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

Robert J Weatheritt[1,3], Toby J Gibson[2] & M Madan Babu[1]

Although 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 modifications. 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]. The mislocalization of several proteins has been documented to have dramatic effects on development[3] and cell morphology[4], and has also been linked to a number of neurodevelopmental and neurodegenerative diseases[4–6]. 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, coexisting in the same cell type, raises the question of why certain proteins are translated at their site of action in distal locations (**Fig. 1a**).

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 proteins[7] and transcripts[13] (**Fig. 1b**,**c**) in neurites of N1E-115 mouse neuroblastoma cells. Both studies, though done in different laboratories, used the same experimental setup 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. 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) proteins[7] with the transport after synthesis (TAS) proteins[13] (**Fig. 1d** and **Supplementary Fig. 1**). We defined the DSS group of proteins as the proteins whose mRNAs had been detected as asymmetrically distributed in the neurites in microarray studies (**Supplementary Note**). We defined the TAS group of proteins as those proteins that had been detected as asymmetrically distributed in proteomic studies of the neurites and whose transcripts had not been identified in the DSS group (Online Methods and **Fig. 1b**).

An initial functional and phenotype analysis of the two groups of proteins revealed that DSS proteins are likely to be functionally distinct from proteins that are transported after synthesis (**Supplementary Note**, and **Supplementary Tables 1** and **2**). This observation prompted us to carry out a more comprehensive analysis. Our genome- and proteome-scale analyses that integrate multiple large-scale data sets (**Fig. 1c** and **Supplementary Table 1**) revealed that DSS 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 rapidly change local protein abundance, indicating a partial decoupling of the response of signaling networks from transcriptional regulation. Our observations were

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. [2]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [3]Present address: The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. Correspondence should be addressed to R.J.W. (rweather@mrc-lmb.cam.ac.uk) or M.M.B. (madanm@mrc-lmb.cam.ac.uk).

Received 4 December 2013; accepted 22 July 2014; published online 24 August 2014; doi:10.1038/nsmb.2876

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 DSS proteins tend to have a larger number of interaction partners than the TAS proteins (**Fig. 2** 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 different members of the same protein complex in the DSS and TAS protein groups (see below and **Supplementary Note**). Further analysis also revealed a significant difference in the structural properties between the two groups (**Fig. 2a**). Although the TAS proteins were enriched in structured globular domains, the DSS proteins had a significant enrichment in IDRs, which are polypeptide segments that lack stable tertiary structure[14] (**Fig. 2a**). The enrichment of IDRs in DSS 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 localization[16].

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 interactions[17]. However, because of the limited number of amino acids involved in binding, the interaction affinity is often weak[17]. Using the Anchor program[18], we observed that putative linear motifs are significantly enriched within IDRs in the DSS 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 proteins[19]. Such segments have recently been shown to mediate molecular phase transitions from small, soluble entities to large macromolecular assemblies[20]. Such assemblies manifest as granules (also referred as cellular bodies), concentrating protein components in spatially restricted regions, thereby increasing the rate of biomolecular interactions[14,21,22]. Compared to the TAS proteins, we observed that the DSS proteins showed a significant enrichment in repeating proline-rich linear motif binding sites and other phase transition–promoting low-complexity polypeptide segments such as (F+G)-rich repeating regions and potential amyloid-forming (Q/N)-rich regions (**Fig. 3a**, Online Methods, **Supplementary Note**, and **Supplementary Fig. 2**). This observation is statistically significant despite the small sample size of the data sets 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

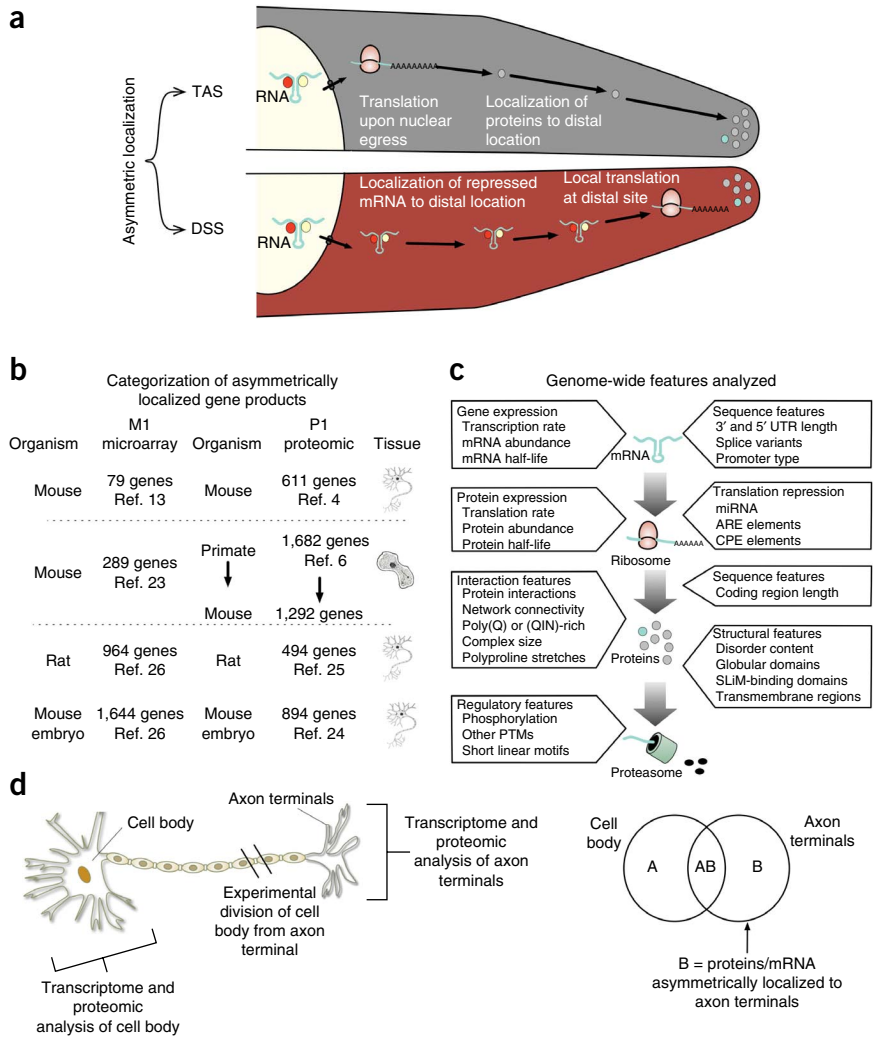

Figure 1 Classification and characterization of TAS and DSS proteins. (**a**) The two major mechanisms for localizing proteins to distal sites in the cell. (**b**) Data sets used to identify groups of DSS and TAS transcripts, as well as DSS and TAS proteins in mouse neuroblastoma cells (N1E-115 neuronal-like cells), fibroblast-like pseudopodia (COS-7 and NIH3T3 cells) and rat sensory neurons. For the fibroblast-like cells data set, mouse genes that are one-to-one orthologs to the primate genes identified in the COS-7 cell line were used in the study (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. This list of proteins was removed from the asymmetrically localized protein data set (P1) to obtain the TAS group of proteins. The transcripts that were asymmetrically localized (M1) were removed from the transcripts whose proteins are asymmetrically localized to obtain the transcripts whose protein products are transported after synthesis (Online Methods). (**c**) The genome-scale data sets used to investigate the differences between the DSS and TAS groups of proteins (**Supplementary Table 1**). UTR, untranslated region; ARE, AU-rich elements; CPE, cytoplasmic polyadenylation element; SLiM, short linear motif. (**d**) An illustration of the concept of asymmetric localization of proteins and mRNA. Note that in neurons, protein transport can take hours or even days to move proteins between locations[4,41].

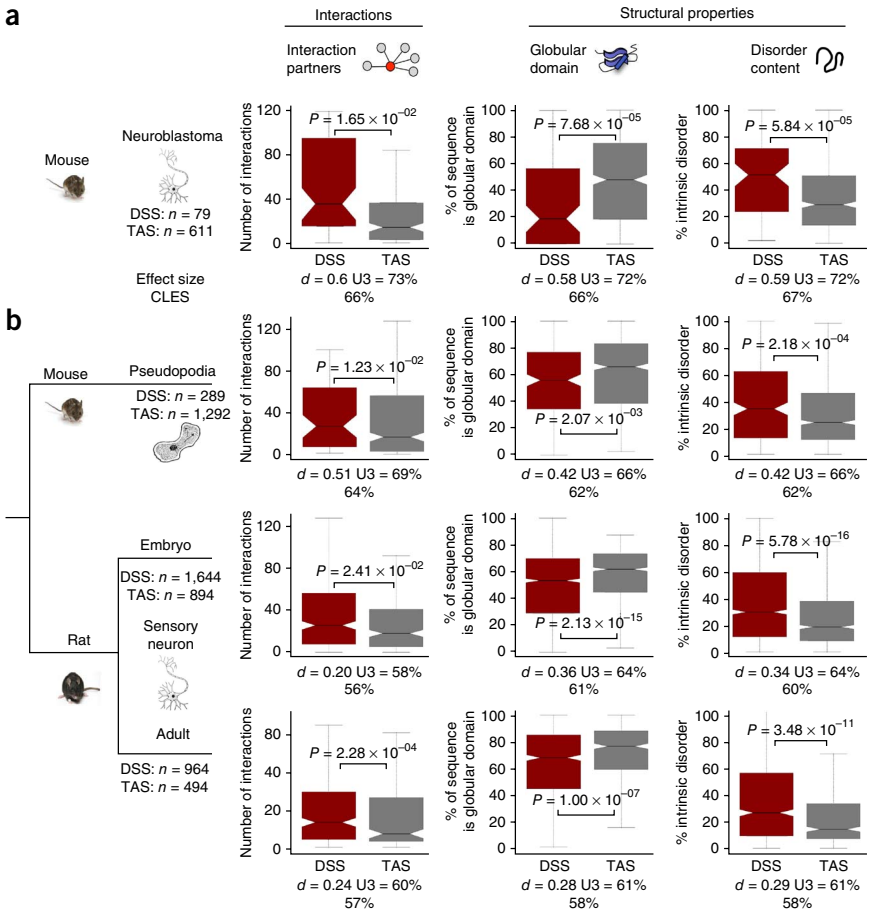

Figure 2 Structural analysis of DSS proteins reveals an enrichment in disordered regions. (**a**,**b**) Distributions of the various structural properties of the DSS and TAS proteins of the mouse neuroblastoma data sets (**a**), the mouse pseudopodia, the rat embryonic sensory neuron data set and the adult sensory neuron data set (**b**). Statistical significance was assessed using the one-tailed Wilcoxon rank-sum test for comparing distributions and the one-tailed Fisher's exact test for comparing enrichments with a false discovery rate correction for multiple testing. The effect size is displayed for each boxplot with a common language effect size (CLES), as well as with the *d* Cohen's U3 statistics (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 dotted line. The notches correspond to ~95% confidence interval for the median.

suggest that translation of asymmetrically localized mRNA at distal sites may dynamically and reversibly influence formation of higher-order assemblies and facilitate the spatial organization of components through interactions mediated by linear motifs in IDRs.

DSS proteins are regulated by post-translational modifications 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' states[19]. We found that despite there being no significant difference in the number of PTM sites in the DSS group of proteins, as compared to the TAS set ($P < 0.26$; Wilcoxon test with correction for multiple testing), DSS proteins were significantly enriched for PTMs within five 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 switches[24]. 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 in disordered regions involving DSS proteins[19,20]. This view is supported by the occurrence of assembly-promoting regions in proteins such as Apc, whose mRNA is asymmetrically localized and whose protein interactions are regulated by phosphorylation[25].

The observed trends are consistent across diverse systems

To test the generality of our observations, we also analyzed data sets from different cell types, developmental stages and another organism (**Figs. 2b** and **3b**, and Online Methods). First we investigated the asymmetrically distributed proteins[8] and transcripts[26] in the pseudopodia of the fibroblast-like cells of the mouse 3T3 fibroblasts[26] and primate COS-7 fibroblasts[8] cells (**Fig. 1b**, Online Methods, **Supplementary Note** and **Supplementary Fig. 1**). Second, we analyzed proteomic[27,28] and transcriptomic[29] data of the sensory neurons of adult rats (3–5 month old)[28] and embryonic rats (embryonic day 16 (E16) to postnatal day 1 (−P1))[27]. 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 were all qualitatively consistent with our findings in mouse neuroblastoma cells reported above (**Figs. 2b** and **3b**, and **Supplementary Note**). Therefore, we identified the same trends in different cell types that are polarized for functionally distinct reasons (for example, pseudopodia of fibroblasts and axons of neurons), which suggests that the observed trends are not simply explained due to the specific properties of proteins involved in forming the neurites in neuroblastoma cells. Furthermore, the consistency of results across data sets from multiple studies and after controlling for occurrences of transmembrane domains (**Supplementary Fig. 3**) suggests that the observed trends are unlikely to be an artifact. Although the observations we report here are consistent with a number of well-known DSS proteins such as Shank1 (ref. 16) (**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 cells[30], we used the aforementioned data set on asymmetric localization in fibroblast-like cells. This data set allowed us to investigate the different stages of gene expression in the asymmetrically localized gene products in fibroblast-like cells. This analysis revealed

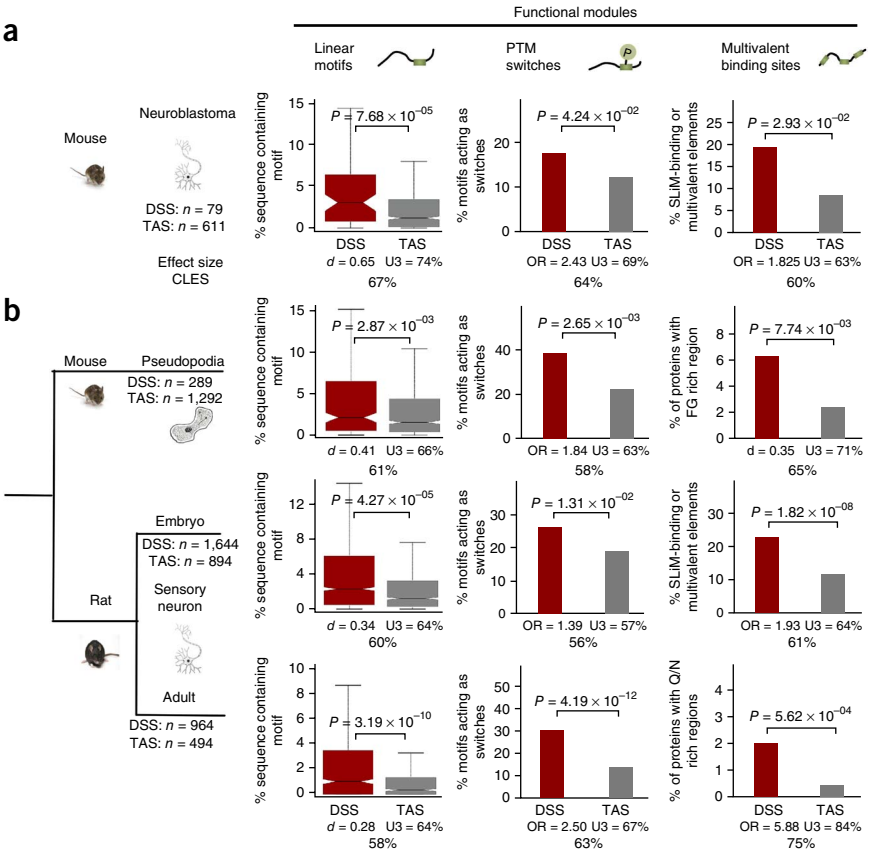

Figure 3 Analysis of DSS proteins reveals an enrichment for linear motifs, phase-transition (i.e., higher-order assembly) promoting segments and PTM sites that act as molecular switches. (**a**,**b**) Distributions of the various regulatory and structural properties of the DSS and TAS proteins of the mouse neuroblastoma data sets (**a**), the mouse pseudopodia, the rat embryonic sensory neuron data set and the adult sensory neuron data set (**b**). The effect size is displayed for each boxplot with a common language metric, Cohen's U3 and odd-ratio (OR) statistic. See **Figure 2** and Online Methods for description of boxplots and statistical tests used.

that the transcripts of DSS proteins have a tendency to display a significantly lower transcription rate, a shorter half-life and a lower abundance than transcripts of the TAS proteins, in line with the properties of proteins with high IDR content[31] (**Fig. 4a**). Analysis of the 3′ untranslated region of transcripts encoding DSS proteins revealed enrichment for mRNA elements that either repress the translation of these transcripts or cause their rapid degradation (**Supplementary Figs. 4** and **5**, 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 sites[32]. In contrast, we found no significant difference in transcript variability by alternative splicing or alternative promoter usage between the DSS and TAS genes (**Supplementary Note**). At the protein level, the DSS proteins also had a significantly lower abundance (**Fig. 4a**), and a shorter half-life than TAS proteins, despite no discernible difference in translation rate, which is also in line with their high disordered content[33] (**Fig. 4a**). Given the distinct structural, biophysical and interaction properties of DSS 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 behavior[34]. Given the relatively lower abundance of DSS proteins, we investigated how such proteins can contribute to cell regulation and signaling. Owing to limited availability of proteome-wide time-series data sets, we used data for the mouse orthologs from non-mouse cell lines for this analysis (**Supplementary Note**). First, we investigated a proteome-wide study that had focused on temporal changes in protein abundance after the stimulation of the extracellular signal-regulated kinase (ERK) pathway[35] in a PC12 rat cell line. We found that DSS proteins displayed a significant increase in relative abundance at the 30 min post-stimulation time point, as compared to the 5-min time point (possibly owing to the decoupling of translation from transcription and transport), which suggests that 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 cells[36] (**Fig. 4c**). We found that DSS proteins maintained a more robust and long-lasting phosphorylation state upon stimulation compared to TAS proteins (**Supplementary Fig. 6**). Furthermore, we found similar trends when analyzing time-series phosphorylation data from two other cells lines, SCC-9 (ref. 37) and HeLa[38] cells (**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 DSS 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 alone[41]. The rapid change in

Table 1 Examples of experimentally validated DSS proteins

DSS signaling proteins	Function of local translation	Ref.
SMAD 1 / 5 / 8	Tissue patterning	53
Shank1	Dendrite formation	16
Beta-catenin	Axonal branching	54
RANBP1	Neurite regeneration	55
PAR3-alpha	Axonal elongation	56
Oskar	Establishment of anterior-posterior axis	3
Nanos	Establishment of anterior-posterior axis	57

These examples of proteins that had been experimentally validated and functionally characterized as being synthesized at distal sites are consistent with our observations.

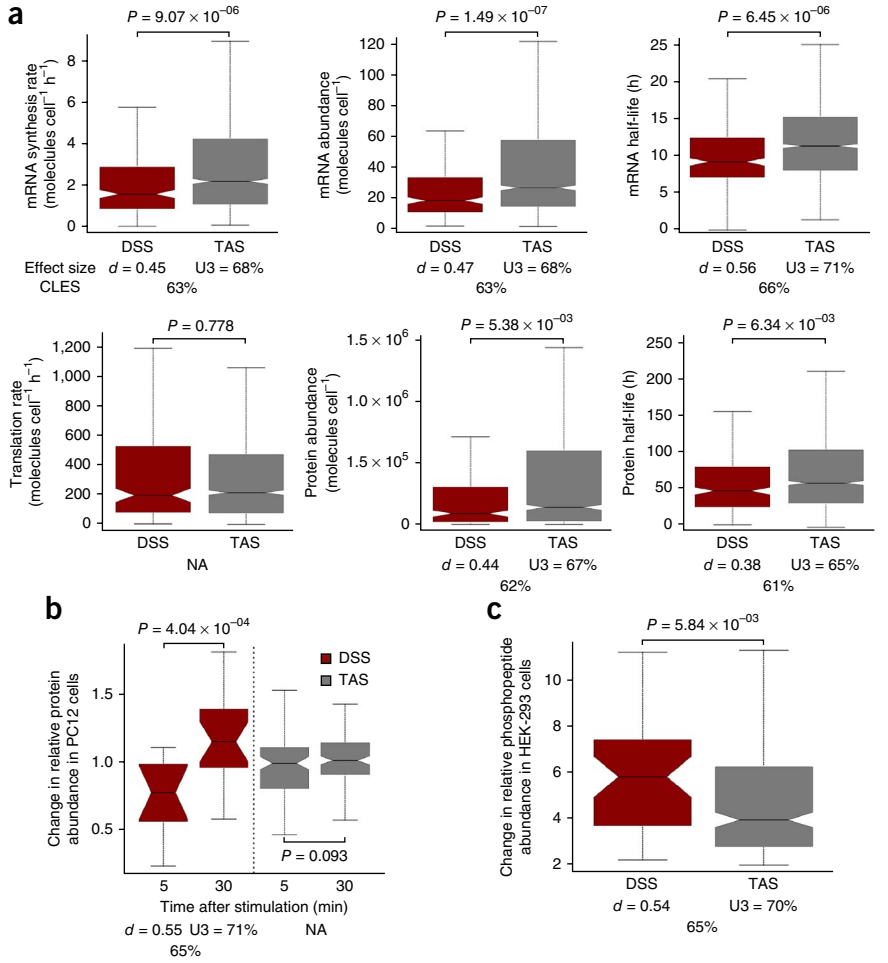

Figure 4 Dynamic regulation of DSS transcripts and proteins. (**a**) Genome-wide quantitative measurements of gene expression of DSS (n = 289) and TAS (n = 1,292) proteins in mouse fibroblast cells. DSS transcripts and proteins have a lower abundance and shorter half-lives suggesting tighter temporal regulation of DSS transcripts and proteins. NA; not applicable. (**b**) Changes in the abundance of TAS and DSS proteins in PC12 cells at 5 min and 30 min compared to 0 min and 5 min, respectively, after activation of the ERK pathway as measured by difference in the normalized peptide counts from the previous time point (data from ref. 35). (**c**) Phosphopeptide abundance in HEK-293 cells at 3 and 15 min (samples merged) after stimulation with angiotensin, compared to the control sample (0 min), as calculated in the original paper. A relative increase in the abundance of regulated phosphopeptides in the DSS group of proteins was observed between the control sample and the stimulated samples, as compared to TAS proteins (Online Methods). Only regulated phosphopeptides samples (twofold 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.

abundance and modification state may also permit nonlinear 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 (**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 in IDRs. The misregulation[31] or overexpression of proteins containing linear motifs has been suggested to result in toxicity[42] owing 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

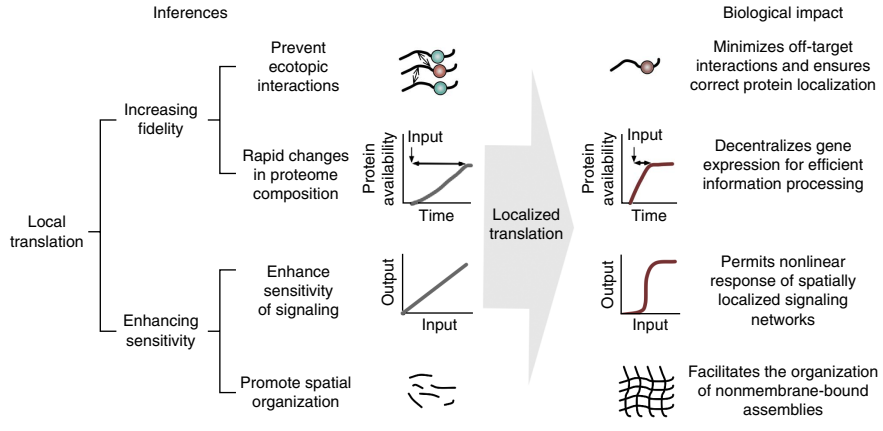

Figure 5 An overview of the potential advantages conferred by distal-site protein synthesis, inferred from our analysis. Turquoise and red filled circle represents off-target and correct interaction partners, respectively. Wavy lines represent a disordered region within a distal site synthesis protein. Grey and red line in graphs represents profiles of the TAS and DSS group of proteins, respectively.

dynamic changes in proteome composition by altering concentration of these proteins in spatially restricted locations in a rapid manner. This may be achieved by regulating translational derepression of the localized transcripts[11], by ribosomal activation through distinct signaling pathways[45] or through specialized ribosomes[46]. This type of regulation would rapidly increase the abundance of specific proteins upon stimulation in distinct subcellular 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 owing 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 signaling[4] wherein protein products of asymmetrically localized transcripts (for example, transcription factors; **Supplementary Table 2a**) can be rapidly synthesized at a distal location and transported back to regulate processes in a different locations (for example, transcription in the nucleus). The importance of mRNA localization in maintaining the fidelity of cell regulation is supported by observations identifying mRNAs localized to their target organelles[12,40,47]. Such highly specific directed transport of mRNA to subcellular organelles, as well as preliminary data from our assessment of mRNAs localized to the mitochondria in yeast (**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 propagation[48] and enables many regulatory proteins to be present in low copy numbers[30]. We propose that localized translation of the DSS 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 centers[49] 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 (**Supplementary Note**) as well as the emerging evidence that the asymmetrically localized transcript may itself act as a nucleator of cellular assemblies[10]. 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 ultrasensitive (nonlinear) 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 syndrome[51]. We propose that together with temporal cues, such as signal integration via post-translational modification[52], the spatial control of proteins by localized translation of asymmetrically distributed transcripts is an important aspect of cell signaling.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

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

AUTHOR CONTRIBUTIONS

R.J.W. collected the data sets and performed the analysis; R.J.W. and M.M.B. designed the study; R.J.W., T.J.G. and M.M.B. conceived the idea and wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Martin, K.C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
2. Scott, J.D. & Pawson, T. Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326**, 1220–1224 (2009).
3. Ephrussi, A., Dickinson, L.K. & Lehmann, R. Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* **66**, 37–50 (1991).
4. Holt, C.E. & Bullock, S.L. Subcellular mRNA localization in animal cells and why it matters. *Science* **326**, 1212–1216 (2009).
5. Jung, H., Gkogkas, C.G., Sonenberg, N. & Holt, C.E. Remote control of gene function by local translation. *Cell* **157**, 26–40 (2014).
6. Lecuyer, E. *et al.* Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174–187 (2007).
7. Pertz, O.C. *et al.* Spatial mapping of the neurite and soma proteomes reveals a functional Cdc42/Rac regulatory network. *Proc. Natl. Acad. Sci. USA* **105**, 1931–1936 (2008).
8. Wang, Y. *et al.* Profiling signaling polarity in chemotactic cells. *Proc. Natl. Acad. Sci. USA* **104**, 8328–8333 (2007).
9. Cajigas, I.J. *et al.* The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* **74**, 453–466 (2012).
10. Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S. & Amster-Choder, O. Translation-independent localization of mRNA in *E. coli*. *Science* **331**, 1081–1084 (2011).
11. Huttelmaier, S. *et al.* Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. *Nature* **438**, 512–515 (2005).
12. Weis, B.L., Schleiff, E. & Zerges, W. Protein targeting to subcellular organelles via MRNA localization. *Biochim. Biophys. Acta* **1833**, 260–273 (2013).
13. Feltrin, D. *et al.* Growth cone MKK7 mRNA targeting regulates MAP1b-dependent microtubule bundling to control neurite elongation. *PLoS Biol.* **10**, e1001439 (2012).
14. van der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
15. Van Roey, K. *et al.* Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* **114**, 6733–6778 (2014).
16. Bockers, T.M. *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.* **26**, 182–190 (2004).
17. Davey, N.E. *et al.* Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).
18. Dosztanyi, Z., Meszaros, B. & Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745–2746 (2009).
19. Van Roey, K., Dinkel, H., Weatheritt, R.J., Gibson, T.J. & Davey, N.E. The switches. ELM Resource: A Compendium of Conditional Regulatory Interaction Interfaces. *Sci. Signal.* **6**, rs7 (2013).

20. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
21. Cumberworth, A., Lamour, G., Babu, M.M. & Gsponer, J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem. J.* **454**, 361–369 (2013).
22. Wu, H. Higher-order assemblies in a new paradigm of signal transduction. *Cell* **153**, 287–292 (2013).
23. Takenawa, T. & Suetsugu, S. The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. *Nat. Rev. Mol. Cell Biol.* **8**, 37–48 (2007).
24. Van Roey, K., Gibson, T.J. & Davey, N.E. Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.* **22**, 378–385 (2012).
25. Honnappa, S. *et al.* An EB1-binding motif acts as a microtubule tip localization signal. *Cell* **138**, 366–376 (2009).
26. Mili, S., Moissoglu, K. & Macara, I.G. Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature* **453**, 115–119 (2008).
27. Nozumi, M. *et al.* Identification of functional marker proteins in the mammalian growth cone. *Proc. Natl. Acad. Sci. USA* **106**, 17211–17216 (2009).
28. Xiong, X. *et al.* Enrichment and proteomic analysis of plasma membrane from rat dorsal root ganglions. *Proteome Sci.* **7**, 41 (2009).
29. Gumy, L.F. *et al.* Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization. *RNA* **17**, 85–98 (2011).
30. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
31. Gsponer, J. & Babu, M.M. Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Reports* **2**, 1425–1437 (2012).
32. Sasaki, Y., Gross, C., Xing, L., Goshima, Y. & Bassell, G.J. Identification of axon-enriched MicroRNAs localized to growth cones of cortical neurons. *Dev. Neurobiol.* **74**, 397–406 (2014).
33. Gsponer, J., Futschik, M.E., Teichmann, S.A. & Babu, M.M. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* **322**, 1365–1368 (2008).
34. Gibson, T.J., Seiler, M. & Veitia, R.A. The transience of transient overexpression. *Nat. Methods* **10**, 715–721 (2013).
35. von Kriegsheim, A. *et al.* Cell fate decisions are specified by the dynamic ERK interactome. *Nat. Cell Biol.* **11**, 1458–1464 (2009).
36. Christensen, G.L. *et al.* Quantitative phosphoproteomics dissection of seven-transmembrane receptor signaling using full and biased agonists. *Mol. Cell. Proteomics* **9**, 1540–1553 (2010).
37. Mausbacher, N., Schreiber, T.B., Machatti, M., Schaab, C. & Daub, H. Proteome-wide analysis of temporal phosphorylation dynamics in lysophosphatidic acid-induced signaling. *Proteomics* **12**, 3485–3498 (2012).
38. Olsen, J.V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
39. Buxbaum, A.R., Wu, B. & Singer, R.H. Single beta-actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* **343**, 419–422 (2014).
40. Yoon, B.C. *et al.* Local translation of extranuclear lamin B promotes axon maintenance. *Cell* **148**, 752–764 (2012).
41. Scott, D.A., Das, U., Tang, Y. & Roy, S. Mechanistic logic underlying the axonal transport of cytosolic proteins. *Neuron* **70**, 441–454 (2011).
42. Vavouri, T., Semple, J.I., Garcia-Verdugo, R. & Lehner, B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **138**, 198–208 (2009).
43. Babu, M.M., van der Lee, R., de Groot, N.S. & Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **21**, 432–440 (2011).
44. Jones, R.B., Gordus, A., Krall, J.A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174 (2006).
45. Roux, P.P. & Topisirovic, I. Regulation of mRNA translation by signaling pathways. *Cold Spring Harb. Perspect. Biol.* **4**, a012252 (2012).
46. Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369 (2012).
47. Marc, P. *et al.* Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep.* **3**, 159–164 (2002).
48. Gibson, T.J. Cell regulation: determined to signal discrete cooperation. *Trends Biochem. Sci.* **34**, 471–482 (2009).
49. Hyman, A.A. & Brangwynne, C.P. Beyond stereospecificity: liquids and mesoscale organization of cytoplasm. *Dev. Cell* **21**, 14–16 (2011).
50. Ladbury, J.E. & Arold, S.T. Noise in cellular signaling pathways: causes and effects. *Trends Biochem. Sci.* **37**, 173–178 (2012).
51. Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
52. Zheng, Y. *et al.* Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature* **499**, 166–171 (2013).
53. Ji, S.J. & Jaffrey, S.R. Intra-axonal translation of SMAD1/5/8 mediates retrograde regulation of trigeminal ganglia subtype specification. *Neuron* **74**, 95–107 (2012).
54. Kundel, M., Jones, K.J., Shin, C.Y. & Wells, D.G. Cytoplasmic polyadenylation element-binding protein regulates neurotrophin-3-dependent beta-catenin mRNA translation in developing hippocampal neurons. *J. Neurosci.* **29**, 13630–13639 (2009).
55. Yudin, D. *et al.* Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve. *Neuron* **59**, 241–252 (2008).
56. Hengst, U., Deglincerti, A., Kim, H.J., Jeon, N.L. & Jaffrey, S.R. Axonal elongation triggered by stimulus-induced local translation of a polarity complex protein. *Nat. Cell Biol.* **11**, 1024–1030 (2009).
57. Gavis, E.R. & Lehmann, R. Localization of nanos RNA controls embryonic polarity. *Cell* **71**, 301–313 (1992).

ONLINE METHODS

Asymmetrically localized protein and transcript data sets. *Procedure to generate the distal-site synthesis and transport after synthesis data sets 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 data sets. 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 polarized regions that were investigated. Furthermore, although RNA molecules themselves may be functional when asymmetrically localized[58] 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 noncoding 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 data set (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) (**Fig. 1b**).

Mouse neuroblastoma. The mouse neuroblastoma group of proteins was extracted from two independent studies investigating the asymmetric distribution of proteins[7] and transcripts[13] in the neurite of neuroblastoma cells. The total RNA from purified neurites analyzed by Affymetrix gene chip technology was extracted from the paper[13]. The original processed proteomic data from the paper were used[7]. Only those proteins or mRNAs enriched in the neurite data set were investigated. In total, there were 79 genes in the DSS data set and 611 genes in the TAS data set (see **Supplementary Table 3**).

Fibroblast-like cells. The mouse pseudopodia transcriptome data sets (M1 in **Fig. 1b**) were extracted from microarray analysis experiments of fibroblast-like cells upon stimulation by either fibronectin or LPA (lysophosphatidic acid)[26] from the GEO database (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10230) and mapped to the latest version of the EnsEMBL mouse genome (version 68; http://www.ensembl.org/) using the Bioconductor package in R (DSS data set 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 LPA[8] 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 database[59] (see **Supplementary Table 4** 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 in refs. 30,60. In total, 289 genes were in the DSS protein data set and 1,292 genes in the TAS protein data set.

Rat sensory neurons. The DSS sensory neuron data set was extracted from a microarray transcriptome profiling of the growth cones for both adult and embryo rats[29]. In the same manner as the mouse neuroblastoma database and mouse pseudopodia data set, this was extracted from GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22638) and mapped to the latest version of the EnsEMBL mouse genome using the Bioconductor package in R. The TAS protein data set for adult rat was extracted from a proteome survey of the plasma membrane of DRG sensory neurons by capillary column liquid chromatography coupled with tandem mass spectrometry (CapLC-MS/MS). All cell body only proteins were removed from data set[28]. The embryonic protein data set was extracted from a proteome study undertaken in the sensory neurons extracted from rat forebrain growth cones at 1–2 d after birth[27]. The same procedure as described in fibroblast-like cells was applied to obtain the final rat dorsal root ganglion (DRG) data sets. In total, 1,644 genes were in the embryo DSS data set, 946 in the adult DSS data set, 894 genes in the embryonic TAS data set and 494 genes in the adult TAS data set.

Yeast mitochondrial data set. The data for both the DSS and TAS data sets were extracted from a data set generated by using DNA microarrays probed with mRNA populations associated with free and mitochondrion-bound polysomes[47]. Those mRNAs identified with a mitochondrial localization of mRNA (MLR) score[47] of greater than 80 were considered as DSS. Those mRNAs confidently identified by the MLR score as not associated with mitochondrion-bound polysomes (score less than 40 (ref. 47)) and that exhibited a mitochondrial localization signal (as annotated by UniProt) were considered as TAS proteins.

Analysis. *Sequence data and scripts.* All sequence data were extracted from EnsEMBL (version 68; http://www.ensembl.org/) BioMart[61], 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 DAVID[62] (http://david.abcc.ncifcrf.gov/) and phenotypic term analysis was performed using MouseMine[63] (http://www.mousemine.org/). The functional analysis uses a combination of the DSS and TAS data sets as the background population of asymmetric proteins (**Supplementary Tables 2a**,**c**). Further functional analysis combined all mouse data sets 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 were extracted from the STRING database[64] (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 database[65] (http://mips.helmholtz-muenchen.de/genre/proj/corum).

Analysis of structural and functional protein segments. The sequence and structural analysis of IDRs and putative linear motifs was undertaken using the IUPred[66] (long mode; http://iupred.enzim.hu/) and ANCHOR[18] (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 five 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) database[17]. Structured domains were extracted for all data sets from the Pfam proteomic-wide surveys[68] (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) was investigated in three ways: 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 iELM[67]) or a polyproline region (identified using regular expression: P+.?P{2,}.?P+; ref. 18).

The second type of phase transition module is the (Q/N)-rich regions or poly(Q) 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 ten lysine or glutamic residues in the same region. Poly(Q) regions are defined as having at least eight glutamines in a ten-residue stretch.

The third type of phase transition–promoting segments are the repeating instances of phenylalanine amino acids occurring next to a glycine residue within an IDR[73–75]. These FG repeat domains had been initially identified in nuclear pore proteins[73] but have recently also been identified as being important for P-body formation and phase transitions in other proteins[76]. 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. PTMs were extracted from the PhosphoSite database[77] (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 ANCHOR[19] 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 switches[18]. Transmembrane regions and external PTMs (for example, glycosylation sites and disulfide bridges) were extracted from annotations within the UniProt database[78] (in November 2012; http://www.uniprot.org/).

Analysis of functional mRNA modules. The mRNA sequence data were extracted from EnsEMBL BioMart[61]. miRNA target sites were extracted from the miRANDA for rat (http://www.microrna.org/)[79] and TargetScan[80] for mouse (release 6.2; http://www.targetscan.org/). AU-rich elements (ARE) were downloaded from the ARED organism database for both rat and mouse[81]

doi:10.1038/nsmb.2876

(http://brp.kfshrc.edu.sa/ARED/). Cytoplasmic polyadenylation element (CPE) sites were downloaded from a genome-wide analysis performed in ref. 82.

Analysis of proteomic data. Both the temporal abundance data, as well as the phosphoproteomic data were collected from studies that used stable isotope labeling by amino acids in culture (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 were 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 data sets available for temporal abundance data is relatively limited. Therefore data were extracted from a study[35] in PC12 rat cells. The asymmetrically localized proteins of mouse were mapped using 1:1 orthologs using InParanoid[59]. 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 normalized[35].

Similarly, the temporal phosphoproteomic data[38] were obtained from HEK-293 cells[36], SCC-9 cells[37] and HeLa cells[38], and were mapped to obtain the mouse orthologs using InParanoid[59]. 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 states[36]. For the HEK-293 cells, the two conditions are before and after stimulation (3 and 15 min time points; samples merged in **Fig. 3c** and unmerged in **Supplementary Fig. 6**) with angiotensin[36]. The phosphopeptide ratios in the original paper in SCC-9 cells at 3 min, 10 min and 30 min after stimulation with lysophosphatidic acid were compared to the control sample (0 min)[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 (twofold 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 ref. 38 for details).

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 false discovery rate 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 in ref. 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'. As Wilcoxon test is a non-parametric test, based on recommendations in ref. 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 data set and $\hat{p}$ is the probability that a score randomly drawn from population *a* will be greater than a score randomly drawn from population *b*.

58. Han, T.W. *et al.* Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
59. Ostlund, G. *et al.* InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38**, D196–D203 (2010).
60. Schwanhausser, B. *et al.* Corrigendum: Global quantification of mammalian gene expression control. *Nature* **495**, 126–127 (2013).
61. Kinsella, R.J. *et al.* Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**, bar030 (2011).
62. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).
63. Bult, C.J., Eppig, J.T., Blake, J.A., Kadin, J.A. & Richardson, J.E. The mouse genome database: genotypes, phenotypes, and models of human disease. *Nucleic Acids Res.* **41**, D885–D891 (2013).
64. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815 (2013).
65. Ruepp, A. *et al.* CORUM: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Res.* **38**, D497–D501 (2010).
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.* **347**, 827–839 (2005).
67. Weatheritt, R.J., Luck, K., Petsalaki, E., Davey, N.E. & Gibson, T.J. The identification of short linear motif-mediated interfaces within the human interactome. *Bioinformatics* **28**, 976–982 (2012).
68. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–D301 (2012).
69. Alberti, S., Halfmann, R., King, O., Kapila, A. & Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* **137**, 146–158 (2009).
70. Halfmann, R. *et al.* Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol. Cell* **43**, 72–84 (2011).
71. Michelitsch, M.D. & Weissman, J.S. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. USA* **97**, 11910–11915 (2000).
72. Toombs, J.A., McCarty, B.R. & Ross, E.D. Compositional determinants of prion formation in yeast. *Mol. Cell. Biol.* **30**, 319–332 (2010).
73. Ader, C. *et al.* Amyloid-like interactions within nucleoporin FG hydrogels. *Proc. Natl. Acad. Sci. USA* **107**, 6281–6285 (2010).
74. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
75. Updike, D.L., Hachey, S.J., Kreher, J. & Strome, S. P granules extend the nuclear pore complex environment in the C. elegans germ line. *J. Cell Biol.* **192**, 939–948 (2011).
76. Malinovska, L., Kroschwald, S. & Alberti, S. Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim. Biophys. Acta* **1834**, 918–931 (2013).
77. Hornbeck, P.V. *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.* **40**, D261–D270 (2012).
78. UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71–D75 (2012).
79. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
80. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
81. Halees, A.S., El-Badrawi, R. & Khabar, K.S. ARED organism: expansion of ARED reveals AU-rich element cluster variations between human and mouse. *Nucleic Acids Res.* **36**, D137–D140 (2008).
82. Pique, M., Lopez, J.M., Foissac, S., Guigo, R. & Mendez, R. A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434–448 (2008).
83. Gupta, R., Bhattacharyya, A., Agosto-Perez, F.J., Wickramasinghe, P. & Davuluri, R.V. MPromDb update 2010: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-seq experimental data. *Nucleic Acids Res.* **39**, D92–D97 (2011).
84. McGraw, K.O. & Wong, S.P. A common language effect size statistic. *Psychol. Bull.* **111**, 361–365 (1992).
85. Grissom, R.J. & Kim, J.J. *Effect Sizes for Research: Univariate and Multivariate Applications* (Routledge, 2012).