LncRNAs interacting with the translation machinery contribute to human neuronal differentiation

Katerina Douka^{1,2}, Isabel Birds^{1,2}, Dapeng Wang², Sophie Clayton^{1,3}, Abigail Byford^{1,3}, Elton J. R. Vasconcelos², Mary J. O'Connell⁵, Jim Deuchars^{2,3}, Adrian Whitehouse^{1,2,4} and Julie L. Aspden^{1,2} #

¹ School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, LS2 9JT, UK.

² LeedsOmics, University of Leeds, UK

³ School of Biomedical Sciences, Faculty of Biological Sciences, University of Leeds, LS2 9JT, UK.

⁴ Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.

⁵ School of Life Sciences, Faculty of Medicine and Health Sciences, The University of Nottingham, Nottingham, NG7 2UH, UK.

corresponding author: j.aspden@leeds.ac.uk

Abstract

LncRNAs are less conserved, yet more tissue and developmental-stage specific than mRNAs and are particularly enriched in the nervous system of *Drosophila melanogaster*, mouse and human. The function of cytoplasmic lncRNAs and their potential translation remains poorly understood. Here we performed Poly-Ribo-Seq to understand the interaction of lncRNAs with the translation machinery and the functional consequences during neuronal differentiation of SH-SH5Y cells. We discovered 237 cytoplasmic lncRNAs upregulated during early neuronal differentiation, most of which are associated with polysome complexes. The majority are cytoplasmically enriched and are intergenic or anti-sense. In addition, we find 45 small ORFs in lncRNAs to be actively translated, 17 specifically upon differentiation. 11 of these smORFs exhibit high sequence conservation across *Hominidae* suggesting they are under strong selective constraint with putative function in this clade. We discover LINC01116 is induced upon differentiation and contains an 87 codon smORF, which we detect as translated, with increased ribosome profiling signal upon differentiation. The LINC01116 peptide exhibits a cytoplasmic distribution and is detected in neurites. Knockdown of LINC01116 results in significant reduction of neurite length in differentiated cells indicating it contributes to neuronal differentiation. Our findings indicate lncRNAs are a source of non-canonical peptides and contribute to neuronal function.

Introduction

Long non-coding RNAs (lncRNAs) are >200nt and thought to lack the potential to encode proteins. lncRNAs are well known to play key roles in development and differentiation via several

mechanisms (Dimartino et al., 2018; Tsagakis et al., 2020), such as the base-pairing of lnc-31 with Rock1 mRNA, to promote its translation during myoblast differentiation (Dimartino et al., 2018). ~40% of human lncRNAs are specifically expressed in the brain (Derrien et al., 2012), where they display precise spatiotemporal expression profiles (Ponting et al., 2009). A subset of nuclear neuronal lncRNAs have been found to regulate neuronal differentiation in mouse and human (Carelli et al., 2019; Chodroff et al., 2010; Lin et al., 2014; Winzi, 2018). Until recently, the majority of work has focused on nuclear lncRNA functions, following the belief that most lncRNAs were retained in the nucleus (Derrien et al., 2012; Djebali et al., 2012). However, it has become increasingly clear that many lncRNAs are exported to the cytoplasm (Carlevaro-Fita et al., 2016) and have specific cytoplasmic functions in post-transcriptional gene regulation whilst some possess specific neuronal roles. For example, UCHL1-AS is detected in dopaminergic neurons and promotes translation of UCHL1 mRNA by increasing its association with heavy polysomes (Carrieri et al., 2012). Whereas BACE1-AS transcript, which is significantly upregulated in the brain of Alzheimer's disease patients, base-pairs with beta-secretase-1 (BACE1) mRNA, stabilising it by masking miR-485-5p binding sites (Faghihi et al., 2010). Alternatively, BC200 represses translation initiation in dendrites by disrupting the formation of pre-initiation 48S complexes (Wang et al., 2002).

Translation regulation is essential during neuronal differentiation (Mohammad et al., 2019), with the translation of non-canonical ORFs, e.g. upstream ORFs (uORFs), playing important roles (Blair et al., 2017; Fujii et al., 2017; Rodriguez et al., 2019). Ribosome profiling in a range of organisms and tissue types has revealed the translation of a variety of non-canonical ORFs including small ORFs (smORFs) <100 codons in length (Aspden et al., 2014; Blair et al., 2017; Duncan & Mata, 2014; Fujii et al., 2017; Guo et al., 2010; Ingolia et al., 2013; Rodriguez et al., 2019). Although these translation events remain controversial, it is clear that lncRNAs can interact with the translation machinery. Limited ribosome profiling signal found on smORFs might be the result of sporadic binding of a single ribosome and may not necessarily correspond to active translation. Poly-Ribo-Seq was previously developed to distinguish those lncRNAs that are bound by multiple ribosomes, and therefore actively translated, from such a background signal. Predicting which lncRNA-smORFs are translated has proven difficult, as attributes to predict protein-coding ability are ineffective on small non-canonical ORFs. A small but growing number of smORF peptides translated from lncRNAs have been found to exhibit cellular and organismal functions (Anderson et al., 2015; Chen et al., 2020; Magny et al., 2013; Pueyo & Couso, 2008; Spencer et al., 2020; Wang et al., 2020).

To study the dynamic interactions of lncRNAs with the translation machinery and identify actively translated cytoplasmic lncRNAs during early neuronal differentiation, we perform Poly-Ribo-Seq on SH-SY5Y cells (human neuronal cell line). During differentiation a global decrease in translation

is evident and ribosomal protein mRNAs are specifically excluded from the polysomes during differentiation as part of this down-regulation. We detect ~800-900 lncRNA genes in the cytoplasm of SH-SY5Y cells. Specifically, 237 lncRNAs are upregulated and 100 downregulated during differentiation, 58-70% of which interact with polysome complexes. These interactions are dynamic during differentiation. Ribo-Seq identified 45 actively translated smORFs in lncRNAs, several of which are regulated during differentiation. A subset of these translation events was validated by transfection tagging experiments and analysis of publicly available mass spectrometry data. High levels of sequence conservation across *Homimidae* indicates that the resulting lncRNA-smORF peptides are produced in other great apes and likely possess cellular functions. One of the lncRNA genes that we discover to be significantly upregulated during differentiation is LINC01116, which is cytoplasmically enriched and associated with polysomes. Our Ribo-Seq analysis reveals the translation of an 87aa peptide from LINC01116 with increased ribosome profiling coverage upon differentiation. The LINC01116 peptide exhibits a cytoplasmic distribution and is detected in neurites. We reveal that LINC01116 contributes to neuronal differentiation and is required for proper neurite lengthening.

Results

Differentiation of SH-SY5Y cells with retinoic acid results in reduced translation levels

To dissect the importance of cytoplasmic lncRNAs and their ribosome associations in early neuronal differentiation we profiled the differentiation of SH-SY5Y cells with retinoic acid (RA) for 3 days. This treatment results in neuronal differentiation as indicated by neurite elongation (Sup 1A), which can be seen by immunostaining for neuronal β III-tubulin (TuJ1) (Fig 1A). Quantification of neurite length reveals significant elongation upon RA treatment (Fig 1B). There is also increased expression of neuronal markers; more cells express c-Fos upon differentiation (Sup 1B, C). There is a concomitant reduction in cell proliferation, as seen by reduced number of Ki67+ cells (Sup 1D, Fig 1C) as well as a reduction in the levels of pluripotency marker SOX2 (Sup 1E). These differentiated cells exhibit characteristics of outer radial glia neural progenitors and are driven towards a noradrenergic phenotype as indicated by increased Dopamine Beta Hydroxylase like Monooxygenase Protein 1 (MOXD1) expression (Sup 1F) (Pollen et al., 2015; Xin et al., 2004). Translational output assessment by polysome profiles (Fig 1D) reveals that differentiation results in down-regulation of global translation. Quantification of translation complexes across the sucrose gradients indicates that levels of polysomes are reduced with respect to 80S monosomes (Sup 1G) resulting in a reduced polysome to monosome ratio (Fig 1E). This down-regulation of translation is accompanied by a shift of ribosomal protein (RP) mRNAs from polysomes to monosomes; e.g. Rpl26 mRNA (Fig 1F, Sup 1H), Rps28 (Sup 1I) and Rpl37

(Sup 1J), as measured by RT-qPCR across gradient fractions. This reduced synthesis of RPs has previously been reported during neuronal differentiation (Blair et al., 2017; Chau et al., 2018).

Poly-Ribo-Seq reveals differences in RNA expression and polysome association upon differentiation

To profile RNA, ribosome association and translational changes upon differentiation we employed Poly-Ribo-Seq (Aspden et al., 2014) with some minor modifications to adapt to human neuronal cells (Fig 2A). This adaptation of ribosome profiling (Ribo-Seq) can detect which RNAs are cytoplasmic, polysome associated and translated (Fig 2A). We sequenced i) poly-A selected cytoplasmic RNA, ‘Total’ RNA-Seq, ii) Polysome-associated poly-A RNAs, ‘Polysome’ RNA-Seq, and iii) ribosomal footprints, Ribo-Seq, from control and RA differentiated cells, with three biological replicates (Materials and Methods for details).

PCA of the 3 Total and 3 Polysome RNA-Seq for both Control and RA treatment shows that RA treated samples cluster separately from Control samples and biological replicates generally cluster together (Sup 2A). To profile differences as a result of differentiation we compared Control and RA samples, performing differential expression analysis at the gene level for both Polysomal RNA-Seq samples (main figures) and Total Cytoplasmic RNA-Seq samples (supplemental figures). To dissect the ability of RNAs to associate with translation complexes we compared the “Polysomal” and “Total” RNA populations for both Control and RA conditions, again through differential expression analysis at the gene level. Differences in expression between Control and RA are greater than those seen between Total and Polysome populations. Interestingly the variation in gene expression between differentiated (RA) Total and Polysome is greater than the variation between Control Total and Control Polysome (Sup 2A).

Analysis of the two types of RNA-Seq revealed large changes in RNA expression in response to differentiation, as previously determined (Blair et al., 2017). In fact, 897 protein-coding genes are upregulated upon RA treatment and 732 downregulated, within the Polysomes (Fig 2B). A very similar pattern is seen from Total; 936 up-regulated and 691 down-regulated genes (Sup 2B), and 78% of the significantly affected protein-coding genes are in common between Total and Polysome populations (Sup 2C&D). GO term analysis of these differentiation-regulated protein-coding genes indicates that functions corresponding to neuronal differentiation are upregulated (Fig 2C). There is an enrichment of ribosome and translation functions in those genes downregulated upon neuronal differentiation (Fig 2D). A clearer enrichment of genes with neuronal functions is evident in GO analysis of Total RNA-Seq (Sup 2E), e.g. synaptic transmission and similar enrichment of translation functions are in the downregulated set (Sup 2F). Expression of genes required for neuronal differentiation is increased or

induced upon RA treatment, e.g. Secretogranin, which is involved in packaging of neuropeptides into secretory vesicles (Sossin & Scheller, 1991) (Fig 2E).

Comparison of Total and Polysome RNA-Seq indicates that some mRNAs are enriched in polysomes whilst others are depleted. Ribosomal protein mRNAs are specifically depleted from polysomes upon differentiation, as seen by their altered relative levels between Total and Polysomes in Control (Fig 2F) and RA (Fig 2G). This is consistent with a global downregulation in translation (Fig 1D) and shift of RP mRNAs out of the polysomes (Fig 1F). By comparing these polysome enriched and depleted mRNAs between Control and RA the dynamic nature of polysome association during differentiation is revealed (Fig 2H). More mRNAs become depleted during differentiation, including more RP mRNAs. GO term analysis of those genes enriched or depleted in the polysomes (Sup 2G, H, I) supports this depletion of RP mRNAs upon differentiation (Sup 2I).

LncRNAs are induced and associated with polysomes during differentiation

To understand the potential role and regulation of cytoplasmic lncRNAs we analysed their expression levels in Total and Polysome populations upon differentiation (Sup 3A). We detect large numbers of lncRNAs present in the cytoplasm. 801 lncRNA genes expressed and present (i.e. have RPKMs ≥ 1) in the cytoplasm under control conditions and 916 lncRNA genes in differentiated cells. Interestingly, 237 lncRNA genes were upregulated during differentiation whilst only 82 were downregulated, within Polysome fractions (Fig 3A). A very similar pattern was seen when considering Total RNA-seq with 178 lncRNA genes up-regulated and 100 down-regulated (Sup 3B). The majority of these differences were identified in both Total and Polysome populations (70% of the upregulated and 58% of the downregulated lncRNAs were in common between total and polysome) (Sup 3C&D). Interestingly when we look in more detail at these lncRNA regulated during differentiation the majority of differentiation-regulated lncRNAs are either intergenic or/and anti-sense lncRNAs (Fig 3 B&C-Polysome, Sup 3 E&F-Total). Direct comparison of levels of lncRNAs within the whole cytoplasm and within polysome fractions indicates the vast majority lncRNAs (Control: 98% and RA: 99% are neither polysome enriched nor depleted (Sup 4 A&B). A small number (Control: 12 and RA: 10) of lncRNAs are specifically depleted from the polysomes including several lncRNAs derived from small nucleolar RNA host genes. Interestingly there is a smaller proportion of anti-sense lncRNAs within these polysome-depleted populations (Sup 4 C&D) compared to differences in expression during differentiation (Fig 3 B&C).

Significant induction of specific polysome-associated lncRNAs during differentiation, such as DLGAP1-AS2 is suggestive of a regulatory role during neuronal differentiation (Fig 3 A&D). We validated these differentiation-induced changes in a subset of 7 lncRNAs (Fig 2E). By RT-qPCR the

expression of all candidate lncRNAs were significantly upregulated upon differentiation, as was determined by RNA-Seq analysis. Fold-changes were highly correlative between RNA-Seq and RT-qPCR (Sup 4E). To enable us to focus on lncRNAs with potential neuronal functions we selected candidate lncRNAs that exhibited the highest fold increase in levels upon differentiation. The majority of these candidate lncRNAs (6/7) are specifically enriched in the cytoplasm, rather than the nucleus (Fig 3E, Sup 4F) in contrast to the known nuclear lncRNA Xist. Only 2/7 lncRNAs tested showed changes in cytoplasmic enrichment upon differentiation (Sup 4G), the majority do not (5/7). When these candidate lncRNAs were profiled across sucrose gradient fractions they were found to associate with polysome complexes within the cytoplasm. LINC02143, which is induced >22-fold during differentiation (Fig 3E), is highly enriched in the cytoplasm compared to the nucleus (Fig 3F) and is found in monosomes and small polysomes (Fig 3G). It is also just as upregulated in polysomes (Fig 3A) as in the total cytoplasm (Sup 3B) during differentiation. Unlike DLGAP1-AS2, DLGAP1-AS1 is enriched in the cytoplasm (Fig 3F) and associates with small polysomes, as well as ribosomal subunits (Sup 4H).

Another lncRNA whose levels significantly increase during differentiation is LINC01116 (Fig 3 A&E), which is involved in the progression of glioblastoma (GBM) (Brodie et al., 2017). LINC01116 is enriched in the cytoplasm (Fig 3F), detected at high levels in the 80S (monosome) fraction and in small polysomal complexes (Fig 3H). Upon differentiation there is an increase in the amount of LINC01116 present in disomes, compared to Control. This is consistent with the upregulation of LINC01116 transcript in the polysomes detected by RNA-Seq, indicating a functional interaction of LINC01116 with polysomes during differentiation. In fact, the majority of LINC01116 transcript was found to associate with polysomes in both undifferentiated (Control) cells (66%) and upon differentiation (RA-57%), suggesting it could either be translated or associating with translation complexes (Fig 3H).

Translation of lncRNA-smORFs during differentiation

To better understand the association of lncRNA with polysome complexes and their potential translation we analysed ribosome footprinting from our Poly-Ribo-Seq experiments (Fig 2A). Framing analysis reveals good framing, specifically at footprint lengths of 31 and 33 nt (Fig 4A) and discrete footprint lengths (Sup 5A). On average 95% of ribosome footprints mapped to CDSs, whilst in RNA-Seq this was only 53% (Sup 5B). Metagene analysis of 31 nt length reads indicates that framing is high within ORFs, and that the footprint signal outside of ORFs is mostly within 5'-UTRs rather than 3'-UTRs, where there is a sharp drop off in footprints at the stop codon (Sup 5C). Together these attributes indicate that our Ribo-Seq quality is high and represents genuine translation events. Since 31 and 33nt reads exhibited high triplet-periodicity they were selected for downstream analysis of translation events to identify ORFs that are translated (Fig 4B). Overall, we are able to detect 16,282 protein-

coding ORFs translated in Control and 12,745 in RA, 10,014 of which are translated in both conditions (Table 1). Interestingly we are also able to detect the translation of 71 uORFs and 5 dORFs (Table 1).

To monitor translational changes we performed differential translation analysis (Chothani et al., 2019). Differential ribosome occupancy on ORFs (Δ RPF), differential mRNA expression (Δ RNA) and differential translation efficiency (Δ TE) were calculated (Chothani et al., 2019). The majority of changes in protein production are driven by transcriptional regulation during differentiation, with no change in TE (Fig 4C). We detected 3 protein-coding CDSs being exclusively regulated by translation, upon differentiation (Fig 4C). One of these is centromere protein S (CENPS), which is expressed at low levels in neuroblastoma tumours and is thought to reduce cell growth (Carén et al., 2005; Krona et al., 2004). We were also able to identify CDSs that are known to be regulated during differentiation such as cytochrome P450 26A1 and cytochrome P450 26B1 (CYP26A1, CYP26B1), involved in RA metabolism (Maden, 2007) (Fig 4D). These ORFs are transcriptionally and translationally upregulated, and their TE is significantly increased upon differentiation.

Importantly, using our pipeline (Fig 4B) we also detect the translation of 45 ORFs within lncRNAs, 28 in control and 23 during differentiation, 6 of these were translated in both conditions (Fig 4E). Interestingly, TEs for these lncRNA-smORFs are similar to those of protein-coding ORFs (Fig 4F), providing further evidence that these are genuine translation events. The dORFs that we detect are translated at much lower efficiencies (Fig 4F). This makes sense considering that ribosomes will likely have to reinitiate after the main ORF, which will occur at a low efficiency. One of the lncRNA-smORFs we identify to be translated in differentiated cells, has previously been characterised, the 84 aa CRNDP (Szafron et al., 2015). Analysis of these 45 translated smORFs from lncRNAs shows that they are <300 aa in length with a median size of 56 aa in Control and 64 aa in RA (Fig 4G, Sup 5D).

Previous analysis has indicated that smORF peptides exhibit specific amino acid usage, which suggests they are both genuine proteins and show enrichment for transmembrane alpha-helices (Song et al., 2010). Therefore, we profiled the amino acid composition of our lncRNAs-smORFs, uORFs and dORFs compared to protein-coding ORFs and expected frequency (Fig 4H). lncRNA-smORFs exhibit similar frequencies to known protein-coding ORFs. Specifically, smORFs possess lower than expected arginine levels, but not as low as known protein-coding ORFs. Amino acid usage does not suggest that any smORF groups have propensity to form transmembrane alpha-helices (Aspden et al., 2014).

From within the set of lncRNAs we identified as induced during differentiation, several contained translated smORFs. One of these is LINC01116, whose translated ORF exhibits high framing (Fig 4I). In fact, ~80% of reads that map to this ORF are in frame 2, whereas outside this ORF, the few reads mapping to the remaining lncRNA sequence are far more equally distributed between the three

possible frames (Fig 4I). This smORF is 71-codons, and its start codon is 608nt into the lncRNA. Ribosome profiling signal is substantially higher upon differentiation, mainly as a result of increased lncRNA transcript abundance (Sup 5E).

Peptide synthesis from smORFs in lncRNAs during differentiation

Our pipeline is highly stringent, i.e., there are many additional ORFs that display good framing but do not pass our thresholds for numbers of footprinting reads or exhibit background signal in rest of the lncRNA. Therefore, we are confident these translation events are taking place. To validate peptide synthesis, we have taken two complementary strategies; mass spectrometry analysis and transfection of ORF tagging constructs. Mass spectrometry generally only detects 90% of translated proteins from protein-coding genes (Zubarev, 2013). Analysis of previously published mass spectrometry datasets from SH-SY5Y (undifferentiated and RA-treated) cells (Brenig et al., 2020; Murillo et al., 2018) supports the production of peptides from 6 of our translated uORFs (2 Control, 4 RA) and 8 lncRNA-smORFs (4 Control and 4 RA) (Fig 5A). This represents ~8% (uORFs) and 18% (lncRNA-smORFs) of what is detected in Poly-Ribo-Seq. This is to be expected given the small size of these peptides and therefore reduced chance of producing peptides >8aa from digestion for mass spectrometry detection (Saghatelian & Couso, 2015).

To validate translation of our lncRNA-smORFs that were not identified in previous mass spectrometry but were detected as translated by our Poly-Ribo-Seq analyses, we have also used a transfection tagging approach. We cloned the lncRNA sequence to the 5' of the putative ORF, termed the 5'-UTR and the smORF itself, without its stop codon, with a C terminal 3x FLAG tag (Fig 5B). The start codon of this FLAG tag is deleted, so any FLAG signal is the result of translation from the lncRNA-smORF. Two candidate lncRNA-smORFs were selected that did not have mass spectrometry support; LINC000478 and LINC01116. A 37 codon smORF was detected as translated from LINC00478 in both conditions (Fig 5C). Transfection of LINC00478-smORF-FLAG into undifferentiated SH-SY5Y cells produced FLAG signal in both nuclear and cytoplasmic compartments (Fig 5D & Sup 6A). FLAG signal was also seen when LINC00478-smORF-FLAG transfected SH-SY5Y cells were treated with RA (Sup 6A). Interestingly this RA FLAG signal was only ever detected in the nucleus. Similar results were seen in HEK293 cells (Sup 6B), but because of the higher transfectional efficiency in HEK293 compared with SH-SY5Y cells, we detect FLAG signal in more cells.

Poly-Ribo-Seq detected increased levels of LINC01116 during differentiation (4-fold in Control vs RA polysome) and translation of a 71-codon smORF during differentiation (Fig 4I). Tagging of LINC01116-smORF (Fig 5E) generated FLAG signal in the cytoplasm of SH-SY5Y cells, which is localized to neurites (Fig 5F), suggesting it could play a role in differentiation. FLAG signal was also present in

LINC01116 transfections in HEK293 cells (Sup 6C), but because of the higher transfectional efficiency in HEK293 compared with SH-SY5Y cells, we detect FLAG signal in more cells.

LINC01116-smORF has two potential ATG start codons (Fig 5E and Sup 6D). When these start codons were assessed for similarity to the Kozak sequence consensus, using NetStart1.0 (Pedersen & Nielsen, 1997), both exhibited scores >0.5, indicating both are in good context and therefore either could be used to initiate translation ($AUG_1 = 0.545$, $AUG_2 = 0.645$). Given the scanning model of translation initiation it seems likely that the first AUG would be used. To determine if the first start codon is actually used it was mutated (Fig 5E). No FLAG signal was present in transfections where the 5' start codon was mutated ($\Delta 1$) (Fig 5G). This suggests that the first start codon is necessary for the translation of the LINC01116-smORF and the resulting peptide is 87aa long. Although FLAG signal is present in a low number of cells, no transfection controls and $\Delta 1$ indicate that FLAG signal is dependent on translation of the LINC01116-smORF (Sup 6E).

Translated lncRNA-smORFs exhibit sequence conservation

Another indicator of coding potential and also of peptide functionality is the sequence conservation of our translated lncRNA-smORFs. To ensure detection of sequence conservation for these short smORFs irrespective of annotation in other genomes, we used three complementary BLAST strategies using transcript nt sequence, smORF nt sequence and protein aa sequence.

Initial searches using the entire lncRNA transcript sequence (nt) and BLASTn (Altschul et al., 1990), returned results for ~78% of translated lncRNAs, many of which had short alignment lengths of 30-100 nt. Although some of these results may represent conservation of the smORFs, many are due to small areas of sequence overlap along the rest of the lncRNA. This is a well-documented issue when investigating lncRNAs; they rarely exhibit the same levels of conservation as mRNAs (Johnsson et al., 2014). A more common finding is that there are “modules” of higher sequence conservation across lncRNA transcripts, as described for Xist lncRNA (Brockdorff, 2018). To accommodate this, a second round of searches were performed, on the initial search results, using the nt sequence of the smORF (BLASTn) (Altschul et al., 1990). This filtering step resulted in fewer, higher quality results, with 14 of 18 remaining lncRNA-smORFs passing manual cross validation. For the majority of these lncRNA-smORFs, conservation is high across the body of the smORF, and drops off across the rest of the transcript. The ENST00000442428.1 (AL162386.2)-smORF exhibits high sequence conservation when compared to gorilla and orangutan (*Pongo abelii*), with 100% and 99% ORF nt sequence identity respectively (Figure 5H). When entire transcripts are aligned, this percentage sequence identity drops to 74% with gorilla (ENSGGOT00000060708.1), and 65% with orangutan (ENSPPYT00000022401.2), indicating the ORF is the most conserved part of these transcripts.

To further corroborate these results, a tBLASTn (Altschul et al., 1990) search of the lncRNA-smORF aa sequences was performed. By translating transcript databases into aa in all 6 frames, tBLASTn removes the noise of synonymous substitutions, which can have a significant effect, particularly in smORFs (Ladoukakis et al., 2011). For the majority of smORFs, the same results were returned, and evidence of conservation was found for a further three lncRNA-smORFs (ENST00000454935.1_477_633, ENST00000557660.5_42_186, ENST00000453910.5_151_262) that appear to have undergone some frameshift mutations.

At the aa sequence level, ~18% of the translated lncRNA-smORFs returned homologous proteins from BLASTp searches (Altschul et al., 1990). All the returned proteins were unreviewed, uncharacterised proteins, with no evidence at protein, transcript or homology levels in the Uniprot database (Bateman et al., 2019). This potentially suggests that automated annotation pipelines recognised the coding potential of these smORFs, unlike in the more curated human genome annotation. Overall, the combination of these 3 layers of sequence conservation analysis reveal that 11 of our translated lncRNA-smORFs exhibit sequence conservation across *Hominidae*, 3 additional smORFs are also found in gibbons (Fig 5I), with evidence for 2 translated smORFs detectable at the greater evolutionary distance of human to mouse, but not found in all apes (Sup 6E).

LINC01116 contributes to neuronal differentiation

To dissect the potential role of LINC01116 and its translation during differentiation we knocked down LINC01116 using a siRNA pool in both undifferentiated and differentiated SH-SY5Y cells (Fig 6A). LINC01116 knockdown initially had a limited effect on cell viability, which recovered to no effect after 2 days (Sup 7A). Interestingly, LINC01116 knockdown resulted in a significant reduction of neurite length in RA treated SH-SY5Y cells (Fig 6B, zoom in Fig 6C), compared to scrambled siRNA treated SH-SY5Y (Fig 6D). There was no effect of the knockdown in undifferentiated cells (Fig 6D). This suggests that LINC01116 is involved in the regulation of neuritic processes formation. To examine potential effects of LINC01116 knockdown further on differentiation we assessed the expression levels of the noradrenergic marker MOXD1, which is important in neural crest development. LINC01116 siRNA knockdown, upon differentiation resulted in a reduction of MOXD1 expression levels, further indicating a role of LINC01116 in neuronal differentiation (Fig 6E). However, LINC01116 knockdown had no effect on proliferation, as measured by % of Ki67+ cells (Sup 7B & Fig 6F) or cell cycle, as measured by E2F1 mRNA RT-qPCR (Fig 6G). LINC01116 likely functions early in the differentiation pathway since its levels are significantly upregulated within the first 24 hours of RA induced differentiation (Sup 7C). Expression of LINC01116 then declines rapidly by day 8 (Fig 7D). Together these results suggest that LINC01116 functions during early differentiation, contributing to neurite process formation.

Discussion

Cytoplasmic lncRNAs have recently emerged as important regulators of several cellular processes, including the regulation of translation, in a variety of tissues (Carrieri et al., 2012; Dimartino et al., 2018). Moreover, the active translation of smORFs from within lncRNAs has been characterized in a range of organisms (Anderson et al., 2015; Aspden et al., 2014; Chen et al., 2020; Magny et al., 2013; Pueyo & Couso, 2008; Wang et al., 2020). Translational regulation is a key process during neuronal differentiation (Blair et al., 2017; Chau et al., 2018), therefore we reasoned that lncRNAs may functionally interact with polysomes during differentiation.

Having established that early neuronal differentiation of human neuroblastoma cells (SH-SY5Y) results in reduced global translation, we sought to further investigate the relationship of lncRNAs with the translation machinery during differentiation. We detected the presence of ~800-900 lncRNA genes expressed and their transcripts present in the cytoplasm. 85-90% of these cytoplasmic lncRNAs are associated with polysome complexes, suggesting that they are either being translated, or regulating the translation of the mRNAs with which they interact. Moreover, the association of lncRNAs with polysomes is dynamic during differentiation, as shown by the differential polysome enrichment of lncRNAs in Control and RA treated cells. The lncRNAs that dynamically interact with polysomes upon neuronal differentiation mainly belong to the 'intergenic' and 'anti-sense' categories of lncRNAs. These results reveal that many lncRNAs are present in the cytoplasm, enriched there, and are associated with translation complexes.

An example of a polysome-associated lncRNA that we characterized in more detail is LINC02143. It is an intergenic lncRNA with no known function, which we find induced upon differentiation. It is detected in 80S and small polysome fractions, indicating it interacts with the translation machinery. A number of anti-sense polysome-associated lncRNAs appear to be upregulated upon differentiation. Amongst them is DLGAP1-AS1, which is anti-sense to Disks large-associated protein 1 (DLGAP1), itself involved in chemical synaptic transmission. DLGAP1-AS1 interacts with actively translating polysomes both in Control and upon differentiation. The lncRNAs depleted from the polysomes have fewer anti-sense lncRNAs relative to other populations, suggesting that anti-sense lncRNAs are preferentially localised to polysomes. These polysome associated anti-sense lncRNAs could potentially regulate the translation of their 'sense' mRNA, through base-pairing, as is the case with BACE1-AS (Faghihi et al., 2010) and UCHL1-AS (Carrieri et al., 2012). This contrasts with the general view that anti-sense lncRNAs regulate gene expression in the nucleus at the point of transcription or splicing (Beckedorff et al., 2013).

Differential translation analysis revealed that upon RA induced neuronal differentiation of SH-SY5Y for 3 days, a small number of ORFs are exclusively translationally regulated. For the majority of ORFs, translation regulation is driven by transcription regulation. This result indicates that at this early timepoint, transcriptional control, rather than translational control is the driving force of SH-SY5Y differentiation. This is not surprising, because according to a previous study of human embryonic stem cells (hESCs) differentiation into neural progenitor cells (NPCs) only 12% of the detected transcripts were exclusively translationally regulated (Blair et al., 2017).

Ribosome profiling of the actively translating polysomes allowed us to distinguish between the lncRNAs that associate with the polysome complexes and those that are being actively translated. We were able to identify 45 translated lncRNA-smORFs in undifferentiated and differentiated cells. Only ~5% of polysome-associated lncRNAs pass our stringent thresholds for translation. The translated lncRNA-smORFs exhibit high levels of triplet periodicity and their translational efficiencies are similar to those of protein-coding genes. We can therefore be confident that these are real translation events leading to the production of substantial peptide levels rather than background, spurious events (Bazzini et al., 2014; Guttman et al., 2013; Patraquim et al., 2020; Ruiz-Orera & Alba, 2019). The size distribution of our novel translated ORFs, indicates that the majority are indeed smORFs (<100aa). The general pattern we identified is that dORFs>lncRNA-smORF>uORFs in size. This is consistent with previous studies where a wide range of peptides lengths were discovered (Aspden et al., 2014; Chong et al., 2020). Amino acid composition of these translated smORFs supports the fact they are translated into peptides. However, it does not suggest they are enriched for transmembrane alpha-helices. This is in contrast to the smORFs characterized in *D. melanogaster* (Aspden, 2014).

Overall, we have independent evidence for peptide synthesis for 12/45 lncRNA-smORFs. 8 of these from published mass spectrometry data from Control and RA-differentiated cells SH-SY5Y cells (Brenig et al., 2020). In general, we find smORFs translated in the same treatment as these mass spectrometry datasets detect peptides (7/8). An 18% mass spectrometry detection level may seem low but given the limitations of detecting small peptides by mass spectrometry this represents a substantial level of validation. Two translation events were validated by FLAG tagging transfection assay: LINC01116 and LINC00478 lncRNA-smORFs. The production of 2/45 lncRNA-smORF peptides is corroborated by previous studies in non-neuronal cells. HAND2-AS1 (translated in Control and RA) is translated in human and rodent heart and encodes for an integral membrane component of the endoplasmic reticulum (van Heesch et al., 2019). CRNDE, which is only translated upon differentiation, encodes for a previously characterised nuclear peptide (CRNDEP) (Szafron et al., 2015). The translation of these smORFs in multiple cell types provides substantial support for the production of peptides and their potential function.

We also discovered that 24% of the lncRNA-smORFs we find translated show sequence conservation across the *Hominidae*. This suggests that the other great apes have the potential to translate very similar peptides. This provides additional evidence to indicate that these translation events are not translational noise. Of course, it will be interesting to uncover the function of these small peptides in the future.

Here we have discovered that LINC01116 produces an 87aa peptide that exhibits cytoplasmic localisation, and specifically is detected in near the cell membrane and in neuritic processes. The upregulation of LINC01116 expression upon differentiation as well as the existing evidence of its expression in the developing and adult brain and spinal cord (Consortium, 2013; Lindsay et al., 2016), coupled with the localisation of its peptide, led us to further investigate its potential role in differentiation. Knockdown of LINC01116 upon differentiation appears to impede neurite outgrowth and results in the reduction of the mRNA levels of the noradrenergic marker MOXD1. Our data suggest that LINC01116 is involved in the regulation of neuronal differentiation, consistent with the fact that it is moderately expressed in the developing human forebrain and highly expressed in the developing human midbrain and spinal cord (Lindsay et al., 2016). LINC01116 was previously found to be involved two other cancer models; in the progression of glioblastoma (Brodie et al., 2017) and it is upregulated in gefitinib resistant non-small cell lung cancer cells (Wang He et al., 2020). siRNA knockdown of LINC01116 in both these cell types results in decreased expression of stem-cell markers (Nanog, SOX2 and Oct4) and reduced cell proliferation. This suggests LINC01116 promotes cell proliferation in these systems, indicating that the downstream effects of LINC01116 may vary according to cell type. Interestingly however, knockdown of LINC01116 also inhibited migration of glioma stem cells (GSCs) (Brodie et al., 2017), while overexpression of LINC01116 promoted invasion and migration of gastric cancer cells (Su et al., 2019). This suggests a potential role of LINC01116 in the formation of cell membrane protrusions, which is consistent with the role we have discovered for LINC01116 in neurite development.

Our findings indicate that many lncRNAs are localised in the cytoplasm and play important roles here. Given the large number of lncRNAs associated with polysomes we anticipate that many more lncRNAs will be found to regulate translation. We have identified 121 novel translation events, many of which are regulated during differentiation. Only ~5% of polysome-associated lncRNAs are translated. The lncRNA-smORFs we discover here represent a general population whose products have not yet been characterized. As we have found for LINC01116, lncRNAs and the small peptides encoded therein have the potential to contribute to important cellular functions and contribute to development and disease.

Materials and Methods

Cell culture

Human neuroblastoma SH-SY5Y cells, were cultured in Dulbecco's Modified Eagle Medium (DMEM 4.5g/L Glucose with L-Glutamine) supplemented with 1% (v/v) Penicillin/Streptomycin and 10% Fetal Bovine Serum (FBS) at 37°C, 5% CO₂. Neural induction commenced at passage 4 and was performed as described previously (Forster et al., 2016; Korecka et al., 2013) with minor alterations. All trans Retinoic Acid (RA, Sigma) was added to cells 24h after plating, at final concentration of 30µM.

Immunocytochemistry

Cells were seeded on Poly-D-Lysine/mouse laminin coated 12mm round coverslips (Corning BioCoat™ Cellware) and fixed with 4% paraformaldehyde (PFA) (Affymetrix) for 20 min at room temperature (RT). A permeabilization step (0.1% Triton-X for 10 min at RT) was performed prior to blocking, followed blocking at RT in blocking buffer (3% BSA, 1XPBS or 5% NGS, 1XPBS and 0.1% Triton-X) for 30 min. Primary antibodies (Antibody Table) were applied in (3% BSA 1X PBS or 0.5% NGS, 1X PBS, 0.1% Triton-X and incubated at RT for 2h or at 4° C overnight. Cells were washed and labelled with Alexa 488, Alexa 555, or Alexa 633 at 1:500 dilution for 2h at RT in 0.5% NGS, 1X PBS, 0.1% Triton-X. Cells were mounted in VECTASHIELD mounting medium, analysed using LSM 700 confocal microscope (Zeiss) ImageJ.

cDNA synthesis and quantitative Real Time PCR (RT-qPCR)

Equal amounts of RNA (whole cell, nuclear and cytoplasmic lysates) or equal volumes (polysome fractions) were subject to cDNA synthesis, using qScript (Quanta Bio) according to manufacturer's instructions. qPCR was performed using the CFX Connect™ thermal cycler and quantification using SYBR Green fluorescent dye (PowerUp™ SYBR Green Master Mix, Thermo Fisher SCIENTIFIC). Primers were designed to anneal to exon-exon junctions, where possible, or to common exons between alternative transcripts (Primer Table). Target mRNA and lncRNA levels were assessed by absolute quantification, by the means of standard curve or relative quantification, using the $\Delta\Delta C_q$ method.

Polysome Profiling

RA was added to SH-SY5Y cells 3 days prior to harvesting. Cells were treated with cycloheximide (Sigma) at 100µg/ml for 3 min at 37°C, washed (1X PBS, 100µg/ml cycloheximide) and trypsinised for 5 min at 37°C. Subsequently, cells were pelleted, washed (1X PBS, 100µg/ml cycloheximide), and resuspended in ice cold lysis buffer; 50mM Tris-HCl pH8, 150mM NaCl, 10mM

MgCl₂, 1mM DTT, 1% IGEPAL, 100µg/ml cycloheximide, Turbo DNase 24U/µL (Invitrogen), RNasin Plus RNase Inhibitor 90U (Promega), cOmplete Protease Inhibitor (Roche), for 45 min. Cells were then subjected to centrifugation at 17,000 xg for 5 min, to pellet nuclei. Cytoplasmic lysates were loaded onto 18%-60% sucrose gradients (~70 x10⁶ cells per gradient) at 4°C and subjected to ultracentrifugation (121,355 x g_{avg} 3.5h, 4°C) in SW-40 rotor. Gradients were fractionated using GRADIENT STATION (Biocomp) and absorbance at 254 nm was monitored using a Biorad detector.

Poly-Ribo-Seq

~20% of cytoplasmic lysate was kept for PolyA selection (Total RNA control) and ~80% was loaded onto 18%- 60% sucrose gradients (~70x10⁶ cells per gradient) at 4°C and subjected to ultracentrifugation (121,355 x g_{avg} 3.5h, 4°C) in SW-40 rotor. Polysome fractions were pooled from control and from differentiated cells. ~25% polysomes were kept for polyA selection (Polysome-associated RNA). The remaining 75% was diluted in 100mM Tris-HCl pH8, 30mM NaCl, 10mM MgCl₂. RNaseI (EN601, 10U/µl 0.7-1U/million cells) was subsequently added incubated overnight at 4°C. RNaseI was deactivated using SUPERase inhibitor (200U/gradient) for 5min at 4°C. Samples were concentrated using 30 kDa molecular weight cut-off columns (Merck) and loaded on sucrose cushion (1M sucrose, 50mM Tris-HCl pH8, 150mM NaCl, 10mM MgCl₂, 40U RNase Inhibitor) and subjected to ultracentrifugation at 204,428 x g_{avg} at 4°C for 4h (TLA110). Pellets were resuspended in TRIzol (Ambion, Life Technologies) and processed for RNA purification.

RNA purification from cytoplasmic lysates and RNaseI footprinted samples was performed by Trizol RNA extraction, following manufacturer's instructions. RNA purification from polysome fractions was performed by isopropanol precipitation, followed by TURBO DNase treatment (ThermoFisher) (according to manufacturer's instructions), acidic phenol/chloroform RNA purification and ethanol precipitation at -80°C overnight. RNA concentration was determined by Nano-drop 2000 software. Two rounds of polyA selection from total cytoplasmic lysate and polysome fractions were performed using oligo (dT) Dynabeads (Invitrogen) according to manufacturer's instructions. PolyA RNA was fragmented by alkaline hydrolysis. 28–34 nt ribosome footprints and 50–80 nt mRNA fragments were gel purified in 10% (w/v) polyacrylamide-TBE-Urea gel at 300V for 3.5h in 1X TBE. Ribosome footprints were subjected to rRNA depletion (Illumina RiboZero rRNA removal kit).

5' stranded libraries were constructed using NEB Next Multiplex Small RNA Library Prep. Resulting cDNA was PCR amplified and gel purified prior to sequencing. Libraries were subjected to 75bp single end RNA Seq using NextSeq500 Illumina sequencer, High Output Kit v2.5 (75 Cycles) (Next Generation Sequencing Facility, Faculty of Medicine, University of Leeds).

RNA-Seq data analysis

RNA-Seq reads were trimmed with Cutadapt (Martin, 2011) and filtered with fastq_quality_filter (Hannon, 2010) to filter out the reads of low quality (90% of the read to have a phred score above 20). Filtered reads were mapped (Liao et al., 2013) to human genome reference (the lncRNA GENCODE (Frankish et al., 2019) annotation added to mRNA annotation from UCSC (Haeussler et al., 2019) human genome assembly (hg19) from iGenomes) with Rsubread (Liao et al., 2013) and uniquely mapped reads were reported. Bam file sorting and indexing was performed with SAMtools (Li et al., 2009). Subsequently summarised read counts for all genes were calculated using featureCounts (Liao et al., 2014). For normalization, RPKM values were calculated. Differential expression analysis was conducted between with DESeq2 (Love et al., 2014) based on the two cut-offs $p^{adj} < 0.05$ and the absolute value of $\log^2\text{FoldChange} > 1$. Gene ontology analysis was performed with GOrilla (*Gene Ontology enrichment analysis and visualization tool*) (Eden et al., 2009).

Cytoplasmic/Nuclear fractionation of SH-SY5Y cells.

Cells were harvested (as above) and washed with 1X PBS. Cells were lysed in whole cell lysis buffer (see table) (500 μ L buffer per 10^6 cells) on ice for 30min. Whole cell lysate aliquots were removed and remainder subjected to centrifugation at 1,600 x g for 8 min to pellet nuclei. Nuclear and cytoplasmic fractions were subjected to two further clearing steps by centrifugation (3,000 x g and 10,000 x g respectively). Nuclei were lysed in RIPA buffer (table). ~10% of both nuclear and cytoplasmic lysates were used for Western Blot and ~90% subjected to RNA extraction (ZYMO R1055).

Western Blot

Samples were diluted in 4X Laemmli sample buffer (Biorad) (277.8 mM Tris-HCl, pH 6.8, 4.4% LDS, 44.4% (v/v) glycerol, 0.02% bromophenol blue), 5% β -mercaptoethanol (Sigma) was added prior to heating at 95°C for 5 min and loaded on 10% SDS gels. Gel electrophoresis was performed using the BioRad Mini-PROTEAN 3 gel electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA). Proteins were transferred to nitrocellulose membranes (AmershamTM ProtranTM) and blocked with 5% fat-free milk powder in 1XPBS, 0.05% Tween-20 (Sigma) for 1h at RT. Blots were incubated with primary antibodies overnight (Antibody Table). Blots were then washed in PBS-T and incubated with secondary antibody (anti-mouse HRP) at RT for 2h. Membranes were washed 3 times with PBS-T, prior to application of ECL (Biological Industries). Chemiluminescent signal was detected with Chemi-Doc (BIO-RAD). All membranes were probed for β -tubulin as loading control.

Ribo-Seq analysis

Quality reports of Polysome-associated RNA-Seq and Ribo-Seq data were made using Fastqc (v.0.11.8) (Andrews, 2010). Adaptor sequences were trimmed using Cutadapt (v.1.81) (Martin, 2011) with minimum read length of 25bp, and untrimmed outputs retained for Ribo-Seq reads. Low-quality reads (score <20 for 10% or more of read) were then discarded using FASTQ Quality Filter, FASTX-Toolkit (v.0.0.14) (Gordon, 2010). Human rRNA sequences were retrieved from RiboGalaxy (Michel et al., 2016) and high confidence hg38 tRNA sequences from GtRNAdb Release 17 (Chan & Lowe, 2016). One base was removed from 3' end of reads to improve alignment quality, reads originating from rRNA and tRNA were aligned and removed using Bowtie2 (v.2.3.4.3) (Langmead & Salzberg, 2012).

The splice aware aligner STAR (v2.6.1b) (Dobin et al., 2012) was used to map remaining reads to human reference genome Release 30 (GRCh38.p12), from Gencode (Frankish et al., 2019). The STAR (v2.6.1b) (Dobin et al., 2012) genome index was built with a sjdbOverhang of 73. Samtools (v.1.9) (Li et al., 2009) was used to create sorted, indexed bam files of the resulting alignments.

Metaplots of aligned Ribo-seq data were generated using create_metaplots.bash script from Ribotaper (v1.3) (Calviello et al., 2016) pipeline. These show the distance between the 5' ends of Ribo-seq and annotated start and stop codons from CCDS ORFs, allowing the locations of P-sites to be inferred. Read lengths exhibiting the best triplet periodicity were selected for each replicate, along with appropriate offsets (Sup 5).

Actively translated smORFs were then identified using Ribotaper (v1.3) (Calviello et al., 2016). Initially, this requires an exon to contain more than 5 P-sites in order to pass to quality control steps. Identified ORFs were then required to have a 3-nt periodic pattern of Ribo-seq reads, with 50% or more of the P-sites in-frame. In the case of multiple start codons, the most upstream in-frame start codon with a minimum of five P-sites in between it and the next ATG was selected.

ORFs for which >30% of the Ribo-seq coverage was only supported by multimapping reads were also subsequently filtered. For a smORF to be considered actively translated in a condition, we required that it be identified in at least two of the three biological replicates for the condition.

Translation efficiency calculation

Translational efficiency (TE) was estimated for all translated ORFs in each condition, where TE was equal to the mean number of P sites per ORF, normalised by the median P sites per ORF per replicate, divided by the mean number of RNA sites per ORF, normalised by the median RNA sites per ORF per replicate.

Differential translation efficiency analysis

Differential translation efficiency analysis was performed using the Δ TE (Chothani et al., 2019) alternate protocol. However, our analysis was based on read counts from all identified translated ORFs from the Ribo-seq and polysome-associated RNA-seq, not the whole CDS as described in the protocol.

Amino acid composition

For each of our ORF sets (protein coding, lncRNA-smORF, uORF and dORFs), the average amino acid compositions were calculated. Random control expected frequencies were taken from King and Jukes (King & Jukes, 1969).

Mass spectrometry analysis

Two published SH-SY5Y cell mass proteomics datasets were analysed: PXD010776 (Murillo et al., 2018) and PXD014381 (Brenig et al., 2020). Binary raw files (*.raw) were downloaded from PRIDE then converted to human-readable MGF format using ThermoRawFileParser (Hulstaert et al., 2020). The amino acid sequences of our translated uORFs, dORFs and lncRNA-smORFs were added to the whole *Homo sapiens* proteome dataset (20,379 entries) downloaded from UniProtKB (Bateman et al., 2019) on Nov/2019. The new FASTA file was then used as a customized database on Comet (v2019.01.2) (Eng et al., 2013) search engine runs that scanned all MS/MS files (*.mgf) against it.

Default settings in Comet were used with the following exceptions according to the MS/MS data type. iTRAQ-4plex (PXD010776): decoy_search = 1, peptide_mass_tolerance= 10.00, fragment_bin_tol = 0.1, fragment_bin_offset = 0.0, theoretical_fragment_ions = 0, spectrum_batch_size = 15000, clear_mz_range = 113.5-117.5, add_Nterm_peptide = 144.10253, add_K_lysine = 144.10253, minimum_peaks = 8. Label-free (PXD014381): decoy_search = 1, peptide_mass_tolerance = 10.00, fragment_bin_tol = 0.02, fragment_bin_offset = 0.0, theoretical_fragment_ions = 0, spectrum_batch_size = 15000.

CometUI (Eng et al., 2013) was employed for assessing Comet's *.xml output files and setting a false discovery rate (FDR) threshold of 10% per peptide identification. This FDR threshold was selected due to expected low abundance levels of the target smORFs.

Conservation analysis

Protein, cDNA and ncRNA sequence data for *H.sapiens*, *P.abelii*, *P.paniscus*, *P.troglodytes*, *G.gorilla*, *N.leucogenys*, and *M.musculus* were obtained from Ensembl (release 100, (Yates et al., 2020)).

A sequence homology search was performed using the 45 translated lncRNA peptide sequences, against each non-human species protein database using BLASTp (e value - 0.001) (Altschul

et al., 1990). As these are small peptides, results were filtered to remove anything with <75% identity, unless a result(s) was the lowest e value hit for a given query in each species. Results were returned for 12 lncRNA peptides, and manually cross validated using Ensembl Genome Browser and multiple sequence alignments performed using ClustalOmega (Sievers et al., 2011), run using default parameters using the msa package in R (Bodenhofer et al., 2015), to give 8 peptides with evidence of sequence conservation.

The transcript sequences of the 45 translated lncRNAs were searched against transcriptome databases created by combining the cDNA and ncRNA data for each species, using BLASTn (e value - 0.001) (Altschul et al., 1990). Results of this BLAST were used to filter the initial BLAST databases. ORF portions of the 45 translated lncRNAs were extracted, and searched against these filtered databases using BLASTn (e value - 0.001)(Altschul et al., 1990).

It was confirmed that all the lncRNA ORFs returned their genes of origin in *H.sapiens*. The remaining species returned results for 18 of the lncRNA ORF queries. These were cross validated as above, resulting in 14 lncRNA ORFs with evidence of sequence conservation based on transcript sequences.

A final search was performed using the 45 translated lncRNA peptide sequences against the transcriptome databases using tBLASTn (e value - 0.001) (Altschul et al., 1990). To reduce redundancy, results were filtered to select the transcripts(s) with lowest e value for each gene. It was confirmed that all ORFs returned their genes of origin in *H.sapiens*.

The remaining species returned results for 21 of the lncRNA peptide queries. As some queries had many spurious results, they were further filtered to select the transcripts(s) with lowest e value for each query in each species. These were cross validated as above, resulting in 16 lncRNA peptides with evidence of sequence conservation based on transcript sequences.

We combined evidence from all these approaches into a final dataset consisting of 17 lncRNA small ORFs with evidence of conservation.

Plasmid construction/Cloning

5'-UTRs and CDSs of putative smORFs (lacking stop codon) were generated by PCR (Table 3), using NEB High Fidelity DNA Polymerase (Q5). C-terminal 3xFLAG tag was incorporated within the reverse primer (Table 3 marked with green) by PCR and products were cloned into NheI and EcoRV restriction sites (Table 3: marked with blue) of pcDNA3.1-hygro vector (Addgene, kindly provided by Mark Richards-Bayliss group, University of Leeds). Start codon mutations were generated by site directed mutagenesis (Q5 Site directed mutagenesis kit, NEB).

Transfections and Microscopy

Plasmid transfections were performed using Lipofectamine 3000 (Thermo) following the manufacturer's instructions. After 48 h, the cells were fixed for 20 min with 4% paraformaldehyde, washed with 1X PBS, 0.1% Triton X-100 (PBS-T) and processed for immunocytochemistry as previously described. Imaging was conducted using EVOS fluorescent microscope.

siRNA knockdown

siRNA knockdown was performed using Lincode siRNA SMARTpool (Dharmacon) (LINC01116 transcript: R-027999-00-0005 SMARTpool). Lincode Non-targeting Pool (D-001810-10) was used as scrambled control. Cells were seeded in 24-well plates (10^5 cells/well,) and siRNA where transfected using RNAiMAX lipofectamine (ThermoFisher) as per manufacturer's instructions.

General statistics and plots

Statistical analyses were performed in R (R Core Team, 2019), using packages including stringr (Wickham, 2019), dplyr (Wickham, 2017), tidyr (Wickham, 2017), protr (Xiao et al., 2015) ggplot (Wickham, 2016), ggstatsplot (Patil, 2018), knitr (Xie, 2020), seqinr (Charif D, 2007), ggbeeswarm (Clarke & Sherrill-Mix, 2017), and EnhancedVolcano (Blighe et al., 2018)

Experimental values from independent samples with equal variances, were assessed using 2-tailed unpaired Student's t-test. The results are shown as mean $\pm$ SEM values of 3 independent replicates. The exact P values are described and specified in each figure legend. P values < 0.05 were considered statistically significant.

Data availability

Poly-Ribo-Seq datasets will be deposited in GEO.

Tables

Table 1: Buffers

Polysome lysis buffer (1mL/70 million cells)	
--	--

Component	final concentration
Tris-HCl pH8	50mM
NaCl	150mM
MgCl ₂	10mM
DTT	1mM
IGEPAL	1%
cycloheximide	100µg/mL
Turbo DNase	24U/mL
RNase Inhibitor (RNaseInPlus)	90U
cOmplete Protease Inhibitor (Roche)	0.33%
ddH ₂ O	
RNaseI footprinting buffer	
Component	final concentration
Tris-HCl pH8	100mM
NaCl	30mM
MgCl ₂	10mM
RNaseI (EN0601-Thermo)	0.8-1U/million cells
SuperRNase Inhibitor (Ambion)	3U/million cells
Cytoplasmic/Nuclear fractionation buffer	
Component	final concentration
Phosphate Buffer Saline (PBS)	1X
Triton-X	1%
RNase Inhibitor (RNaseInPlus)	40U/mL
ddH ₂ O	
Nuclei lysis buffer-RIPA based	
Component	final concentration
NaCl	150mM
IGEPAL	1%

Sodium Deoxycholate (DOC)	0.5%
SDS	0.1%
Tris-HCl pH 7.4	25mM
cOmplete Protease Inhibitor (Roche)	1X
ddH ₂ O	

677

678

679

680 Table 2: qPCR primers

qPCR		
gene/transcript	Forward primer sequence	Reverse primer sequence
MOXD1	GGAAGCCGAAAAGCCAAGTG	TCGAAAATGACGCAGCCTGA
NTN4	CGAGTGCAGAACCTGCAAGTGT	CATCTGGAGCTGAGAAGGGTC
E2F1	TGGAGCAAGAACCGCTGTTGT	GGGAAAGGCTGATGAACTCCT
SOX2	ACATGAACGGCTGGAGCAA	GTAGGACATGCTGTAGGTGGG
GAPDH	CATCCTGGGCTACACTGAGC	GTCAAAGGTGGAGGAGTGGG
LINC01116	TCTAAGAATGGGTCTCACTCTGC	CCAGGCATGGTGGCTCAC
LINC02143	AACCTTTGCAGTAGCTCCTGG	GGATGAGGAGACTGAGACTGAGAG
AC254633.1	GTGACTCACCTCCCAGACTTC	TGCTGTGCAGCCAGCGTC
DLGAP1-AS1	TCTGAGAGCCAGCGAACTTT	AGCCTGTTGCGTCATGTGAT
DLGAP1-AS2	CCCAGGACACAGACAAGACC	ATGCACGCTCTCTGACAGCA
SERPINB9P1	AGTCAGCGAGTGGACAAAGC	GACTCCATGCTGCGGTTTTTC
AC090001.1	GTGCCCATGAGGGAGAACAC	GACAAGAAGTCAGGAGGTAGACA
SNAP25-AS1	AGCCATGGAAGTCAAATGCTG	AGGCATTTTTGCTGTCTTTCCTC
XIST	GGCTCCTCTTGACATTCTGAG	AGCTTGGCCAGATTCTCAAAG

681

682

683

684

685 Table 3: Cloning Primers

Transcript	Forward primer sequence	Reverse primer sequence
LINC001116 -smORF	GGCGGTGCTAGCGCGAGCCACGGGCCTC	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGTTTTTAAGCTGACTTGT C
LINC01116- whole	GGCGGTGCTAGCGCGAGCCACGGGCCTC	ACCGCCGATATCCAATTACTGCATTCACGTATTCTTC
LINC00478	GGCGGTGCTAGCCTCCTGTCGTTAAGATAAATTCTCCA	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGTAGTAAATGCTCTCT GA
LINC01116- mutant (ATG1)	CTTCTAAGAAAAGGTCTCACTCTGC	CAATTCAGTTGTCTTCTAATAC
LINC01116- mutant (ATG2)	GGCACCATCAAAGCTCACTGCAGC	ACCACACTCCAGCCTGGG
LINC01116- mutant (both ATG)	GCTGGAGTGTGGTGGCACCATCAAAGCTCACTGCAGCCTTGA A	CTGGGTGATGGCAGAGTGAGACCTTTTCTTAGAAGCAATTCAGTTGTC

686

687

688

689

690

691

692

693

694

695

696

Table 4: Antibodies

Protein	Species	Company	Usage (ICC/WB)	Dilution
betaIII-tubulin (TuJ1)	rabbit	ProteinTech	ICC	1:50
ki67	mouse	Dako-Agilent	ICC	1:100
c-Fos	rabbit	Santa Cruz	ICC	1:50
Tubulin B	mouse	DSHB	both	1:5000 for WB, 1:200 for ICC
NXF1	mouse	abcam	WB	1:5000
H3K27me3	mouse	abcam	WB	1:1000
FLAG	mouse	Sigma	both	1:1000
hnRNPK	rabbit	abcam	ICC	1:100
anti-mouse IgG-HRP	goat	New England Biolabs	WB	1:5000
anti-rabbit IgG-HRP	goat	New England Biolabs	WB	1:5000
Alexa anti-mouse IgG-488	goat	Thermofisher	ICC	1:500
Alexa anti-rabbit IgG-488	goat	Thermofisher	ICC	1:500
Alexa anti-rabbit 633	goat	Thermofisher	ICC	1:500

Acknowledgements

We wish to thank Dr Eric Hewitt for providing the SH-SY5Y cells; Dr Iosifina Sampson and Dr Mark Richards (Bayliss group) for providing HEK293 cells and mammalian vectors. We thank the Next Generation Sequencing facility, at St James University Hospital, Leeds, UK for performing Next Generation Sequencing and Imaging Facility, Faculty of Biological Sciences, University of Leeds, UK for their assistance in confocal microscopy. Parts of this work were undertaken on ARC3, part of the High Performance Computing facilities at the University of Leeds, UK.

Author contributions

KD designed and performed experiments, acquired, analysed and interpreted data, drafted and revised the manuscript. IB designed and performed experiments, acquired, analysed and interpreted data, drafted and revised the manuscript. DW analysed, interpreted data and revised the manuscript. SC acquired and analysed data, and revised the manuscript. AB acquired and analysed data, and revised the manuscript. EJRV analysed data and revised the manuscript. MOC helped interpret portions of the data and critique the output for important intellectual content, and revised

the manuscript. JD helped design experiments, interpret portions of the data and critique the output for important intellectual content, and revised the manuscript. AW helped design experiments, interpret portions of the data and critique the output for important intellectual content, and revised the manuscript. JLA conceived the work, interpreted data, drafted and revised the manuscript.

Funding

Katerina Douka's PhD was funded by Leeds Anniversary Research Scholarship (LARS). Julie Aspden is funded by the University of Leeds (University Academic Fellow scheme). This work was funded by MRC (MR/N000471/1).

References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). BASIC LOCAL ALIGNMENT SEARCH TOOL [Article]. *Journal of Molecular Biology*, 215(3), 403-410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
- Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., & Olson, E. N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, 160(4), 595-606.
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. In (Version v.0.11.8)
- Aspden, J. L., Eyre-Walker, Y. C., Phillips, R. J., Amin, U., Mumtaz, M. A., Brocard, M., & Couso, J. P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *Elife*, 3, e03528. <https://doi.org/10.7554/eLife.03528>
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Alpi, E., Bely, B., Bingley, M., Britto, R., Bursteinas, B., Busiello, G., Bye-A-Jee, H., Da Silva, A., De Giorgi, M., Dogan, T., Castro, L. G., Garmiri, P., Georgiou, G., Gonzales, D., Gonzales, L., Hatton-Ellis, E., Ignatchenko, A., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Nightingale, A., Onwubiko, J., Palka, B., Pichler, K., Pundir, S., Qi, G. Y., Raj, S., Renaux, A., Lopez, M. R., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Xenarios, I., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Coudert, E., Cuhe, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T. B., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C. M., Chen, Y. X., Cowart, J., Garavelli, J. S., Huang, H. Z.,

- 759 Laiho, K., McGarvey, P., Natale, D. A., Ross, K., Vinayaka, C. R., Wang, Q. H., Wang, Y.
760 Q., Yeh, L. S., Zhang, J., & UniProt, C. (2019). UniProt: a worldwide hub of protein
761 knowledge [Article]. *Nucleic Acids Research*, 47(D1), D506-D515.
762 <https://doi.org/10.1093/nar/gky1049>
- 763 Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E.
764 S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., & Giraldez, A. J. (2014).
765 Identification of small ORFs in vertebrates using ribosome footprinting and
766 evolutionary conservation. *EMBO J*, 33(9), 981-993.
767 <https://doi.org/10.1002/emboj.201488411>
- 768 Beckedorff, F. C., Ayupe, A. C., Crocci-Souza, R., Amaral, M. S., Nakaya, H. I., Soltys, D. T.,
769 Menck, C. F., Reis, E. M., & Verjovski-Almeida, S. (2013). The intronic long noncoding
770 RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of
771 RASSF1A and increasing cell proliferation. *PLoS Genet*, 9(8), e1003705.
772 <https://doi.org/10.1371/journal.pgen.1003705>
- 773 Blair, J. D., Hockemeyer, D., Doudna, J. A., Bateup, H. S., & Floor, S. N. (2017). Widespread
774 Translational Remodeling during Human Neuronal Differentiation. *Cell Reports*,
775 21(7), 2005--2016. <https://doi.org/10.1016/j.celrep.2017.10.095>
- 776 Blighe, K., S., Rana, M., & Lewis. (2018). *EnhancedVolcano: Publication-ready volcano plots*
777 *with enhanced colouring and labeling*. In
- 778 Bodenhofer, U., Bonatesta, E., Horejs-Kainrath, C., & Hochreiter, S. (2015). msa: an R
779 package for multiple sequence alignment [Article]. *Bioinformatics*, 31(24), 3997-
780 3999. <https://doi.org/10.1093/bioinformatics/btv494>
- 781 Brenig, K., Grube, L., Schwarzländer, M., Köhrer, K., Stühler, K., & Poschmann, G. (2020). The
782 Proteomic Landscape of Cysteine Oxidation That Underpins Retinoic Acid-Induced
783 Neuronal Differentiation. *J Proteome Res*, 19(5), 1923-1940.
784 <https://doi.org/10.1021/acs.jproteome.9b00752>
- 785 Brockdorff, N. (2018). Local Tandem Repeat Expansion in Xist RNA as a Model for the
786 Functionalisation of ncRNA. *Non-Coding RNA*, 4(4), 28.
- 787 Brodie, S., Lee, H. K., Jiang, W., Cazacu, S., Xiang, C., Poisson, L. M., Datta, I., Kalkanis, S.,
788 Ginsberg, D., & Brodie, C. (2017). The novel long non-coding RNA TALNEC2, regulates
789 tumor cell growth and the stemness and radiation response of glioma stem cells.
790 *Oncotarget*, 8(19), 31785--31801. <https://doi.org/10.18632/oncotarget.15991>
- 791 Calviello, L., Mukherjee, N., Wyler, E., Zaubner, H., Hirsekorn, A., Selbach, M., Landthaler, M.,
792 Obermayer, B., & Ohler, U. (2016). Detecting actively translated open reading frames
793 in ribosome profiling data [Article]. *Nature Methods*, 13(2), 165-+.
794 <https://doi.org/10.1038/nmeth.3688>
- 795 Carelli, S., Giallongo, T., Rey, F., Latorre, E., Bordini, M., Mazzucchelli, S., Gorio, M. C.,
796 Pansarasa, O., Provenzani, A., & Cereda, C. (2019). HuR interacts with lincBRN1a and
797 lincBRN1b during neuronal stem cells differentiation. *RNA Biology*, 16(10), 1471--
798 1485. <https://doi.org/10.1080/15476286.2019.1637698>
- 799 Carlevaro-Fita, J., Rahim, A., Guigó, R., Vardy, L. A., & Johnson, R. (2016). Cytoplasmic long
800 noncoding RNAs are frequently bound to and degraded at ribosomes in human cells.
801 *Rna*, 22(6), 867--882. <https://doi.org/10.1261/rna.053561.115>
- 802 Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I.,
803 Collavin, L., Santoro, C., Forrest, A. R. R., Piero, C., Biffo, S., Stupka, E., & Gustincich,
804 S. (2012). Long non-coding antisense RNA controls Uchl1 translation through an

embedded SINEB2 repeat. *Nature*, 491(7424), 454--457.
<https://doi.org/10.1038/nature11508>

Carén, H., Ejeskär, K., Fransson, S., Hesson, L., Latif, F., Sjöberg, R. M., Krona, C., & Martinsson, T. (2005). A cluster of genes located in 1p36 are down-regulated in neuroblastomas with poor prognosis, but not due to CpG island methylation. *Mol Cancer*, 4(1), 10. <https://doi.org/10.1186/1476-4598-4-10>

Chan, P. P., & Lowe, T. M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, 44(D1), D184-D189. <https://doi.org/10.1093/nar/gkv1309>

Charif D, L. J. (2007). *SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis*. In

Chau, K. F., Shannon, M. L., Fame, R. M., Fonseca, E., Mullan, H., Johnson, M. B., Sendamarai, A. K., Springel, M. W., Laurent, B., & Lehtinen, M. K. (2018). Downregulation of ribosome biogenesis during early forebrain development. *Elife*, 7.

Chen, J., Brunner, A.-D., Cogan, J. Z., Nuez, J. K., Fields, A. P., Adamson, B., Itzhak, D. N., Li, J. Y., Mann, M., Leonetti, M. D., & Weissman, J. S. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science*, 367(6482), 1140--1146. <https://doi.org/10.1126/science.aay0262>

Chodroff, R. A., Goodstadt, L., Sirey, T. M., Oliver, P. L., Davies, K. E., Green, E. D., Molnár, Z., & Ponting, C. P. (2010). Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes.

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B. J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., Xenarios, I., Zhang, L., Trono, D., Harari, A., Ohler, U., Coukos, G., & Bassani-Sternberg, M. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun*, 11(1), 1293.

Chothani, S., Adami, E., Ouyang, J. F., Viswanathan, S., Hubner, N., Cook, S. A., Schafer, S., & Rackham, O. J. L. (2019). deltaTE: Detection of Translationally Regulated Genes by Integrative Analysis of Ribo-seq and RNA-seq Data. *Current Protocols in Molecular Biology*, 129(1), e108.

Clarke, E., & Sherrill-Mix, S. c. (2017). *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. In <https://CRAN.R-project.org/package=ggbeeswarm>

Consortium, G. T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*. <https://doi.org/10.1038/ng.2653>

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., Thomas, M., Davis, C. A., Shiekhhattar, R., Gingeras, T. R., Hubbard, T. J., Notredame, C., Harrow, J., & Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*, 22(9), 1775-1789.

Dimartino, D., Colantoni, A., Ballarino, M., Martone, J., Mariani, D., Danner, J., Bruckmann, A., Meister, G., Morlando, M., & Bozzoni, I. (2018). The Long Non-coding RNA Inc-31 Interacts with Rock1 mRNA and Mediates Its YB-1-Dependent Translation. *Cell Reports*, 23(3), 733--740. <https://doi.org/10.1016/j.celrep.2018.03.101>

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A.,

Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Bar, N. S., Batut, P., Bell, K., Bell, I., Chakraborty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O. J., Park, E., Persaud, K., Preall, J. B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L. H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S. E., Hannon, G., Giddings, M. C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., & Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, 489(7414), 101-108.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. <https://doi.org/10.1093/bioinformatics/bts635>

Duncan, C. D. S., & Mata, J. (2014). The translational landscape of fission-yeast meiosis and sporulation. *Nature Structural & Molecular Biology*, 21(7), 641-647. <https://doi.org/10.1038/nsmb.2843>

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10(1), 48. <https://doi.org/10.1186/1471-2105-10-48>

Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool [Article]. *Proteomics*, 13(1), 22-24. <https://doi.org/10.1002/pmic.201200439>

Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M. P., Nalls, M. A., Cookson, M. R., St-Laurent, G., 3rd, & Wahlestedt, C. (2010). Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol*, 11(5), R56.

Forster, J. I., Kglisberger, S., Trefois, C., Boyd, O., Baumuratov, A. S., Buck, L., Balling, R., & Antony, P. M. A. (2016). Characterization of Differentiated SH-SY5Y as Neuronal Screening Model Reveals Increased Oxidative Vulnerability. *Journal of Biomolecular Screening*, 21(5), 496-509. <https://doi.org/10.1177/1087057115625190>

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., Gonzalez, J. M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., G Izuogu, O., Lagarde, J., Martin, F. J., Martínez, L., Mohanan, S., Muir, P., Navarro, F. C. P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B. M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczyńska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J. S., Gerstein, M., Guigó, R., Hubbard, T. J. P., Kellis, M., Paten, B., Reymond, A., Tress, M. L., & Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*, 47(D1), D766-D773. <https://doi.org/10.1093/nar/gky955>

Fujii, K., Shi, Z., Zhulyn, O., Denans, N., & Barna, M. (2017). Pervasive translational regulation of the cell signalling circuitry underlies mammalian development. *Nature Communications*, 8(1), 14443. <https://doi.org/10.1038/ncomms14443>

Gordon, A. (2010). *FASTQ/A short-reads pre-processing tools*. In (Version v.0.0.14)

- 899 Guo, H., Ingolia, N. T., Weissman, J. S., & Bartel, D. P. (2010). Mammalian microRNAs
900 predominantly act to decrease target mRNA levels. *Nature*, 466(7308), 835--840.
901 <https://doi.org/10.1038/nature09267>
- 902 Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. (2013). Ribosome
903 profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*,
904 154(1), 240-251. <https://doi.org/10.1016/j.cell.2013.06.009>
- 905 Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., Lee, C. M.,
906 Lee, B. T., Hinrichs, A. S., Gonzalez, J. N., Gibson, D., Diekhans, M., Clawson, H.,
907 Casper, J., Barber, G. P., Haussler, D., Kuhn, R. M., & Kent, W. J. (2019). The UCSC
908 Genome Browser database: 2019 update. *Nucleic acids research*, 47(D1), D853--
909 D858. <https://doi.org/10.1093/nar/gky1095>
- 910 Hannon, G. J. (2010). FASTX-Toolkit.
- 911 Hulstaert, N., Shofstahl, J., Sachsenberg, T., Walzer, M., Barsnes, H., Martens, L., & Perez-
912 Riverol, Y. (2020). ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW
913 File Conversion. *Journal of Proteome Research*, 19(1), 537-542.
914 <https://doi.org/10.1021/acs.jproteome.9b00328>
- 915 Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2013). Genome-
916 Wide Annotation and Quantitation of Translation by Ribosome Profiling. 4.18.11--
917 14.18.19. <https://doi.org/10.1002/0471142727.mb0418s103>
- 918 Johnsson, P., Lipovich, L., Grander, D., & Morris, K. V. (2014). Evolutionary conservation of
919 long non-coding RNAs; sequence, structure, function [Review]. *Biochimica Et*
920 *Biophysica Acta-General Subjects*, 1840(3), 1063-1071.
921 <https://doi.org/10.1016/j.bbagen.2013.10.035>
- 922 King, J. L., & Jukes, T. H. (1969). NON-DARWINIAN EVOLUTION [Article]. *Science*, 164(3881),
923 788-+. <https://doi.org/10.1126/science.164.3881.788>
- 924 Korecka, J. A., van Kesteren, R. E., Blaas, E., Spitzer, S. O., Kamstra, J. H., Smit, A. B., Swaab,
925 D. F., Verhaagen, J., & Bossers, K. (2013). Phenotypic Characterization of Retinoic
926 Acid Differentiated SH-SY5Y Cells by Transcriptional Profiling. *PLoS ONE*, 8(5),
927 e63862. <https://doi.org/10.1371/journal.pone.0063862>
- 928 Krona, C., Ejeskär, K., Carén, H., Abel, F., Sjöberg, R. M., & Martinsson, T. (2004). A novel
929 1p36.2 located gene, APITD1, with tumour-suppressive properties and a putative
930 p53-binding domain, shows low expression in neuroblastoma tumours. *Br J Cancer*,
931 91(6), 1119-1130. <https://doi.org/10.1038/sj.bjc.6602083>
- 932 Ladoukakis, E., Pereira, V., Magny, E. G., Eyre-Walker, A., & Couso, J. P. (2011). Hundreds of
933 putatively functional small open reading frames in Drosophila [Article]. *Genome*
934 *Biology*, 12(11), 17, Article R118. <https://doi.org/10.1186/gb-2011-12-11-r118>
- 935 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
936 *Methods*, 9(4), 357-U354. <https://doi.org/10.1038/nmeth.1923>
- 937 Lewandowski, J. P., Dumbović, G., Watson, A. R., Hwang, T., Jacobs-Palmer, E., Chang, N.,
938 Much, C., Turner, K., Kirby, C., Schulz, J. F., Müller, C.-L., Rubinstein, N. D., Groff, A.
939 F., Liapis, S. C., Gerhardinger, C., Hubner, N., van Heesch, S., Hoekstra, H. E.,
940 Sauvageau, M., & Rinn, J. L. (2019). The *Tug1* Locus is Essential for Male
941 Fertility. *bioRxiv*, 562066. <https://doi.org/10.1101/562066>
- 942 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., &
943 Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools.
944 *Bioinformatics (Oxford, England)*, 25(16), 2078--2079.
945 <https://doi.org/10.1093/bioinformatics/btp352>

- 946 Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read
947 mapping by seed-and-vote. *Nucleic acids research*, 41(10), e108--e108.
948 <https://doi.org/10.1093/nar/gkt214>
- 949 Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program
950 for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*,
951 30(7), 923--930. <https://doi.org/10.1093/bioinformatics/btt656>
- 952 Lin, N., Chang, K.-Y., Li, Z., Gates, K., Rana, Z. A., Dang, J., Zhang, D., Han, T., Yang, C.-S.,
953 Cunningham, T. J., Head, S. R., Duester, G., Dong, P. D. S., & Rana, T. M. (2014). An
954 Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and
955 Neural Lineage Commitment. *Molecular Cell*, 53(6), 1005--1019.
956 <https://doi.org/https://doi.org/10.1016/j.molcel.2014.01.021>
- 957 Lindsay, S. J., Xu, Y., Lisgo, S. N., Harkin, L. F., Copp, A. J., Gerrelli, D., Clowry, G. J., Talbot, A.,
958 Keogh, M. J., Coxhead, J., Santibanez-Koref, M., & Chinnery, P. F. (2016). HDBR
959 Expression: A Unique Resource for Global and Individual Gene Expression Studies
960 during Early Human Brain Development. *Frontiers in Neuroanatomy*.
961 <https://doi.org/10.3389/fnana.2016.00086>
- 962 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and
963 dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
964 <https://doi.org/10.1186/s13059-014-0550-8>
- 965 Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the
966 nervous system. *Nature Reviews Neuroscience*, 8(10), 755-765.
967 <https://doi.org/10.1038/nrn2212>
- 968 Magny, E. G., Pueyo, J. I., Pearl, F. M., Cespedes, M. A., Niven, J. E., Bishop, S. A., & Couso, J.
969 P. (2013). Conserved regulation of cardiac calcium uptake by peptides encoded in
970 small open reading frames. In *Science* (Vol. 341, pp. 1116-1120).
971 <https://doi.org/10.1126/science.1238802>
- 972 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing
973 reads. 2011, 17(1), 3. <https://doi.org/10.14806/ej.17.1.200>
- 974 Michel, A. M., Mullan, J. P. A., Velayudhan, V., O'Connor, P. B. F., Donohue, C. A., & Baranov,
975 P. V. (2016). RiboGalaxy: A browser based platform for the alignment, analysis and
976 visualization of ribosome profiling data [Article]. *Rna Biology*, 13(3), 316-319.
977 <https://doi.org/10.1080/15476286.2016.1141862>
- 978 Mohammad, L., Wiseman, J., Erickson, S., & Yang, G. (2019). Protein Synthesis and
979 Translational Control in Neural Stem Cell Development and Neurogenesis.
980 <https://doi.org/10.1093/oxfordhb/9780190686307.013.21>
- 981 Murillo, J. R., Pla, I., Goto-Silva, L., Nogueira, F. C. S., Domont, G. B., Perez-Riverol, Y.,
982 Sánchez, A., & Junqueira, M. (2018). Mass spectrometry evaluation of a
983 neuroblastoma SH-SY5Y cell culture protocol. *Anal Biochem*, 559, 51-54.
984 <https://doi.org/10.1016/j.ab.2018.08.013>
- 985 Patil, I. (2018). ggstatsplot: 'ggplot2' Based Plots with Statistical Details. In [https://cran.r-](https://cran.r-project.org/web/packages/ggstatsplot/index.html)
986 [project.org/web/packages/ggstatsplot/index.html](https://cran.r-project.org/web/packages/ggstatsplot/index.html)
- 987 Patraquim, P., Mumtaz, M. A. S., Pueyo, J. I., Aspden, J. L., & Couso, J. P. (2020).
988 Developmental regulation of canonical and small ORF translation from mRNAs.
989 *Genome Biol*, 21(1), 128. <https://doi.org/10.1186/s13059-020-02011-5>
- 990 Pedersen, A. G., & Nielsen, H. (1997). Neural network prediction of translation initiation
991 sites in eukaryotes: perspectives for EST and genome analysis. *Proc Int Conf Intell*
992 *Syst Mol Biol*, 5, 226-233.

- 993 Pollen, A. A., Nowakowski, T. J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C. R.,
994 Shuga, J., Liu, S. J., Oldham, M. C., Diaz, A., Lim, D. A., Leyrat, A. A., West, J. A., &
995 Kriegstein, A. R. (2015). Molecular Identity of Human Outer Radial Glia during
996 Cortical Development. *Cell*, 163(1), 55–67.
997 <https://doi.org/10.1016/j.cell.2015.09.004>
- 998 Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long Noncoding
999 RNAs. *Cell*, 136(4), 629–641. <https://doi.org/10.1016/j.cell.2009.02.006>
- 1000 Pueyo, J. I., & Couso, J. P. (2008). The 11-aminoacid long Tarsal-less peptides trigger a cell
1001 signal in Drosophila leg development. In *Dev Biol* (Vol. 324, pp. 192–201).
1002 <https://doi.org/10.1016/j.ydbio.2008.08.025>
- 1003 Rodriguez, C. M., Chun, S. Y., Mills, R. E., & Todd, P. K. (2019). Translation of upstream open
1004 reading frames in a model of neuronal differentiation. *BMC Genomics*, 20(1), 391.
1005 <https://doi.org/10.1186/s12864-019-5775-1>
- 1006 Ruiz-Orera, J., & Alba, M. M. (2019). Conserved regions in long non-coding RNAs contain
1007 abundant translation and protein–RNA interaction signatures. *NAR Genomics and*
1008 *Bioinformatics*, 1(1), e2–e2. <https://doi.org/10.1093/nargab/lqz002>
- 1009 Saghatelian, A., & Couso, J. P. (2015). Discovery and characterization of smORF-encoded
1010 bioactive polypeptides. *Nat Chem Biol*, 11(12), 909–916.
1011 <https://doi.org/10.1038/nchembio.1964>
- 1012 Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W. Z., Lopez, R., McWilliam, H.,
1013 Remmert, M., Soding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable
1014 generation of high-quality protein multiple sequence alignments using Clustal
1015 Omega [Article]. *Molecular Systems Biology*, 7, 6, Article 539.
1016 <https://doi.org/10.1038/msb.2011.75>
- 1017 Song, E. J., Werner, S. L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J. W.,
1018 Elledge, S. J., Kirschner, M. W., & Rape, M. (2010). The Prp19 complex and the
1019 Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the
1020 spliceosome. *Genes Dev*, 24(13), 1434–1447. <https://doi.org/10.1101/gad.1925010>
- 1021 Sossin, W. S., & Scheller, R. H. (1991). Biosynthesis and sorting of neuropeptides. *Curr Opin*
1022 *Neurobiol*, 1(1), 79–83. [https://doi.org/10.1016/0959-4388\(91\)90013-w](https://doi.org/10.1016/0959-4388(91)90013-w)
- 1023 Spencer, H. L., Sanders, R., Boulberdaa, M., Meloni, M., Cochrane, A., Spiroski, A.-M.,
1024 Mountford, J., Emanuelli, C., Caporali, A., Brittan, M., Rodor, J., & Baker, A. H. (2020).
1025 The LINC00961 transcript and its encoded micropeptide, small regulatory
1026 polypeptide of amino acid response, regulate endothelial cell function.
1027 *Cardiovascular Research*. <https://doi.org/10.1093/cvr/cvaa008>
- 1028 Su, X., Zhang, J., Luo, X., Yang, W., Liu, Y., Liu, Y., & Shan, Z. (2019). LncRNA LINC01116
1029 Promotes Cancer Cell Proliferation, Migration And Invasion In Gastric Cancer By
1030 Positively Interacting With lncRNA CASC11. *OncoTargets and Therapy*, Volume 12,
1031 8117–8123. <https://doi.org/10.2147/OTT.S208133>
- 1032 Szafron, L. M., Balcerak, A., Grzybowska, E. A., Pienkowska-Grela, B., Felisiak-Golabek, A.,
1033 Podgorska, A., Kulesza, M., Nowak, N., Pomorski, P., Wysocki, J., Rubel, T., Dansonka-
1034 Mieszkowska, A., Konopka, B., Lukasik, M., & Kupryjanczyk, J. (2015). The Novel Gene
1035 CRNDE Encodes a Nuclear Peptide (CRNDEP) Which Is Overexpressed in Highly
1036 Proliferating Tissues. *PloS one*, 10(5), e0127475–e0127475.
1037 <https://doi.org/10.1371/journal.pone.0127475>

- 1038 Tsagakakis, I., Douka, K., Birds, I., & Aspden, J. L. (2020). Long non-coding RNAs in
1039 development and disease: Conservation to mechanisms. *J Pathol.*
1040 <https://doi.org/10.1002/path.5405>
- 1041 van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., Kirchner,
1042 M., Maatz, H., Blachut, S., Sandmann, C. L., Kanda, M., Worth, C. L., Schafer, S.,
1043 Calviello, L., Merriott, R., Patone, G., Hummel, O., Wyler, E., Obermayer, B., Mücke,
1044 M. B., Lindberg, E. L., Trnka, F., Memczak, S., Schilling, M., Felkin, L. E., Barton, P. J.
1045 R., Quaife, N. M., Vanezis, K., Diecke, S., Mukai, M., Mah, N., Oh, S. J., Kurtz, A.,
1046 Schramm, C., Schwinge, D., Sebode, M., Harakalova, M., Asselbergs, F. W., Vink, A.,
1047 de Weger, R. A., Viswanathan, S., Widjaja, A. A., Gärtner-Rommel, A., Milting, H., Dos
1048 Remedios, C., Knosalla, C., Mertins, P., Landthaler, M., Vingron, M., Linke, W. A.,
1049 Seidman, J. G., Seidman, C. E., Rajewsky, N., Ohler, U., Cook, S. A., & Hubner, N.
1050 (2019). The Translational Landscape of the Human Heart. In *Cell* (Vol. 178, pp. 242-
1051 260 e229). © 2019 Elsevier Inc. <https://doi.org/10.1016/j.cell.2019.05.010>
- 1052 Wang, H., Iacoangeli, A., Popp, S., Muslimov, I. A., Hiroaki, I., Sonenberg, N., Lomakin, I. B., &
1053 Tiedge, H. (2002). Dendritic BC1 RNA: Functional Role in Regulation of Translation
1054 Initiation. *Journal of Neuroscience*, 22(23), 10232--10241.
1055 <https://doi.org/10.1523/JNEUROSCI.22-23-10232.2002>
- 1056 Wang, L., Fan, J., Han, L., Qi, H., Wang, Y., Wang, H., Chen, S., Du, L., Li, S., Zhang, Y., Tang,
1057 W., Ge, G., Pan, W., Hu, P., & Cheng, H. (2020). The micropeptide LEMP plays an
1058 evolutionarily conserved role in myogenesis. *Cell Death & Disease*, 11(5), 357.
1059 <https://doi.org/10.1038/s41419-020-2570-5>
- 1060 Wickham, H. (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. In
1061 <https://CRAN.R-project.org/package=stringr>
- 1062 Winzi, M. a. (2018). The long noncoding RNA IncR492 inhibits neural differentiation of
1063 murine embryonic stem cells. *Plos One*, 13(1), e0191682.
1064 <https://doi.org/10.1371/journal.pone.0191682>
- 1065 Xiao, N., Cao, D. S., Zhu, M. F., & Xu, Q. S. (2015). protr/ProtrWeb: R package and web server
1066 for generating various numerical representation schemes of protein sequences
1067 [Article; Proceedings Paper]. *Bioinformatics*, 31(11), 1857-1859.
1068 <https://doi.org/10.1093/bioinformatics/btv042>
- 1069 Xie, Y. (2020). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. In
1070 Xin, X., Mains, R. E., & Eipper, B. A. (2004). Monooxygenase X, a Member of the Copper-
1071 dependent Monooxygenase Family Localized to the Endoplasmic Reticulum. *Journal*
1072 *of Biological Chemistry*, 279(46), 48159--48167.
1073 <https://doi.org/10.1074/jbc.M407486200>
- 1074 Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I.
1075 M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugan, J. C., Cummins, C.,
1076 Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C. G., Gil, L., Grego, T., Haggerty,
1077 L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., Kay, M.,
1078 Lavidas, I., Le, T., Lemos, D., Martinez, J. G., Maurel, T., McDowall, M., McMahon, A.,
1079 Mohanan, S., Moore, B., Nuhn, M., Oheh, D. N., Parker, A., Parton, A., Patricio, M.,
1080 Sakthivel, M. P., Salam, A. I. A., Schmitt, B. M., Schuilenburg, H., Sheppard, D.,
1081 Sycheva, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Vullo, A., Walts, B.,
1082 Winterbottom, A., Zadissa, A., Chakiachvili, M., Flint, B., Frankish, A., Hunt, S. E.,
1083 Ilsley, G., Kostadima, M., Langridge, N., Loveland, J. E., Martin, F. J., Morales, J.,
1084 Mudge, J. M., Muffato, M., Perry, E., Ruffier, M., Trevanion, S. J., Cunningham, F.,

1085 Howe, K. L., Zerbino, D. R., & Flicek, P. (2020). Ensembl 2020 [Article]. *Nucleic Acids*
1086 *Research*, 48(D1), D682-D688. <https://doi.org/10.1093/nar/gkz966>
1087 Zubarev, R. A. (2013). The challenge of the proteome dynamic range and its implications for
1088 in-depth proteomics. *Proteomics*, 13(5), 723-726.
1089 <https://doi.org/10.1002/pmic.201200451>
1090

Figure 1

Figure 1: Differentiation of SH-SY5Y results in a reduction in the level of active translation.

Differentiation of SH-SY5Y cells with trans-retinoic acid (RA) for 3 days results in neurite elongation as seen by (A) representative immunofluorescence after staining for TuJ1 (β III-tubulin) (arrowheads mark elongated neurites) and (B) quantification of neurite length (Student's t-test, $n=3$, $p<0.01$). (C) RA treatment also results in reduction of proliferation, indicated by the reduction of ki67+ cells (Student's t-test, $n=1000$, $p<0.001$). (D) Sucrose gradient UV absorbance profiles of undifferentiated (Control) and differentiated (RA) cells reveals global translation is reduced upon differentiation. Peaks correspond to ribosomal subunits (40S, 60S), monosome (80S) and polysomes. (E) Quantification of Polysome/Monosome (P/M) ratio indicates reduction of active translation is significant upon RA treatment (student's t-test, $n=3$, $p<0.001$). (F) RPL26 ribosomal protein mRNA shift from polysomes to RNPs and ribosomal subunits upon differentiation, as shown by RT-qPCR across gradient (student's t-test, $n=3$, $p<0.05$).

Sup 1

Sup Figure 1: Differentiation of SH-SY5Y results in a reduction in the level of active translation.

(A) Light microscopy reveals that undifferentiated SH-SY5Y cells grow in clumps and upon neural induction with RA they extend their neurites. (B) Immunofluorescence indicates c-Fos differentiation marker is expressed at higher levels in differentiated cells. (C) Quantification of c-Fos⁺ cells in Control and RA (Student's t-test, n=1000, p<0.05). (D) Immunofluorescence shows ki67 proliferation marker expression is reduced upon differentiation. (E) RT-qPCR of pluripotency/proliferation marker SOX2 shows reduced levels upon differentiation. (F) RT-qPCR of the differentiation marker MOXD1 indicates differentiation towards outer radial glia neural precursors (Student's t-test, n=3, p<0.05). (G) Quantification of polysome (Student's t-test, n=3, p<0.05) and monosome (Student's t-test, n=3, p<0.01) areas upon differentiation. (H) RT-qPCR of RPL26 mRNA distribution (Control; orange, RA; green) (with GAPDH mRNA controls; Control; light grey, RA; dark grey), in RNPs and ribosomal subunits (RNPs, 40S, 60S), monosome (80S) and polysomes (Student's t-test, n=3, p<0.05). RT-qPCR across gradients of RPS28 (I) and RPL37 (J) mRNAs respectively (Student's t-test, n=3, p<0.05).

Figure 2

Figure 2: Differentiation results in global RNA changes

(A) Schematic of Poly-Ribo-Seq. (B) Volcano plot of protein-coding gene level changes in response to differentiation (polysome-associated); with 897 protein-coding genes significantly upregulated and 732 significantly downregulated ($\log_2$ fold-change cut-off=1, $p^{\text{adj}} < 0.05$). GO terms enriched in polysome-associated protein-coding genes (C) upregulated and (D) downregulated upon differentiation. (E) Genome browser visualization of RNA-seq reads mapping to neuronal transcript Secretogranin, induced upon differentiation. (F, G) Volcano plots of differential expression of protein-coding transcripts between Total cytoplasm and polysomes for (F) Control and (G) RA, revealing the depletion of RP mRNAs from polysomes upon differentiation. (H) Venn diagrams of overlap between those mRNAs enriched or depleted in the polysomes between Control and RA.

Sup 2

A

B

C

D

E

Upregulated in Total

F

Downregulated in Total

G GO analysis of polysome depleted genes in Control

H GO analysis of polysome enriched genes in Control

I GO analysis of polysome depleted genes in RA

Sup Figure 2: Differentiation results in global RNA changes

(A) PCA analysis of the RNA-Seq datasets shows that Control samples are more similar to each other than to RA treated samples. Control samples cluster separately from RA samples, indicating significant variation in gene expression between Control and differentiated cells. RA total samples cluster separately from RA polysome samples, whereas Control samples do not show different clustering, indicating that the variation in gene expression between RA total and RA polysome is greater than the variation between Control total and Control polysome. (B) Volcano plot depicting the differentially expressed protein-coding genes between Total Control and Total RA populations; 936 protein-coding mRNAs are upregulated and 691 downregulated upon differentiation ($\log_2$ fold-change cut-off=1, $p^{\text{adj}} < 0.05$). Venn diagrams of overlap between those mRNAs identified as up (C) or down (D) regulated between Total and Polysome populations. GO term analysis for total cytoplasmic protein-coding genes (E) upregulated in and (F) downregulated. GO term analysis for protein-coding genes; (G) depleted from polysomes in Control, (H) enriched in polysomes in Control and (I) depleted from polysomes in RA.

Figure 3

Control vs RA Polysome associated lncRNAs

A **B** **C** **D**

DLGAP1-AS2 antisense lncRNA

F **E** **G** **H**

Figure 3: LncRNAs regulated and associated with polysomes

(A) Volcano plot displaying the differentially expressed polysome-associated lncRNAs (labelled by their geneIDs) between Control and RA populations. 237 lncRNAs are upregulated during differentiation and 82 downregulated ($\log_2$ fold-change cutoff=1, $p^{\text{adj}} < 0.05$). Pie charts showing types of lncRNAs (B) upregulated and (C) downregulated upon differentiation (intergenic; anti-sense; sense-overlapping; retained intron; sense-intronic; uncharacterized; NMD target). (D) Example of lncRNA induced upon differentiation, DLGAP1-AS2. RNA-seq coverage for DLGAP1-AS2 in Control and RA conditions. (E) Changes in lncRNA levels upon differentiation validated by RT-qPCR, with XIST lncRNA and GAPDH mRNA as controls (SE is plotted, $n=3$, student's t-test $p < 0.05$ and $p < 0.01$). (F) lncRNAs of interest that are induced are specifically localised to cytoplasm as shown by subcellular fractionation RT-qPCR. XIST lncRNA was used as a nuclear and GAPDH mRNA as a cytoplasmic positive control ($n=3$, SE is plotted, student's t test, $n=3$, $p > 0.05$). RT-qPCR of lncRNAs across sucrose gradient fractions indicates that (G) LINC02143 is found in 80S and small polysome fractions during differentiation (in differentiated cells 5% of the transcripts is detected in 80S (monosome) fraction and 65.6% in small polysome complexes) and (H) LINC01116 is found in 80S and 2-7 polysome fractions both in control and RA treated cells. On average, 66% of the LINC01116 transcripts is detected in the polysome fractions in Control and 57% upon differentiation. ($n=3$, SE is plotted).

A

B

C

D

E

F

Sup Figure 3: Regulation of lncRNA expression and polysome association

(A) Pipeline of lncRNA differential expression analysis. (B) Volcano plot displaying the significantly differentially expressed lncRNAs (labelled by their geneIDs) between Control and RA populations (Total) with $\log_2$ fold-change cutoff=1, $p^{\text{adj}} < 0.05$. Venn diagrams of overlap in differentially expressed lncRNAs between Total and Polysomes datasets for (C) upregulated lncRNAs and (D) downregulated lncRNAs upon differentiation. Pie charts showing breakdown by lncRNA type for lncRNAs (E) upregulated and (F) downregulated upon differentiation (Polysome) (dark purple: intergenic; magenta: anti-sense; purple: sense-overlapping; dark pink: retained intron; light pink: sense-intronic; white: uncharacterized; bright pink: NMD target).

Control Total vs Control polysome-IncRNAs

A

RA Total vs RA polysome-IncRNAs

B

C

LncRNAs depleted from polysomes in RA

D

LncRNAs depleted from polysomes in Control

E

Correlation of log₂ fold-change between qPCR and RNA-Seq (total)

F

G

H

Sup Figure 4: Regulation of lncRNA expression and polysome association

Volcano plots displaying the significantly differentially localised lncRNAs (labelled by their geneIDs) between Total and Polysome populations in (A) Control and (B) RA with $\log_2$ fold-change cutoff=1, $p^{\text{adj}} < 0.05$. Pie charts of lncRNA types present in different populations (C) depleted from polysomes in RA and (D) depleted from polysomes in Control (intergenic; anti-sense; sense-overlapping; retained intron; sense-intronic; uncharacterized; NMD target). (E) Correlation of RT-qPCR data for fold-change for lncRNAs with RNA-Seq analysis (LINC01116, LINC02143, SNAP25-AS1, DLGAP1-AS1, DLGAP1-AS2, SERPINB9P1 and AC254633.1). (F) Western blot confirming cellular fractionation with nuclear marker (H3K27me3) and cytoplasmic markers (beta-tubulin). (G) lncRNAs of interest that are induced are specifically localised to cytoplasm as shown by subcellular localisation RT-qPCR for Control and RA samples. XIST lncRNA was used as a nuclear and GAPDH mRNA as a cytoplasmic positive control (n=3, SE is plotted). (H) RT-qPCR of lncRNAs across sucrose gradient fractions indicates that DLGAP1-AS2 is found in 80S and small polysome fractions during differentiation and fractions both in control and RA treated cells (n=3, SE is plotted) On average, 63% of the transcripts is detected in the polysome fractions in Control and 49% upon differentiation.

Figure 4

Figure 4: Translation of lncRNA smORFs

(A) Poly-Ribo-Seq exhibits triplet periodicity, reflecting genuine translation events. (B) Workflow for identification of translated ORFs from ribo-seq and RNA-seq, see materials and methods for details. (C) Plot of log₂-fold change for each ORF in RNA (RNA-Seq) versus ribosome footprints (Ribo-Seq). DTOs (differentially transcribed ORFs) or forward ORFs are driven by transcriptional regulation, with significant Δ RPF and Δ RNA but no significant Δ TE (blue). DTEOs (differential translation efficiency ORFs) or exclusive ORFs are driven by translational regulation, with significant Δ RPF and Δ TE without any significant Δ RNA (red). DTO & DTEOs (ORFs undergoing differential transcription and translation efficiency) are either intensified; a significant Δ TE acts with a significant Δ RNA, or buffered; a significant Δ TE is completely counteracting a Δ RNA, leading to no significant Δ RPF (pink). (D) View of Poly-Ribo-Seq data for CYP26B1. This is DTG ORF and exhibits significant Δ RNA and Δ RPF. (E) Venn diagram of lncRNA smORFs translated in Control and RA, with overlap. (F) Plots of Translational Efficiencies for protein-coding ORFs, lncRNA-smORFs, dORFs and uORFs. (G) Length distribution of translated ORFs in lncRNAs, dORFs and uORFs (in codons) in Control and RA. (H) Amino acid usage for smORFs compared to protein-coding ORFs and expected frequency. (I) Poly-Ribo-Seq profile for LINC01116 in RA treatment. RNA-Seq (Polysome) reads are grey and ribosome P sites in purple, turquoise and yellow according to frame. Purple lines mark beginning and end of translated smORF. All possible start and stop codons are indicated below. Framing within and outside translated smORF shown on left.

Translated ORFs	Control	RA	Overlap	Total
Protein-Coding ORFs	16,282	12,745	10,014	19,013
uORFs	56	27	12	71
dORFs	4	1	0	5
lncRNA-smORFs	28	23	6	45

Table 1: Translation of small ORFs

Number of ORFs detected as translated in Poly-Ribo-Seq

Sup Fig 5

A

C

B

D

E

Sup Figure 5: Translation of lncRNA smORFs

(A) Triplet periodicity plot for 25-35 nt reads from RA sample, replicate 3. (B) Pie charts of numbers of reads mapping to CDSs and elsewhere for ribosome footprints and RNA-Seq (Total). Values are mean of three Control replicates. (C) Metagene analysis for 33 nt reads from 33 nt reads from replicate x, Control sample. (D) Length distribution of translated ORFs in lncRNAs, dORFs and uORFs (in codons), results from Control and RA combined. (E) Example Poly-Ribo-Seq profile for LINC01116 in control treatment. RNA-Seq (Polysome) reads are grey and ribosome P sites in red, green and blue according to frame. Purple lines mark beginning and end of translated smORF. All possible start and stop codons are indicated below. Framing within and outside translated smORF shown on left.

Sup Table 1

Summary of triplet periodicity data for all 3 replicates, Control and RA, for 31 nt and 33 nt footprint lengths

Control	33nt long reads		
	Frame 0	Frame 1	Frame 2
rep1	22863313	49448301	21722546
rep2	2805820	2664114	2205933
rep3	15264767	27379738	7843708
RA	33nt long reads		
	Frame 0	Frame 1	Frame 2
rep1	17268000	29857839	18325348
rep2	3129165	2453330	2127290
rep3	11751674	21167199	5521907

Control	31nt long reads		
	Frame 0	Frame 1	Frame 2
rep1	24005094	32426374	60624552
rep2	8362106	23604658	23159636
rep3	9924066	17280912	30882341
RA	31nt long reads		
	Frame 0	Frame 1	Frame 2
rep1	15605202	15497506	16166029
rep2	4896645	12171565	11646160
rep3	19042787	45325747	85476963

Figure 5

Figure 5: Peptide production from smORFs in lncRNAs

(A) Venn diagrams showing overlap in uORFs and lncRNA-smORFs detected, between our Poly-Ribo-Seq and publically available mass spectrometry data from SH-SY5Y (purple). Control in blue, RA in pink. (B) Schematic of tagging construct for LINC00478; lncRNA sequence upstream of smORF and smORF, excluding its stop codon, cloned upstream of 3X FLAG, which is lacking its own start codon. FLAG signal is therefore dependent on smORF translation. (C) Poly-Ribo-Seq profile for LINC00478 in RA treatment. RNA-Seq (Polysome) reads are grey and ribosome P sites in purple, turquoise and yellow according to frame. Purple lines mark beginning and end of translated smORF. All possible start and stop codons are indicated below. Framing within and outside translated smORF shown on right. (D) Confocal images of FLAG-tagged LINC00478 peptide in SH-SY5Y cells (Control), showing (i) nuclear and (ii) cytoplasmic distribution, green is FLAG, magenta is hnRNPK (marking nuclei) and blue is DAPI (scale bar is 20µm). (E) Schematic of tagging constructs for LINC01116 (WT and start codon mutant Δ1). (F) Confocal images of FLAG-tagged LINC01116 peptide (WT showing cytoplasmic localisation, near cell membrane and neuritic processes (magnification of insert is 3X). (G) Δ1 start codon mutant, showing no FLAG signal, in SH-SY5Y cells; green is FLAG and blue is DAPI (scale bar is 20µm). (H) Portion of human ENST00000442428 lncRNA nt alignment with gorilla and orangutan nt sequences, showing the smORF. Alignment built in ClustalOmega (Sievers, F. et al. 2011) displayed in JalView (Waterhouse, A.M. et al. 2009). (I) Phylogram with lncRNA-smORFs for which evidence of sequence conservation were found represented as circles, coloured according to how sequence conservation was identified. Phylogram built in iTOL (Letunic, I. et al. 2006) using data from TimeTree (Kumar, S. , et al. 2017), scale in 10 MYA along.

LINC00478 in HEK293

ENST00000502603	ENST00000501117	ENST00000509075	ENST0000045391	ENST00000500479	ENST00000500486	ENST0000044941	ENST0000045493	ENST0000055766	ENST0000056622	ENST0000050294	ENST0000051022	ENST0000054586	ENST0000065116	ENST0000044125	ENST0000042994	ENST0000081220
6.1_122_6_2042	7.7_136_388	0.1_25_283	0.5_151_262	2.6_74_230	9.1_132_360	9.5_342_636	5.1_477_633	0.5_42_186	0.2_141_570	1.5_6_22	1.5_6_2_85	9.1_2_2_12	3.1_67_343	7.1_75_243	0.6_97_595	8.1_220_352

X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Sup Figure 6: Peptide production from smORFs in lncRNAs

(A) Representative confocal images of FLAG-tagged LINC00478 peptide in (i) Control and (ii-iii) RA treated SH-SY5Y cells. LINC00478 exhibits nuclear distribution. (B) Representative confocal images of transfections of FLAG-tagged LINC00478 smORF in HEK293 cells; showing (i-ii) nuclear and (iii) cytoplasmic distribution, green is FLAG, and blue is DAPI (scale bar is 20µm). (C) Representative confocal images of FLAG-tagged LINC01116 smORF (i) WT showing cytoplasmic distribution near cell membrane and filopodia. Green is FLAG and blue is DAPI (scale bar is 16µm). (D) schematic of LINC01116 smORF with 2 start codons annotated. (E) Negative control (no transfection) in (i) SH-SY5Y and (ii) HEK293 cells; green is FLAG and blue is DAPI (scale bar is 20µm). (F) Phylogram with lncRNA-smORFs for which evidence of sequence conservation were found represented as circles, coloured according to how sequence conservation was identified. Each lncRNA-smORF with evidence of sequence conservation is shown along with which species shown conservation. Phylogram built in iTOL (Letunic, I. et al .2006) using data from TimeTree (Kumar, S. , et al. 2017), scale in 10 MYA along.

Figure 6**A** **C** **B**

Control

D **E** **F** **G**

Figure 6: LINC01116 contributes to neuronal differentiation but does not affect cell cycle progression.

(A) LINC01116 was efficiently knocked down prior to differentiation and the knockdown is persistent throughout differentiation. (B) Representative immunofluorescence image of Control and RA SH-SY5Y cells, transfected with siRNA targeting LINC01116 and scrambled control, after staining for Tuj (β III-tubulin) at day 3 post-differentiation (scale bar=200 μ m). White windows magnified in (E). (C) Quantification of neurite length in Control and RA treated cells upon knockdown shows a significant reduction of neurite length in the differentiated cells that lack expression of LINC01116 (N=3 biological replicates, n>100 measurements, student's t-test $p<0.05$). (D) RT-qPCR of differentiation marker MOXD1 in Control and RA treated cells, transfected with siRNA targeting LINC01116 and scrambled control, shows significant reduction of MOXD1 expression in differentiated cells with reduced LINC01116 levels at day 3 post-differentiation (n=3 biological replicates, student's t-test $p<0.05$). (F) Quantification of proliferating cells (% of ki67+ cells) in Control and RA cells shows no effect of LINC01116 knockdown on cell proliferation (N=3 biological replicates, n>100 measurements, student's t-test $p>0.05$). (G) LINC01116 knockdown does not affect cell cycle progression as shown by RT-qPCR targeting cell cycle promoting transcription factor E2F1 (n=3 biological replicates, student's t-test $p>0.05$).

Sup Fig 7

A

B

C

D

Sup Figure 7: LINC01116 expression is upregulated early on upon differentiation and does not affect cell cycle progression.

(A) LINC01116 knockdown is not cytotoxic, as shown by cell viability assay (N=3 biological replicates, n=2 technical duplicates per replicate, student's t-test $p>0.05$). (B) Representative immunofluorescence of Control and RA SH-SY5Y cells, transfected with LINC01116 or scrambled siRNA, after staining for proliferation marker ki67 (arrowheads mark ki67+ cells) at day 3 post differentiation (N=3 biological replicates, n>100 measurements, student's t-test $p>0.05$). Scale bar=100 μ m (C) Expression of LINC01116 lncRNA is upregulated from first day post-differentiation as shown by RT-qPCR (N=3 biological replicates, student's t-test $p<0.05$). (D) LINC01116 expression levels, in phase 1 and phase 2 differentiated SH-SY5Y cells, measured by RT-qPCR. LINC01116 levels decrease ~6-fold during phase 2 differentiation (n=3, standard deviation is plotted).