
Shank1	dendrite formation	Bockers et al. (2004) ¹⁶
Beta-Catenin	axonal branching	Kundel et al. (2009) ⁵⁴
RANBP1	neurite regeneration	Yudin et al. (2008) ⁵⁵
PAR3-alpha	axonal elongation	Hengst et al. (2009) ⁵⁶
oskar	establishment of anterior-posterior axis	Ephrussi et al. (1991) ³
nanos	establishment of anterior-posterior axis	Gavis & Lehmann (1992) ⁵⁷