LncRNAs interacting with the translation machinery contribute to human neuronal differentiation

Katerina Douka[1,2], Isabel Birds[1,2], Dapeng Wang[2], Sophie Clayton[1,3], Abigail Byford[1,3], Elton J. R. Vasconcelos[2], Mary J. O'Connell[5], 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

36 mechanisms (Dimartino et al., 2018; Tsagakis et al., 2020), such as the base-pairing of lnc-31 with
37 Rock1 mRNA, to promote its translation during myoblast differentiation (Dimartino et al., 2018). ~40%
38 of human lncRNAs are specifically expressed in the brain (Derrien et al., 2012), where they display
39 precise spatiotemporal expression profiles (Ponting et al., 2009). A subset of nuclear neuronal lncRNAs
40 have been found to regulate neuronal differentiation in mouse and human (Carelli et al., 2019;
41 Chodroff et al., 2010; Lin et al., 2014; Winzi, 2018). Until recently, the majority of work has focused on
42 nuclear lncRNA functions, following the belief that most lncRNAs were retained in the nucleus (Derrien
43 et al., 2012; Djebali et al., 2012). However, it has become increasingly clear that many lncRNAs are
44 exported to the cytoplasm (Carlevaro-Fita et al., 2016) and have specific cytoplasmic functions in post-
45 transcriptional gene regulation whilst some possess specific neuronal roles. For example, UCHL1-AS is
46 detected in dopaminergic neurons and promotes translation of UCHL1 mRNA by increasing its
47 association with heavy polysomes (Carrieri et al., 2012). Whereas BACE1-AS transcript, which is
48 significantly upregulated in the brain of Alzheimer's disease patients, base-pairs with beta-secretase-
49 1 (BACE1) mRNA, stabilising it by masking miR-485-5p binding sites (Faghihi et al., 2010). Alternatively,
50 BC200 represses translation initiation in dendrites by disrupting the formation of pre-initiation 48S
51 complexes (Wang et al., 2002).
52 Translation regulation is essential during neuronal differentiation (Mohammad et al., 2019),
53 with the translation of non-canonical ORFs, e.g. upstream ORFs (uORFs), playing important roles (Blair
54 et al., 2017; Fujii et al., 2017; Rodriguez et al., 2019). Ribosome profiling in a range of organisms and
55 tissue types has revealed the translation of a variety of non-canonical ORFs including small ORFs
56 (smORFs) <100 codons in length (Aspden et al., 2014; Blair et al., 2017; Duncan & Mata, 2014; Fujii et
57 al., 2017; Guo et al., 2010; Ingolia et al., 2013; Rodriguez et al., 2019). Although these translation
58 events remain controversial, it is clear that lncRNAs can interact with the translation machinery.
59 Limited ribosome profiling signal found on smORFs might be the result of sporadic binding of a single
60 ribosome and may not necessarily correspond to active translation. Poly-Ribo-Seq was previously
61 developed to distinguish those lncRNAs that are bound by multiple ribosomes, and therefore actively
62 translated, from such a background signal. Predicting which lncRNA-smORFs are translated has proven
63 difficult, as attributes to predict protein-coding ability are ineffective on small non-canonical ORFs. A
64 small but growing number of smORF peptides translated from lncRNAs have been found to exhibit
65 cellular and organismal functions (Anderson et al., 2015; Chen et al., 2020; Magny et al., 2013; Pueyo
66 & Couso, 2008; Spencer et al., 2020; Wang et al., 2020).
67 To study the dynamic interactions of lncRNAs with the translation machinery and identify
68 actively translated cytoplasmic lncRNAs during early neuronal differentiation, we perform Poly-Ribo-
69 Seq on SH-SY5Y cells (human neuronal cell line). During differentiation a global decrease in translation

70 is evident and ribosomal protein mRNAs are specifically excluded from the polysomes during
71 differentiation as part of this down-regulation. We detect ~800-900 lncRNA genes in the cytoplasm of
72 SH-SY5Y cells. Specifically, 237 lncRNAs are upregulated and 100 downregulated during
73 differentiation, 58-70% of which interact with polysome complexes. These interactions are dynamic
74 during differentiation. Ribo-Seq identified 45 actively translated smORFs in lncRNAs, several of which
75 are regulated during differentiation. A subset of these translation events was validated by transfection
76 tagging experiments and analysis of publicly available mass spectrometry data. High levels of sequence
77 conservation across *Homimidae* indicates that the resulting lncRNA-smORF peptides are produced in
78 other great apes and likely possess cellular functions. One of the lncRNA genes that we discover to be
79 significantly upregulated during differentiation is LINC01116, which is cytoplasmically enriched and
80 associated with polysomes. Our Ribo-Seq analysis reveals the translation of an 87aa peptide from
81 LINC01116 with increased ribosome profiling coverage upon differentiation. The LINC01116 peptide
82 exhibits a cytoplasmic distribution and is detected in neurites. We reveal that LINC01116 contributes
83 to neuronal differentiation and is required for proper neurite lengthening.
84
85 Results
86 **Differentiation of SH-SY5Y cells with retinoic acid results in reduced translation levels**
87 To dissect the importance of cytoplasmic lncRNAs and their ribosome associations in early
88 neuronal differentiation we profiled the differentiation of SH-SY5Y cells with retinoic acid (RA) for 3
89 days. This treatment results in neuronal differentiation as indicated by neurite elongation (Sup 1A),
90 which can be seen by immunostaining for neuronal βIII-tubulin (TuJ1) (Fig 1A). Quantification of
91 neurite length reveals significant elongation upon RA treatment (Fig 1B). There is also increased
92 expression of neuronal markers; more cells express c-Fos upon differentiation (Sup 1B, C). There is a
93 concomitant reduction in cell proliferation, as seen by reduced number of Ki67+ cells (Sup 1D, Fig 1C)
94 as well as a reduction in the levels of pluripotency marker SOX2 (Sup 1E). These differentiated cells
95 exhibit characteristics of outer radial glia neural progenitors and are driven towards a noradrenergic
96 phenotype as indicated by increased Dopamine Beta Hydroxylase like Monoxygenase Protein 1
97 (MOXD1) expression (Sup 1F) (Pollen et al., 2015; Xin et al., 2004). Translational output assessment by
98 polysome profiles (Fig 1D) reveals that differentiation results in down-regulation of global translation.
99 Quantification of translation complexes across the sucrose gradients indicates that levels of polysomes
100 are reduced with respect to 80S monosomes (Sup 1G) resulting in a reduced polysome to monosome
101 ratio (Fig 1E). This down-regulation of translation is accompanied by a shift of ribosomal protein (RP)
102 mRNAs from polysomes to monosomes; e.g. RpL26 mRNA (Fig 1F, Sup 1H), RpS28 (Sup 1I) and RpL37

103 (Sup 1J), as measured by RT-qPCR across gradient fractions. This reduced synthesis of RPs has
104 previously been reported during neuronal differentiation (Blair et al., 2017; Chau et al., 2018).
105
106 **Poly-Ribo-Seq reveals differences in RNA expression and polysome association upon differentiation**
107 To profile RNA, ribosome association and translational changes upon differentiation we
108 employed Poly-Ribo-Seq (Aspden et al., 2014) with some minor modifications to adapt to human
109 neuronal cells (Fig 2A). This adaptation of ribosome profiling (Ribo-Seq) can detect which RNAs are
110 cytoplasmic, polysome associated and translated (Fig 2A). We sequenced i) poly-A selected
111 cytoplasmic RNA, 'Total' RNA-Seq, ii) Polysome-associated poly-A RNAs, 'Polysome' RNA-Seq, and iii)
112 ribosomal footprints, Ribo-Seq, from control and RA differentiated cells, with three biological
113 replicates (Materials and Methods for details).
114 PCA of the 3 Total and 3 Polysome RNA-Seq for both Control and RA treatment shows that RA
115 treated samples cluster separately from Control samples and biological replicates generally cluster
116 together (Sup 2A). To profile differences as a result of differentiation we compared Control and RA
117 samples, performing differential expression analysis at the gene level for both Polysomal RNA-Seq
118 samples (main figures) and Total Cytoplasmic RNA-Seq samples (supplemental figures). To dissect the
119 ability of RNAs to associate with translation complexes we compared the "Polysomal" and "Total" RNA
120 populations for both Control and RA conditions, again through differential expression analysis at the
121 gene level. Differences in expression between Control and RA are greater than those seen between
122 Total and Polysome populations. Interestingly the variation in gene expression between differentiated
123 (RA) Total and Polysome is greater than the variation between Control Total and Control Polysome
124 (Sup 2A).
125 Analysis of the two types of RNA-Seq revealed large changes in RNA expression in response to
126 differentiation, as previously determined (Blair et al., 2017). In fact, 897 protein-coding genes are
127 upregulated upon RA treatment and 732 downregulated, within the Polysomes (Fig 2B). A very similar
128 pattern is seen from Total; 936 up-regulated and 691 down-regulated genes (Sup 2B), and 78% of the
129 significantly affected protein-coding genes are in common between Total and Polysome populations
130 (Sup 2C&D). GO term analysis of these differentiation-regulated protein-coding genes indicates that
131 functions corresponding to neuronal differentiation are upregulated (Fig 2C). There is an enrichment
132 of ribosome and translation functions in those genes downregulated upon neuronal differentiation
133 (Fig 2D). A clearer enrichment of genes with neuronal functions is evident in GO analysis of Total RNA-
134 Seq (Sup 2E), e.g. synaptic transmission and similar enrichment of translation functions are in the
135 downregulated set (Sup 2F). Expression of genes required for neuronal differentiation is increased or

136 induced upon RA treatment, e.g. Secretogranin, which is involved in packaging of neuropeptides into
137 secretory vesicles (Sossin & Scheller, 1991) (Fig 2E).
138 Comparison of Total and Polysome RNA-Seq indicates that some mRNAs are enriched in
139 polysomes whilst others are depleted. Ribosomal protein mRNAs are specifically depleted from
140 polysomes upon differentiation, as seen by their altered relative levels between Total and Polysomes
141 in Control (Fig 2F) and RA (Fig 2G). This is consistent with a global downregulation in translation (Fig
142 1D) and shift of RP mRNAs out of the polysomes (Fig 1F). By comparing these polysome enriched and
143 depleted mRNAs between Control and RA the dynamic nature of polysome association during
144 differentiation is revealed (Fig 2H). More mRNAs become depleted during differentiation, including
145 more RP mRNAs. GO term analysis of those genes enriched or depleted in the polysomes (Sup 2G, H,
146 I) supports this depletion of RP mRNAs upon differentiation (Sup 2I).
147
148 **LncRNAs are induced and associated with polysomes during differentiation**
149 To understand the potential role and regulation of cytoplasmic lncRNAs we analysed their
150 expression levels in Total and Polysome populations upon differentiation (Sup 3A). We detect large
151 numbers of lncRNAs present in the cytoplasm. 801 lncRNA genes expressed and present (i.e. have
152 RPKMs $\geq$ 1) in the cytoplasm under control conditions and 916 lncRNA genes in differentiated cells.
153 Interestingly, 237 lncRNA genes were upregulated during differentiation whilst only 82 were
154 downregulated, within Polysome fractions (Fig 3A). A very similar pattern was seen when considering
155 Total RNA-seq with 178 lncRNA genes up-regulated and 100 down-regulated (Sup 3B). The majority of
156 these differences were identified in both Total and Polysome populations (70% of the upregulated
157 and 58% of the downregulated lncRNAs were in common between total and polysome) (Sup 3C&D).
158 Interestingly when we look in more detail at these lncRNA regulated during differentiation the
159 majority of differentiation-regulated lncRNAs are either intergenic or/and anti-sense lncRNAs (Fig 3
160 B&C-Polysome, Sup 3 E&F-Total). Direct comparison of levels of lncRNAs within the whole cytoplasm
161 and within polysome fractions indicates the vast majority lncRNAs (Control: 98% and RA: 99% are
162 neither polysome enriched nor depleted (Sup 4 A&B). A small number (Control: 12 and RA: 10) of
163 lncRNAs are specifically depleted from the polysomes including several lncRNAs derived from small
164 nucleolar RNA host genes. Interestingly there is a smaller proportion of anti-sense lncRNAs within
165 these polysome-depleted populations (Sup 4 C&D) compared to differences in expression during
166 differentiation (Fig 3 B&C).
167 Significant induction of specific polysome-associated lncRNAs during differentiation, such as
168 DLGAP1-AS2 is suggestive of a regulatory role during neuronal differentiation (Fig 3 A&D). We
169 validated these differentiation-induced changes in a subset of 7 lncRNAs (Fig 2E). By RT-qPCR the

170 expression of all candidate lncRNAs were significantly upregulated upon differentiation, as was
171 determined by RNA-Seq analysis. Fold-changes were highly correlative between RNA-Seq and RT-qPCR
172 (Sup 4E). To enable us to focus on lncRNAs with potential neuronal functions we selected candidate
173 lncRNAs that exhibited the highest fold increase in levels upon differentiation. The majority of these
174 candidate lncRNAs (6/7) are specifically enriched in the cytoplasm, rather than the nucleus (Fig 3E,
175 Sup 4F) in contrast to the known nuclear lncRNA Xist. Only 2/7 lncRNAs tested showed changes in
176 cytoplasmic enrichment upon differentiation (Sup 4G), the majority do not (5/7). When these
177 candidate lncRNAs were profiled across sucrose gradient fractions they were found to associate with
178 polysome complexes within the cytoplasm. LINC02143, which is induced >22-fold during
179 differentiation (Fig 3E), is highly enriched in the cytoplasm compared to the nucleus (Fig 3F) and is
180 found in monosomes and small polysomes (Fig 3G). It is also just as upregulated in polysomes (Fig 3A)
181 as in the total cytoplasm (Sup 3B) during differentiation. Unlike DLGAP1-AS2, DLGAP1-AS1 is enriched
182 in the cytoplasm (Fig 3F) and associates with small polysomes, as well as ribosomal subunits (Sup 4H).
183 Another lncRNA whose levels significantly increase during differentiation is LINC01116 (Fig 3
184 A&E), which is involved in the progression of glioblastoma (GBM) (Brodie et al., 2017). LINC01116 is
185 enriched in the cytoplasm (Fig 3F), detected at high levels in the 80S (monosome) fraction and in small
186 polysomal complexes (Fig 3H). Upon differentiation there is an increase in the amount of LINC01116
187 present in disomes, compared to Control. This is consistent with the upregulation of LINC01116
188 transcript in the polysomes detected by RNA-Seq, indicating a functional interaction of LINC01116
189 with polysomes during differentiation. In fact, the majority of LINC01116 transcript was found to
190 associate with polysomes in both undifferentiated (Control) cells (66%) and upon differentiation (RA-
191 57%), suggesting it is could either be translated or associating with translation complexes (Fig 3H).
192
193 **Translation of lncRNA-smORFs during differentiation**
194 To better understand the association of lncRNA with polysome complexes and their potential
195 translation we analysed ribosome footprinting from our Poly-Ribo-Seq experiments (Fig 2A). Framing
196 analysis reveals good framing, specifically at footprint lengths of 31 and 33 nt (Fig 4A) and discrete
197 footprint lengths (Sup 5A). On average 95% of ribosome footprints mapped to CDSs, whilst in RNA-
198 Seq this was only 53% (Sup 5B). Metagene analysis of 31 nt length reads indicates that framing is high
199 within ORFs, and that the footprint signal outside of ORFs is mostly within 5'-UTRs rather than 3'-UTRs,
200 where there is a sharp drop off in footprints at the stop codon (Sup 5C). Together these attributes
201 indicate that our Ribo-Seq quality is high and represents genuine translation events. Since 31 and 33nt
202 reads exhibited high triplet-periodicity they were selected for downstream analysis of translation
203 events to identify ORFs that are translated (Fig 4B). Overall, we are able to detect 16,282 protein-

204 coding ORFs translated in Control and 12,745 in RA, 10,014 of which are translated in both conditions
205 (Table 1). Interestingly we are also able to detect the translation of 71 uORFs and 5 dORFs (Table 1).
206
207 To monitor translational changes we performed differential translation analysis (Chothani et
208 al., 2019). Differential ribosome occupancy on ORFs (ΔRPF), differential mRNA expression (ΔRNA) and
209 differential translation efficiency (ΔTE) were calculated (Chothani et al., 2019). The majority of
210 changes in protein production are driven by transcriptional regulation during differentiation, with no
211 change in TE (Fig 4C). We detected 3 protein-coding CDSs being exclusively regulated by translation,
212 upon differentiation (Fig 4C). One of these is centromere protein S (CENPS), which is expressed at low
213 levels in neuroblastoma tumours and is thought to reduce cell growth (Carén et al., 2005; Krona et al.,
214 2004). We were also able to identify CDSs that are known to be regulated during differentiation such
215 as cytochrome P450 26A1 and cytochrome P450 26B1 (CYP26A1, CYP26B1), involved in RA
216 metabolism (Maden, 2007) (Fig 4D). These ORFs are transcriptionally and translationally upregulated,
217 and their TE is significantly increased upon differentiation.
218 Importantly, using our pipeline (Fig 4B) we also detect the translation of 45 ORFs within
219 lncRNAs, 28 in control and 23 during differentiation, 6 of these were translated in both conditions (Fig
220 4E). Interestingly, TEs for these lncRNA-smORFs are similar to those of protein-coding ORFs (Fig 4F),
221 providing further evidence that these are genuine translation events. The dORFs that we detect are
222 translated at much lower efficiencies (Fig 4F). This makes sense considering that ribosomes will likely
223 have to reinitiate after the main ORF, which will occur at a low efficiency. One of the lncRNA-smORFs
224 we identify to be translated in differentiated cells, has previously been characterised, the 84 aa
225 CRNDEP (Szafron et al., 2015). Analysis of these 45 translated smORFs from lncRNAs shows that they
226 are <300 aa in length with a median size of 56 aa in Control and 64 aa in RA (Fig 4G, Sup 5D).
227 Previous analysis has indicated that smORF peptides exhibit specific amino acid usage, which
228 suggests they are both genuine proteins and show enrichment for transmembrane alpha-helices (Song
229 et al., 2010). Therefore, we profiled the amino acid composition of our lncRNAs-smORFs, uORFs and
230 dORFs compared to protein-coding ORFs and expected frequency (Fig 4H). lncRNA-smORFs exhibit
231 similar frequencies to known protein-coding ORFs. Specifically, smORFs possess lower than expected
232 arginine levels, but not as low as known protein-coding ORFs. Amino acid usage does not suggest that
233 any smORF groups have propensity to form transmembrane alpha-helices (Aspden et al., 2014).
234 From within the set of lncRNAs we identified as induced during differentiation, several
235 contained translated smORFs. One of these is LINC01116, whose translated ORF exhibits high framing
236 (Fig 4I). In fact, ~80% of reads that map to this ORF are in frame 2, whereas outside this ORF, the few
237 reads mapping to the remaining lncRNA sequence are far more equally distributed between the three

238 possible frames (Fig 4I). This smORF is 71-codons, and its start codon is 608nt into the lncRNA.
239 Ribosome profiling signal is substantially higher upon differentiation, mainly as a result of increased
240 lncRNA transcript abundance (Sup 5E).
241
242 **Peptide synthesis from smORFs in lncRNAs during differentiation**
243 Our pipeline is highly stringent, i.e., there are many additional ORFs that display good framing
244 but do not pass our thresholds for numbers of footprinting reads or exhibit background signal in rest
245 of the lncRNA. Therefore, we are confident these translation events are taking place. To validate
246 peptide synthesis, we have taken two complementary strategies; mass spectrometry analysis and
247 transfection of ORF tagging constructs. Mass spectrometry generally only detects 90% of translated
248 proteins from protein-coding genes (Zubarev, 2013). Analysis of previously published mass
249 spectrometry datasets from SH-SY5Y (undifferentiated and RA-treated) cells (Brenig et al., 2020;
250 Murillo et al., 2018) supports the production of peptides from 6 of our translated uORFs (2 Control, 4
251 RA) and 8 lncRNA-smORFs (4 Control and 4 RA) (Fig 5A). This represents ~8% (uORFs) and 18% (lncRNA-
252 smORFs) of what is detected in Poly-Ribo-Seq. This is to be expected given the small size of these
253 peptides and therefore reduced chance of producing peptides >8aa from digestion for mass
254 spectrometry detection (Saghatelian & Couso, 2015).
255 To validate translation of our lncRNA-smORFs that were not identified in previous mass
256 spectrometry but were detected as translated by our Poly-Ribo-Seq analyses, we have also used a
257 transfection tagging approach. We cloned the lncRNA sequence to the 5' of the putative ORF, termed
258 the 5'-UTR and the smORF itself, without its stop codon, with a C terminal 3x FLAG tag (Fig 5B). The
259 start codon of this FLAG tag is deleted, so any FLAG signal is the result of translation from the lncRNA-
260 smORF. Two candidate lncRNA-smORFs were selected that did not have mass spectrometry support;
261 LINC000478 and LINC01116. A 37 codon smORF was detected as translated from LINC00478 in both
262 conditions (Fig 5C). Transfection of LINC00478-smORF-FLAG into undifferentiated SH-SY5Y cells
263 produced FLAG signal in both nuclear and cytoplasmic compartments (Fig 5D & Sup 6A). FLAG signal
264 was also seen when LINC00478-smORF-FLAG transfected SH-SY5Y cells were treated with RA (Sup 6A).
265 Interestingly this RA FLAG signal was only ever detected in the nucleus. Similar results were seen in
266 HEK293 cells (Sup 6B), but because of the higher transfectional efficiency in HEK293 compared with
267 SH-SY5Y cells, we detect FLAG signal in more cells.
268 Poly-Ribo-Seq detected increased levels of LINC01116 during differentiation (4-fold in Control
269 vs RA polysome) and translation of a 71-codon smORF during differentiation (Fig 4I). Tagging of
270 LINC01116-smORF (Fig 5E) generated FLAG signal in the cytoplasm of SH-SY5Y cells, which is localized
271 to neurites (Fig 5F), suggesting it could play a role in differentiation. FLAG signal was also present in

272 LINC01116 transfections in HEK293 cells (Sup 6C), but because of the higher transfectional efficiency
273 in HEK293 compared with SH-SY5Y cells, we detect FLAG signal in more cells.
274 LINC01116-smORF has two potential ATG start codons (Fig 5E and Sup 6D). When these start
275 codons were assessed for similarity to the Kozak sequence consensus, using NetStart1.0 (Pedersen &
276 Nielsen, 1997), both exhibited scores >0.5, indicating both are in good context and therefore either
277 could be used to initiate translation (AUG_1= 0.545, AUG_2= 0.645). Given the scanning model of
278 translation initiation it seems likely that the first AUG would be used. To determine if the first start
279 codon is actually used it was mutated (Fig 5E). No FLAG signal was present in transfections where the
280 5' start codon was mutated (Δ1) (Fig 5G). This suggests that the first start codon is necessary for the
281 translation of the LINC01116-smORF and the resulting peptide is 87aa long. Although FLAG signal is
282 present in a low number of cells, no transfection controls and Δ1 indicate that FLAG signal is
283 dependent on translation of the LINC01116-smORF (Sup 6E).
284
285 **Translated lncRNA-smORFs exhibit sequence conservation**
286 Another indicator of coding potential and also of peptide functionality is the sequence
287 conservation of our translated lncRNA-smORFs. To ensure detection of sequence conservation for
288 these short smORFs irrespective of annotation in other genomes, we used three complementary
289 BLAST strategies using transcript nt sequence, smORF nt sequence and protein aa sequence.
290 Initial searches using the entire lncRNA transcript sequence (nt) and BLASTn (Altschul et al.,
291 1990), returned results for ~78% of translated lncRNAs, many of which had short alignment lengths of
292 30-100 nt. Although some of these results may represent conservation of the smORFs, many are due
293 to small areas of sequence overlap along the rest of the lncRNA. This is a well-documented issue when
294 investigating lncRNAs; they rarely exhibit the same levels of conservation as mRNAs (Johnsson et al.,
295 2014). A more common finding is that there are "modules" of higher sequence conservation across
296 lncRNA transcripts, as described for Xist lncRNA (Brockdorff, 2018). To accommodate this, a second
297 round of searches were performed, on the initial search results, using the nt sequence of the smORF
298 (BLASTn) (Altschul et al., 1990). This filtering step resulted in fewer, higher quality results, with 14 of
299 18 remaining lncRNA-smORFs passing manual cross validation. For the majority of the these lncRNA-
300 smORFs, conservation is high across the body of the smORF, and drops off across the rest of the
301 transcript. The ENST00000442428.1 (AL162386.2)-smORF exhibits high sequence conservation when
302 compared to gorilla and orangutan (*Pongo abelii*), with 100% and 99% ORF nt sequence identity
303 respectively (Figure 5H). When entire transcripts are aligned, this percentage sequence identity drops
304 to 74% with gorilla (ENSGGOT00000060708.1), and 65% with orangutan (ENSPPYT00000022401.2),
305 indicating the ORF is the most conserved part of these transcripts.

306 To further corroborate these results, a tBLASTn (Altschul et al., 1990) search of the lncRNA-
307 smORF aa sequences was performed. By translating transcript databases into aa in all 6 frames,
308 tBLASTn removes the noise of synonymous substitutions, which can have a significant effect,
309 particularly in smORFs (Ladoukakis et al., 2011). For the majority of smORFs, the same results were
310 returned, and evidence of conservation was found for a further three lncRNA-smORFs
311 (ENST00000454935.1_477_633, ENST00000557660.5_42_186, ENST00000453910.5_151_262) that
312 appear to have undergone some frameshift mutations.
313 At the aa sequence level, ~18% of the translated lncRNA-smORFs returned homologous
314 proteins from BLASTp searches (Altschul et al., 1990). All the returned proteins were unreviewed,
315 uncharacterised proteins, with no evidence at protein, transcript or homology levels in the Uniprot
316 database (Bateman et al., 2019). This potentially suggests that automated annotation pipelines
317 recognised the coding potential of these smORFs, unlike in the more curated human genome
318 annotation. Overall, the combination of these 3 layers of sequence conservation analysis reveal that
319 11 of our translated lncRNA-smORFs exhibit sequence conservation across *Hominidae*, 3 addition
320 smORFs are also found in gibbons (Fig 5I), with evidence for 2 translated smORFs detectable at the
321 greater evolutionary distance of human to mouse, but not found in all apes (Sup 6E).
322
323 **LINC01116 contributes to neuronal differentiation**
324 To dissect the potential role of LINC01116 and its translation during differentiation we
325 knocked down LINC01116 using a siRNA pool in both undifferentiated and differentiated SH-SY5Y cells
326 (Fig 6A). LINC01116 knockdown initially had a limited effect on cell viability, which recovered to no
327 effect after 2 days (Sup 7A). Interestingly, LINC01116 knockdown resulted in a significant reduction of
328 neurite length in RA treated SH-SY5Y cells (Fig 6B, zoom in Fig 6C), compared to scrambled siRNA
329 treated SH-SY5Y (Fig 6D). There was no effect of the knockdown in undifferentiated cells (Fig 6D). This
330 suggests that LINC01116 is involved in the regulation of neuritic processes formation. To examine
331 potential effects of LINC01116 knockdown further on differentiation we assessed the expression levels
332 of the noradrenergic marker MOXD1, which is important in in neural crest development. LINC01116
333 siRNA knockdown, upon differentiation resulted in a reduction of MOXD1 expression levels, further
334 indicating a role of LINC01116 in neuronal differentiation (Fig 6E). However, LINC0116 knockdown had
335 no effect on proliferation, as measure by % of Ki67+ cells (Sup 7B & Fig 6F) or cell cycle, as measured
336 by E2F1 mRNA RT-qPCR (Fig 6G). LINC01116 likely functions early in the differentiation pathway since
337 its levels are significantly upregulated within the first 24 hours of RA induced differentiation (Sup 7C).
338 Expression of LINC01116 then declines rapidly by day 8 (Fig 7D). Together these results suggest that
339 LINC01116 functions during early differentiation, contributing to neurite process formation.

340

341

342 Discussion

343 Cytoplasmic lncRNAs have recently emerged as important regulators of several cellular
344 processes, including the regulation of translation, in a variety of tissues (Carrieri et al., 2012; Dimartino
345 et al., 2018). Moreover, the active translation of smORFs from within lncRNAs has been characterized
346 in a range of organisms (Anderson et al., 2015; Aspden et al., 2014; Chen et al., 2020; Magny et al.,
347 2013; Pueyo & Couso, 2008; Wang et al., 2020). Translational regulation is a key process during
348 neuronal differentiation (Blair et al., 2017; Chau et al., 2018), therefore we reasoned that lncRNAs
349 may functionally interact with polysomes during differentiation.

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

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

374 Differential translation analysis revealed that upon RA induced neuronal differentiation of SH-
375 SY5Y for 3 days, a small number of ORFs are exclusively translationally regulated. For the majority of
376 ORFs, translation regulation is driven by transcription regulation. This result indicates that at this early
377 timepoint, transcriptional control, rather than translational control is the driving force of SH-SY5Y
378 differentiation. This is not surprising, because according to a previous study of human embryonic stem
379 cells (hESCs) differentiation into neural progenitor cells (NPCs) only 12% of the detected transcripts
380 were exclusively translationally regulated (Blair et al., 2017).
381 Ribosome profiling of the actively translating polysomes allowed us to distinguish between
382 the lncRNAs that associate with the polysome complexes and those that are being actively translated.
383 We were able to identify 45 translated lncRNA-smORFs in undifferentiated and differentiated cells.
384 Only ~5% of polysome-associated lncRNAs pass our stringent thresholds for translation. The translated
385 lncRNA-smORFs exhibit high levels of triplet periodicity and their translational efficiencies are similar
386 to those of protein-coding genes. We can therefore be confident that these are real translation events
387 leading to the production of substantial peptide levels rather than background, spurious events
388 (Bazzini et al., 2014; Guttman et al., 2013; Patraquim et al., 2020; Ruiz-Orera & Alba, 2019). The size
389 distribution of our novel translated ORFs, indicates that the majority are indeed smORFs (<100aa). The
390 general pattern we identified is that dORFs>lncRNA-smORF>uORFs in size. This is consistent with
391 previous studies where a wide range of peptides lengths were discovered (Aspden et al., 2014; Chong
392 et al., 2020). Amino acid composition of these translated smORFs supports the fact they are translated
393 into peptides. However, it does not suggest they are enriched for transmembrane alpha-helices. This
394 is in contrast to the smORFs characterized in *D. melanogaster* (Aspden, 2014).
395 Overall, we have independent evidence for peptide synthesis for 12/45 lncRNA-smORFs. 8 of
396 these from published mass spectrometry data from Control and RA-differentiated cells SH-SY5Y cells
397 (Brenig et al., 2020). In general, we find smORFs translated in the same treatment as these mass
398 spectrometry datasets detect peptides (7/8). An 18% mass spectrometry detection level may seem
399 low but given the limitations of detecting small peptides by mass spectrometry this represents a
400 substantial level of validation. Two translation events were validated by FLAG tagging transfection
401 assay: LINC01116 and LINC00478 lncRNA-smORFs. The production of 2/45 lncRNA-smORF peptides is
402 corroborated by previous studies in non-neuronal cells. HAND2-AS1 (translated in Control and RA) is
403 translated in human and rodent heart and encodes for an integral membrane component of the
404 endoplasmic reticulum (van Heesch et al., 2019). CRNDE, which is only translated upon differentiation,
405 encodes for a previously characterised nuclear peptide (CRNDEP) (Szafron et al., 2015). The translation
406 of these smORFs in multiple cell types provides substantial support for the production of peptides and
407 their potential function.

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

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

433 Our findings indicate that many lncRNAs are localised in the cytoplasm and play important
434 roles here. Given the large number of lncRNAs associated with polysomes we anticipate that many
435 more lncRNAs will be found to regulate translation. We have identified 121 novel translation events,
436 many of which are regulated during differentiation. Only ~5% of polysome-associated lncRNAs are
437 translated. The lncRNA-smORFs we discover here represent a general population whose products
438 have not yet been characterized. As we have found for LINC01116, lncRNAs and the small peptides
439 encoded therein have the potential to contribute to important cellular functions and contribute to
440 development and disease.
441

442 Materials and Methods

443 **Cell culture**

444 Human neuroblastoma SH-SY5Y cells, were cultured in Dulbecco's Modified Eagle Medium
445 (DMEM 4.5g/L Glucose with L-Glutamine) supplemented with 1% (v/v) Penicillin/Streptomycin and
446 10% Fetal Bovine Serum (FBS) at 37°C, 5% CO_2. Neural induction commenced at passage 4 and was
447 performed as described previously (Forster et al., 2016; Korecka et al., 2013) with minor alterations.
448 All trans Retinoic Acid (RA, Sigma) was added to cells 24h after plating, at final concentration of 30µM.
449

450 **Immunocytochemistry**

451 Cells were seeded on Poly-D-Lysine/mouse laminin coated 12mm round coverslips (Corning
452 BioCoat™ Cellware) and fixed with 4% paraformaldehyde (PFA) (Affymetrix) for 20 min at room
453 temperature (RT). A permeabilization step (0.1% Triton-X for 10 min at RT) was performed prior to
454 blocking, followed blocking at RT in blocking buffer (3% BSA, 1XPBS or 5% NGS, 1XPBS and 0.1% Triton-
455 X) for 30 min. Primary antibodies (Antibody Table) were applied in (3% BSA 1X PBS or 0.5% NGS, 1X
456 PBS, 0.1% Triton-X and incubated at RT for 2h or at 4° C overnight. Cells were washed and labelled
457 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-
458 X. Cells were mounted in VECTASHIELD mounting medium, analysed using LSM 700 confocal
459 microscope (Zeiss) ImageJ.
460

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

462 Equal amounts of RNA (whole cell, nuclear and cytoplasmic lysates) or equal volumes
463 (polysome fractions) were subject to cDNA synthesis, using qScript (Quanta Bio) according to
464 manufacturer's instructions. qPCR was performed using the CFX Connect™ thermal cycler and
465 quantification using SYBR Green fluorescent dye (PowerUp™ SYBR Green Master Mix, Thermo Fisher
466 SCIENTIFIC). Primers were designed to anneal to exon-exon junctions, where possible, or to common
467 exons between alternative transcripts (Primer Table). Target mRNA and lncRNA levels were assessed
468 by absolute quantification, by the means of standard curve or relative quantification, using the ΔΔCq
469 method.
470

471 **Polysome Profiling**

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

476 $MgCl_2$, 1mM DTT, 1% IGEPAL, 100µg/ml cycloheximide, Turbo DNase 24U/µL (Invitrogen), RNasin Plus
477 RNase Inhibitor 90U (Promega), cOmplete Protease Inhibitor (Roche), for 45 min. Cells were then
478 subjected to centrifugation at 17,000 xg for 5 min, to pellet nuclei. Cytoplasmic lysates were loaded
479 onto 18%-60% sucrose gradients (~70 $x10^6$ cells per gradient) at 4°C and subjected to
480 ultracentrifugation (121,355 x g_{avg} 3.5h, 4°C) in SW-40 rotor. Gradients were fractionated using
481 GRADIENT STATION (Biocomp) and absorbance at 254 nm was monitored using a Biorad detector.
482
483 **Poly-Ribo-Seq**
484 ~20% of cytoplasmic lysate was kept for PolyA selection (Total RNA control) and ~80% was
485 loaded onto 18%- 60% sucrose gradients (~$70x10^6$ cells per gradient) at 4°C and subjected to
486 ultracentrifugation (121,355 x g_{avg} 3.5h, 4°C) in SW-40 rotor. Polysome fractions were pooled from
487 control and from differentiated cells. ~25% polysomes were kept for polyA selection (Polysome-
488 associated RNA). The remaining 75% was diluted in 100mM Tris-HCl pH8, 30mM NaCl, 10mM $MgCl_2$.
489 RNaseI (EN601, 10U/µl 0.7-1U/million cells) was subsequently added incubated overnight at 4°C.
490 RNaseI was deactivated using SUPERase inhibitor (200U/gradient) for 5min at 4°C. Samples were
491 concentrated using 30 kDa molecular weight cut-off columns (Merck) and loaded on sucrose cushion
492 (1M sucrose, 50mM Tris-HCl pH8, 150mM NaCl, 10mM $MgCl_2$, 40U RNase Inhibitor) and subjected to
493 ultracentrifugation at 204,428 x g_{avg} at 4°C for 4h (TLA110). Pellets were resuspended in TRIzol
494 (Ambion, Life Technologies) and processed for RNA purification.
495 RNA purification from cytoplasmic lysates and RNaseI footprinted samples was performed by
496 Trizol RNA extraction, following manufacturer's instructions. RNA purification from polysome
497 fractions was performed by isopropanol precipitation, followed by TURBO DNase treatment
498 (Thermofisher) (according to manufacturer's instructions), acidic phenol/chloroform RNA purification
499 and ethanol precipitation at -80°C overnight. RNA concentration was determined by Nano-drop 2000
500 software. Two rounds of polyA selection from total cytoplasmic lysate and polysome fractions were
501 performed using oligo (dT) Dynabeads (Invitrogen) according to manufacturer's instructions. PolyA
502 RNA was fragmented by alkaline hydrolysis. 28–34 nt ribosome footprints and 50–80 nt mRNA
503 fragments were gel purified in 10% (w/v) polyacrylamide-TBE-Urea gel at 300V for 3.5h in 1X TBE.
504 Ribosome footprints were subjected to rRNA depletion (Illumina RiboZero rRNA removal kit).
505 5' stranded libraries were constructed using NEB Next Multiplex Small RNA Library Prep.
506 Resulting cDNA was PCR amplified and gel purified prior to sequencing. Libraries were subjected to
507 75bp single end RNA Seq using NextSeq500 Illumina sequencer, High Output Kit v2.5 (75 Cycles) (Next
508 Generation Sequencing Facility, Faculty of Medicine, University of Leeds).
509
510 **RNA-Seq data analysis**

511 RNA-Seq reads were trimmed with Cutadapt (Martin, 2011) and filtered with
512 fastq_quality_filter (Hannon, 2010) to filter out the reads of low quality (90% of the read to have a
513 phred score above 20). Filtered reads were mapped (Liao et al., 2013) to human genome reference
514 (the lncRNA GENCODE (Frankish et al., 2019) annotation added to mRNA annotation from UCSC
515 (Haeussler et al., 2019) human genome assembly (hg19) from iGenomes) with Rsubread (Liao et al.,
516 2013) and uniquely mapped reads were reported. Bam file sorting and indexing was performed with
517 SAMtools (Li et al., 2009). Subsequently summarised read counts for all genes were calculated using
518 featureCounts (Liao et al., 2014). For normalization, RPKM values were calculated. Differential
519 expression analysis was conducted between with DESeq2 (Love et al., 2014) based on the two cut-offs
520 $p^{adj} < 0.05$ and the absolute value of $\log^2$FoldChange > 1. Gene ontology analysis was performed with
521 GOrilla (***G****ene* ***O****ntology en****RI****chment ana****L****ysis and visua****L****iz****A****tion too*l) (Eden et al., 2009).
522
523 **Cytoplasmic/Nuclear fractionation of SH-SY5Y cells.**
524 Cells were harvested (as above) and washed with 1X PBS. Cells were lysed in whole cell lysis
525 buffer (see table) (500µL buffer per 10^6 cells) on ice for 30min. Whole cell lysate aliquots were
526 removed and remainder subjected to centrifugation at 1,600 x g for 8 min to pellet nuclei. Nuclear and
527 cytoplasmic fractions were subjected to two further clearing steps by centrifugation (3,000 x g and
528 10,000 x g respectively). Nuclei were lysed in RIPA buffer (table). ~10% of both nuclear and cytoplasmic
529 lysates were used for Western Blot and ~90% subjected to RNA extraction (ZYMO R1055).
530
531 **Western Blot**
532 Samples were diluted in 4X Laemmli sample buffer (Biorad) (277.8 mM Tris-HCl, pH 6.8, 4.4%
533 LDS, 44.4% (v/v) glycerol, 0.02% bromophenol blue), 5% β-mercaptoethanol (Sigma) was added prior
534 to heating at 95°C for 5 min and loaded on 10% SDS gels. Gel electrophoresis was performed using the
535 BioRad Mini-PROTEAN 3 gel electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA).
536 Proteins were transferred to nitrocellulose membranes (Amersham™ Protran™) and blocked with 5%
537 fat-free milk powder in 1XPBS, 0.05% Tween-20 (Sigma) for 1h at RT. Blots were incubated with
538 primary antibodies overnight (Antibody Table). Blots were then washed in PBS-T and incubated with
539 secondary antibody (anti-mouse HRP) at RT for 2h. Membranes were washed 3 times with PBS-T, prior
540 to application of ECL (Biological Industries). Chemiluminescent signal was detected with Chemi-Doc
541 (BIO-RAD). All membranes were probed for β-tubulin as loading control.
542
543 **Ribo-Seq analysis**

544 Quality reports of Polysome-associated RNA-Seq and Ribo-Seq data were made using Fastqc
545 (v.0.11.8) (Andrews, 2010). Adaptor sequences were trimmed using Cutadapt (v.1.81) (Martin, 2011)
546 with minimum read length of 25bp, and untrimmed outputs retained for Ribo-Seq reads. Low-quality
547 reads (score <20 for 10% or more of read) were then discarded using FASTQ Quality Filter, FASTX-
548 Toolkit (v.0.0.14) (Gordon, 2010). Human rRNA sequences were retrieved from RiboGalaxy (Michel et
549 al., 2016) and high confidence hg38 tRNA sequences from GtRNAdb Release 17 (Chan & Lowe, 2016).
550 One base was removed from 3' end of reads to improve alignment quality, reads originating from
551 rRNA and tRNA were aligned and removed using Bowtie2 (v.2.3.4.3) (Langmead & Salzberg, 2012).
552 The splice aware aligner STAR (v2.6.1b) (Dobin et al., 2012) was used to map remaining reads
553 to human reference genome Release 30 (GRCh38.p12), from Gencode (Frankish et al., 2019). The STAR
554 (v2.6.1b) (Dobin et al., 2012) genome index was built with a sjdbOverhang of 73. Samtools (v.1.9) (Li
555 et al., 2009) was used to create sorted, indexed bam files of the resulting alignments.
556 Metaplots of aligned Ribo-seq data were generated using create_metaplots.bash script from
557 Ribotaper (v1.3) (Calviello et al., 2016) pipeline. These show the distance between the 5' ends of Ribo-
558 seq and annotated start and stop codons from CCDS ORFs, allowing the locations of P-sites to be
559 inferred. Read lengths exhibiting the best triplet periodicity were selected for each replicate, along
560 with appropriate offsets (Sup 5).
561 Actively translated smORFs were then identified using Ribotaper (v1.3) (Calviello et al., 2016).
562 Initially, this requires an exon to contain more than 5 P-sites in order to pass to quality control steps.
563 Identified ORFs were then required to have a 3-nt periodic pattern of Ribo-seq reads, with 50% or
564 more of the P-sites in-frame. In the case of multiple start codons, the most upstream in-frame start
565 codon with a minimum of five P-sites in between it and the next ATG was selected.
566 ORFs for which >30% of the Ribo-seq coverage was only supported by multimapping reads were
567 also subsequently filtered. For a smORF to be considered actively translated in a condition, we
568 required that it be identified in at least two of the three biological replicates for the condition.
569
570 **Translation efficiency calculation**
571 Translational efficiency (TE) was estimated for all translated ORFs in each condition, where TE
572 was equal to the mean number of P sites per ORF, normalised by the median P sites per ORF per
573 replicate, divided by the mean number of RNA sites per ORF, normalised by the median RNA sites per
574 ORF per replicate.
575
576 **Differential translation efficiency analysis**

577 Differential translation efficiency analysis was performed using the ΔTE (Chothani et al., 2019)
578 alternate protocol. However, our analysis was based on read counts from all identified translated ORFs
579 from the Ribo-seq and polysome-associated RNA-seq, not the whole CDS as described in the protocol.
580
581 **Amino acid composition**
582 For each of our ORF sets (protein coding, lncRNA-smORF, uORF and dORFs), the average amino
583 acid compositions were calculated. Random control expected frequencies were taken from King and
584 Jukes (King & Jukes, 1969).
585
586 **Mass spectrometry analysis**
587 Two published SH-SY5Y cell mass proteomics datasets were analysed: PXD010776 (Murillo et
588 al., 2018) and PXD014381 (Brenig et al., 2020). Binary raw files (*.raw) were downloaded from PRIDE
589 then converted to human-readable MGF format using ThemoRawFileParser (Hulstaert et al., 2020).
590 The amino acid sequences of our translated uORFs, dORFs and lncRNA-smORFs were added to the
591 whole *Homo sapiens* proteome dataset (20,379 entries) downloaded from UniProtKB (Bateman et al.,
592 2019) on Nov/2019. The new FASTA file was then used as a customized database on Comet
593 (v2019.01.2) (Eng et al., 2013) search engine runs that scanned all MS/MS files (*.mgf) against it.
594 Default settings in Comet were used with the following exceptions according to the MS/MS
595 data type. iTRAQ-4plex (PXD010776): decoy_search = 1, peptide_mass_tolerance= 10.00,
596 fragment_bin_tol = 0.1, fragment_bin_offset = 0.0, theoretical_fragment_ions = 0,
597 spectrum_batch_size = 15000, clear_mz_range = 113.5-117.5, add_Nterm_peptide = 144.10253,
598 add_K_lysine = 144.10253, minimum_peaks = 8. Label-free (PXD014381): decoy_search = 1,
599 peptide_mass_tolerance = 10.00, fragment_bin_tol = 0.02, fragment_bin_offset = 0.0,
600 theoretical_fragment_ions = 0, spectrum_batch_size = 15000.
601 CometUI (Eng et al., 2013) was employed for assessing Comet's *.xml output files and setting
602 a false discovery rate (FDR) threshold of 10% per peptide identification. This FDR threshold was
603 selected due to expected low abundance levels of the target smORFs.
604
605 **Conservation analysis**
606 Protein, cDNA and ncRNA sequence data for *H.sapiens, P.abelii, P.paniscus, P.troglodytes,*
607 *G.gorilla, N.leucogenys,* and *M.musculus* were obtained from Ensembl (release 100, (Yates et al.,
608 2020).
609 A sequence homology search was performed using the 45 translated lncRNA peptide
610 sequences, against each non-human species protein database using BLASTp (e value - 0.001) (Altschul

611 et al., 1990). As these are small peptides, results were filtered to remove anything with <75% identity,
612 unless a result(s) was the lowest e value hit for a given query in each species. Results were returned
613 for 12 lncRNA peptides, and manually cross validated using Ensembl Genome Browser and multiple
614 sequence alignments performed using ClustalOmega (Sievers et al., 2011), run using default
615 parameters using the msa package in R (Bodenhofer et al., 2015), to give 8 peptides with evidence of
616 sequence conservation.
617 The transcript sequences of the 45 translated lncRNAs were searched against transcriptome
618 databases created by combining the cDNA and ncRNA data for each species, using BLASTn (e value -
619 0.001) (Altschul et al., 1990). Results of this BLAST were used to filter the initial BLAST databases. ORF
620 portions of the 45 translated lncRNAs were extracted, and searched against these filtered databases
621 using BLASTn (e value - 0.001)(Altschul et al., 1990).
622 It was confirmed that all the lncRNA ORFs returned their genes of origin in *H.sapiens*. The
623 remaining species returned results for 18 of the lncRNA ORF queries. These were cross validated as
624 above, resulting in 14 lncRNA ORFs with evidence of sequence conservation based on transcript
625 sequences.
626 A final search was performed using the 45 translated lncRNA peptide sequences against the
627 transcriptome databases using tBLASTn (e value - 0.001) (Altschul et al., 1990). To reduce redundancy,
628 results were filtered to select the transcripts(s) with lowest e value for each gene. It was confirmed
629 that all ORFs returned their genes of origin in *H.sapiens*.
630 The remaining species returned results for 21 of the lncRNA peptide queries. As some queries
631 had many spurious results, they were further filtered to select the transcripts(s) with lowest e value
632 for each query in each species. These were cross validated as above, resulting in 16 lncRNA peptides
633 with evidence of sequence conservation based on transcript sequences.
634 We combined evidence from all these approaches into a final dataset consisting of 17 lncRNA
635 small ORFs with evidence of conservation.
636
637 **Plasmid construction/Cloning**
638 5'-UTRs and CDSs of putative smORFs (lacking stop codon) were generated by PCR (Table 3),
639 using NEB High Fidelity DNA Polymerase (Q5). C-terminal 3xFLAG tag was incorporated within the
640 reverse primer (Table 3 marked with green) by PCR and products were cloned into NheI and EcoRV
641 restriction sites (Table 3: marked with blue) of pcDNA3.1-hygro vector (Addgene, kindly provided by
642 Mark Richards-Bayliss group, University of Leeds). Start codon mutations were generated by site
643 directed mutagenesis (Q5 Site directed mutagenesis kit, NEB).
644

645 **Transfections and Microscopy**

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

650

651 **siRNA knockdown**

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

656

657 **General statistics and plots**

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

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

666

667 **Data availability**

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

669

670

671

672

673

674

675 Tables

676 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%
ddH2O	
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
ddH2O	
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
ddH2O	

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	GACTCCATGCTGCGGTTTTC
AC090001.1	GTGCCCATGAGGGAGAACAC	GACAAGAAGTCAGGAGGTAGACA
SNAP25-AS1	AGCCATGGAAGTCAAATGCTG	AGGCATTTTTGCTGTCTTTCCTC
XIST	GGCTCCTCTTGGACATTCTGAG	AGCTTGGCCAGATTCTCAAAG

681

682

683

Table 3: Cloning Primers

Transcript	Forward primer sequence	Reverse primer sequence
LINC001116 -smORF	GGCGGTGCTAGCGCGAGCCACGGGCCTC	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGTTTTTAAGCTGACTTGTC
LINC01116-whole	GGCGGTGCTAGCGCGAGCCACGGGCCTC	ACCGCCGATATCCAATTACTGCATTCACGTATTCTTC
LINC00478	GGCGGTGCTAGCCTCCTGTCGTTAAGATAAATTCTCCA	GGCGGTGATATCTCACTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCGTAGTAAAATGCTCTCTGA
LINC01116-mutant (ATG1)	CTTCTAAGAAAAGGTCTCACTCTGC	CAATTCAGTTGTCTTCTAATAC
LINC01116-mutant (ATG2)	GGCACCATCAAAGCTCACTGCAGC	ACCACACTCCAGCCTGGG
LINC01116-mutant (both ATG)	GCTGGAGTGTGGTGGCACCATCAAAGCTCACTGCAGCCTTGAA	CTGGGTGATGGCAGAGTGAGACCTTTTCTTAGAAGCAATTCAGTTGTC

697

698 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

699

700

701

702 **Acknowledgements**

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

709

710 **Author contributions**

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

717 the manuscript. JD helped design experiments, interpret portions of the data and critique the output
718 for important intellectual content, and revised the manuscript. AW helped design experiments,
719 interpret portions of the data and critique the output for important intellectual content, and revised
720 the manuscript. JLA conceived the work, interpreted data, drafted and revised the manuscript.
721

722 **Funding**
723 Katerina Douka's PhD was funded by Leeds Anniversary Research Scholarship (LARS). Julie Aspden is
724 funded by the University of Leeds (University Academic Fellow scheme). This work was funded by
725 MRC (MR/N000471/1).
726
727

728 References

729

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

Guo, H., Ingolia, N. T., Weissman, J. S., & Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, *466*(7308), 835--840. https://doi.org/10.1038/nature09267

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*, *154*(1), 240-251. https://doi.org/10.1016/j.cell.2013.06.009

Haeussler, M., Zweig, A. S., Tyner, C., Speir, M. L., Rosenbloom, K. R., Raney, B. J., Lee, C. M., Lee, B. T., Hinrichs, A. S., Gonzalez, J. N., Gibson, D., Diekhans, M., Clawson, H., Casper, J., Barber, G. P., Haussler, D., Kuhn, R. M., & Kent, W. J. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic acids research*, *47*(D1), D853--D858. https://doi.org/10.1093/nar/gky1095

Hannon, G. J. (2010). FASTX-Toolkit.

Hulstaert, N., Shofstahl, J., Sachsenberg, T., Walzer, M., Barsnes, H., Martens, L., & Perez-Riverol, Y. (2020). ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *Journal of Proteome Research*, *19*(1), 537-542. https://doi.org/10.1021/acs.jproteome.9b00328

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., & Weissman, J. S. (2013). Genome-Wide Annotation and Quantitation of Translation by Ribosome Profiling. 4.18.11--14.18.19. https://doi.org/10.1002/0471142727.mb0418s103

Johnsson, P., Lipovich, L., Grander, D., & Morris, K. V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function [Review]. *Biochimica Et Biophysica Acta-General Subjects*, *1840*(3), 1063-1071. https://doi.org/10.1016/j.bbagen.2013.10.035

King, J. L., & Jukes, T. H. (1969). NON-DARWINIAN EVOLUTION [Article]. *Science*, *164*(3881), 788-+. https://doi.org/10.1126/science.164.3881.788

Korecka, J. A., van Kesteren, R. E., Blaas, E., Spitzer, S. O., Kamstra, J. H., Smit, A. B., Swaab, D. F., Verhaagen, J., & Bossers, K. (2013). Phenotypic Characterization of Retinoic Acid Differentiated SH-SY5Y Cells by Transcriptional Profiling. *PLoS ONE*, *8*(5), e63862. https://doi.org/10.1371/journal.pone.0063862

Krona, C., Ejeskär, K., Carén, H., Abel, F., Sjöberg, R. M., & Martinsson, T. (2004). A novel 1p36.2 located gene, APITD1, with tumour-suppressive properties and a putative p53-binding domain, shows low expression in neuroblastoma tumours. *Br J Cancer*, *91*(6), 1119-1130. https://doi.org/10.1038/sj.bjc.6602083

Ladoukakis, E., Pereira, V., Magny, E. G., Eyre-Walker, A., & Couso, J. P. (2011). Hundreds of putatively functional small open reading frames in Drosophila [Article]. *Genome Biology*, *12*(11), 17, Article R118. https://doi.org/10.1186/gb-2011-12-11-r118

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357-U354. https://doi.org/10.1038/nmeth.1923

Lewandowski, J. P., Dumbović, G., Watson, A. R., Hwang, T., Jacobs-Palmer, E., Chang, N., Much, C., Turner, K., Kirby, C., Schulz, J. F., Müller, C.-L., Rubinstein, N. D., Groff, A. F., Liapis, S. C., Gerhardinger, C., Hubner, N., van Heesch, S., Hoekstra, H. E., Sauvageau, M., & Rinn, J. L. (2019). The Tug1 Locus is Essential for Male Fertility. *bioRxiv*, 562066. https://doi.org/10.1101/562066

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078--2079. 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-
986 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.

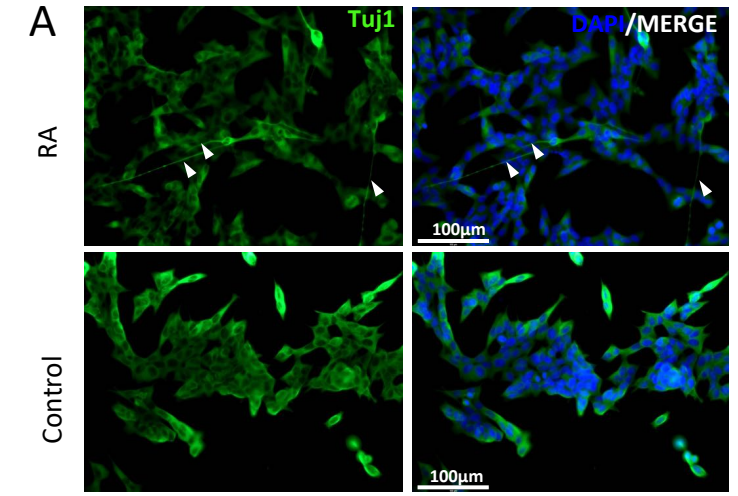
993 Pollen, A. A., Nowakowski, T. J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C. R.,
994 Shuga, J., Liu, S. J., Oldham, M. C., Diaz, A., Lim, D. A., Leyrat, A. A., West, J. A., &
995 Kriegstein, A. R. (2015). Molecular Identity of Human Outer Radial Glia during
996 Cortical Development. *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
1023 Spencer, H. L., Sanders, R., Boulberdaa, M., Meloni, M., Cochrane, A., Spiroski, A.-M.,
1024 Mountford, J., Emanueli, 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

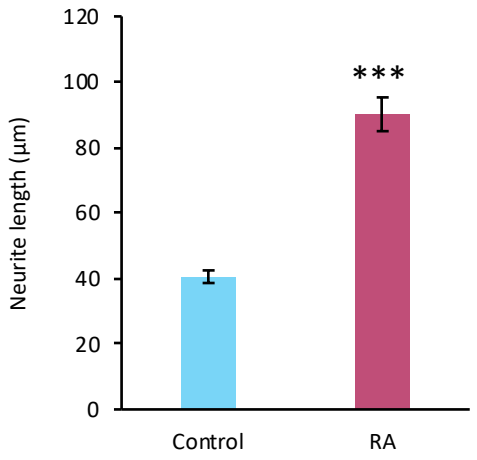
1038 Tsagakis, I., Douka, K., Birds, I., & Aspden, J. L. (2020). Long non-coding RNAs in
1039 development and disease: Conservation to mechanisms. *J Pathol*.
1040 https://doi.org/10.1002/path.5405
1041 van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., Kirchner,
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 lncR492 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.,

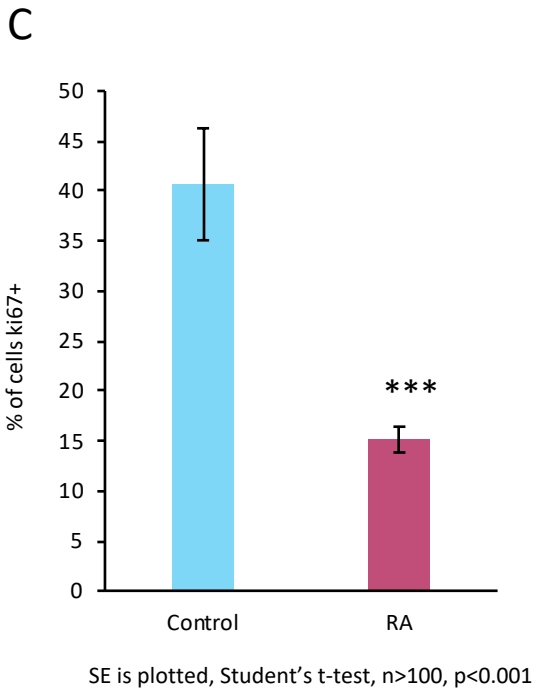
Howe, K. L., Zerbino, D. R., & Flicek, P. (2020). Ensembl 2020 [Article]. *Nucleic Acids Research*, *48*(D1), D682-D688. https://doi.org/10.1093/nar/gkz966

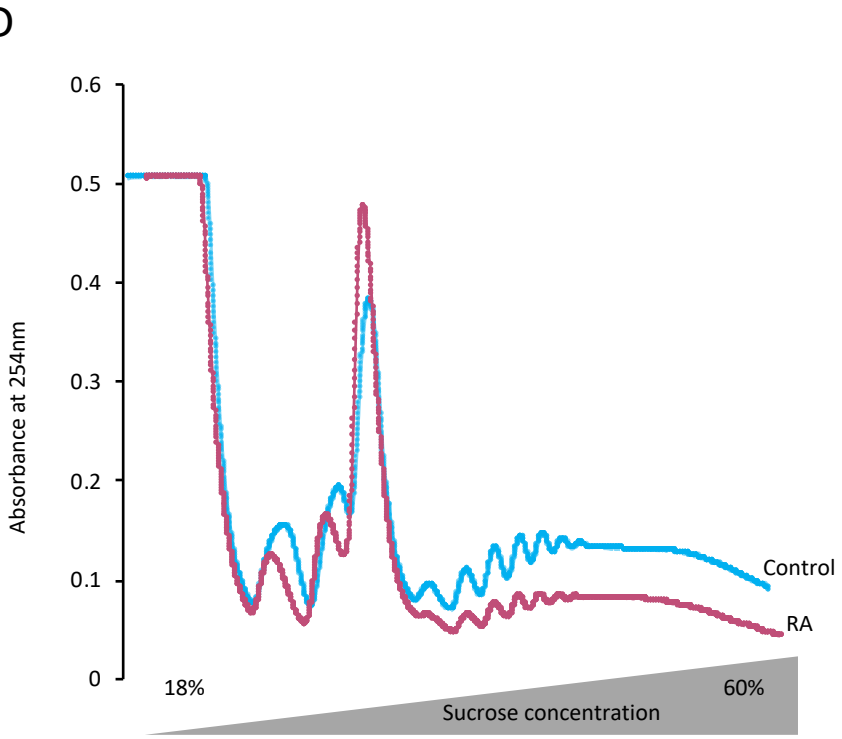
Zubarev, R. A. (2013). The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics*, *13*(5), 723-726. https://doi.org/10.1002/pmic.201200451

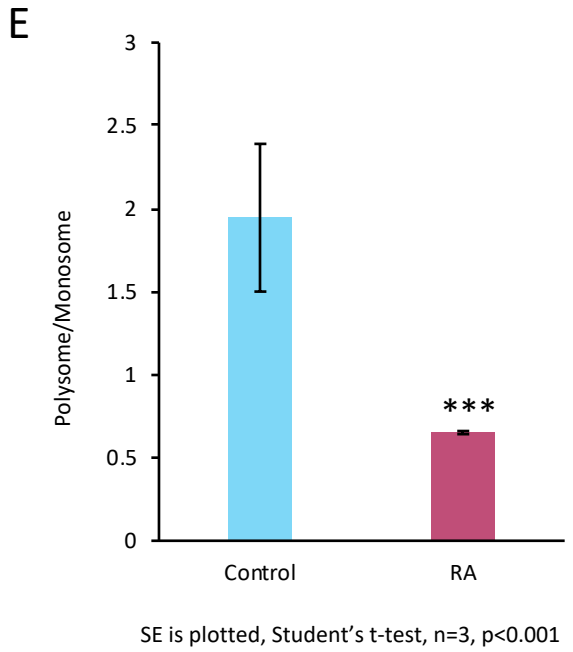
Figure 1


A
Tuj1
DAPI/MERGE
RA
Control
100µm
100µm


B
Neurite length (µm)
120
100
80
60
40
20
0
***
Control
RA


C
% of cells ki67+
50
45
40
35
30
25
20
15
10
5
0
***
Control
RA
SE is plotted, Student's t-test, n>100, p<0.001


D
Absorbance at 254nm
0.6
0.5
0.4
0.3
0.2
0.1
0
Control
RA
18%
60%
Sucrose concentration


E
Polysome/Monosome
3
2.5
2
1.5
1
0.5
0
***
Control
RA
SE is plotted, Student's t-test, n=3, p<0.001

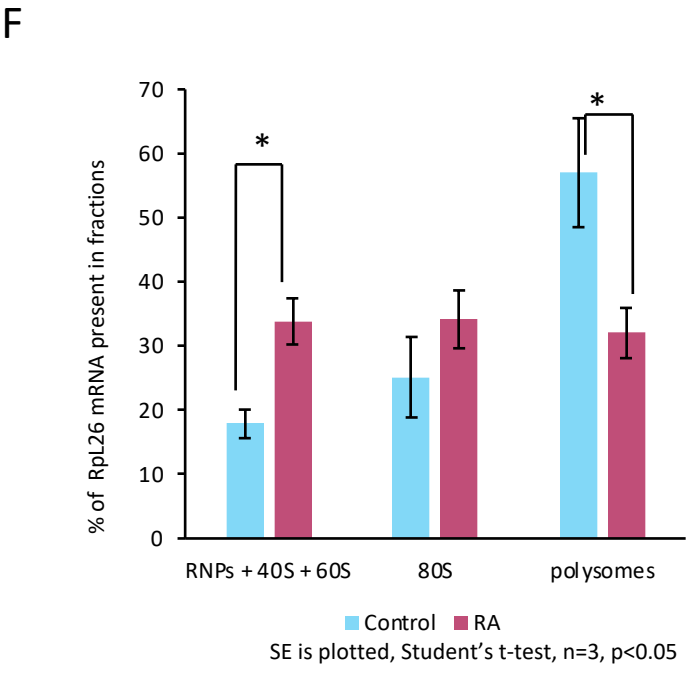
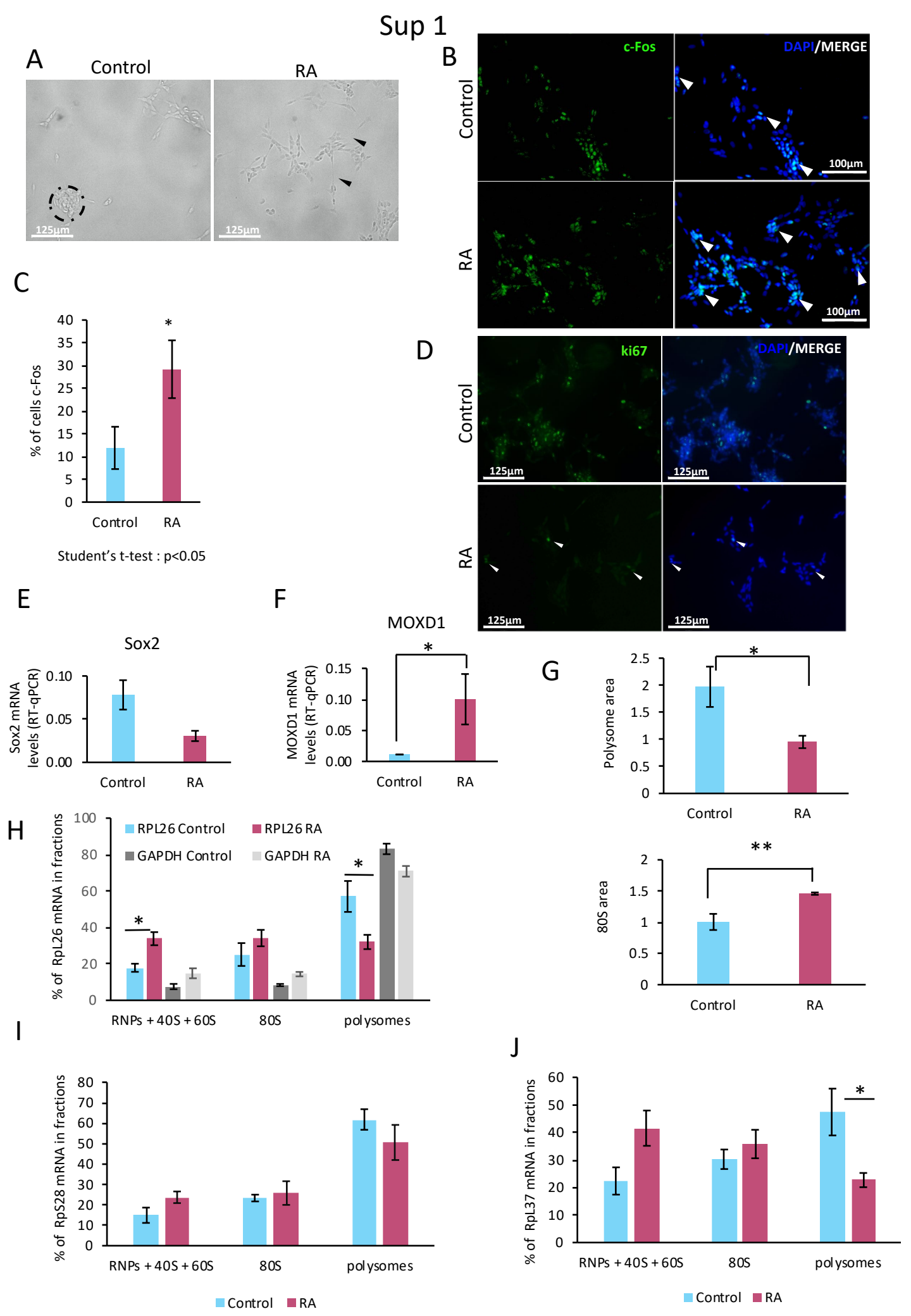

F
% of RpL26 mRNA present in fractions
70
60
50
40
30
20
10
0
*
*
RNPs + 40S + 60S
80S
polysomes
Control
RA
SE is plotted, Student's t-test, n=3, p<0.05

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 (βIII-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, $p<0.05$).

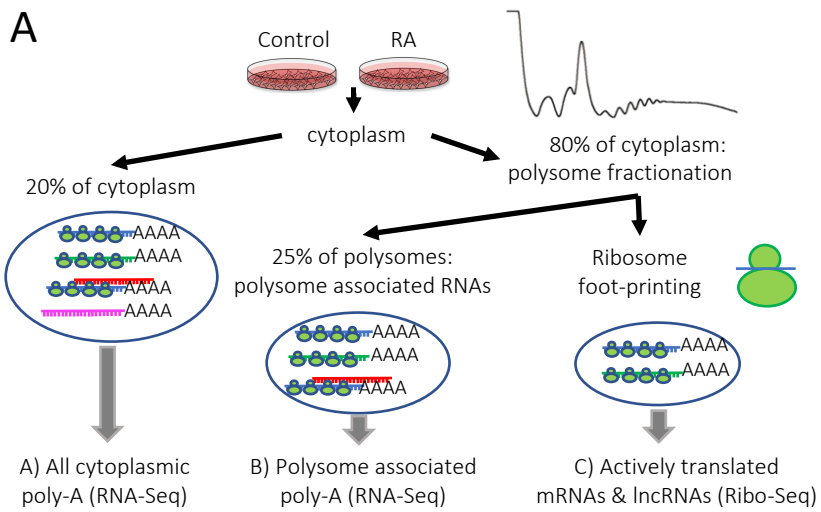
Sup 1

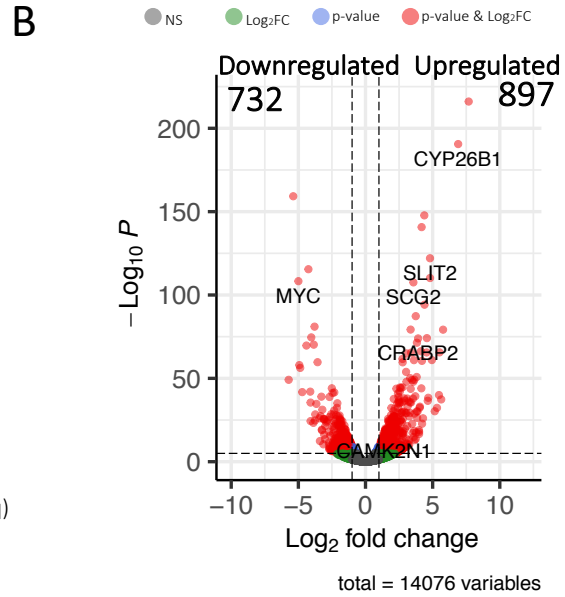

A
Control
RA
125µm
125µm
B
c-Fos
DAPI/MERGE
Control
RA
100µm
100µm
C
% of cells c-Fos
40
35
30
25
20
15
10
5
0
*
Control
RA
Student's t-test : p<0.05
D
ki67
DAPI/MERGE
Control
RA
125µm
125µm
125µm
125µm
E
Sox2
Sox2 mRNA levels (RT-qPCR)
0.10
0.05
0.00
Control
RA
F
MOXD1
MOXD1 mRNA levels (RT-qPCR)
0.15
0.10
0.05
0.00
*
Control
RA
G
Polysome area
2.5
2
1.5
1
0.5
0
*
Control
RA
80S area
2
1.5
1
0.5
0
**
Control
RA
H
% of RpL26 mRNA in fractions
100
80
60
40
20
0
RPL26 Control
RPL26 RA
GAPDH Control
GAPDH RA
*
*
RNPs + 40S + 60S
80S
polysomes
I
% of RpS28 mRNA in fractions
80
70
60
50
40
30
20
10
0
RNPs + 40S + 60S
80S
polysomes
Control
RA
J
% of RpL37 mRNA in fractions
60
50
40
30
20
10
0
*
RNPs + 40S + 60S
80S
polysomes
Control
RA

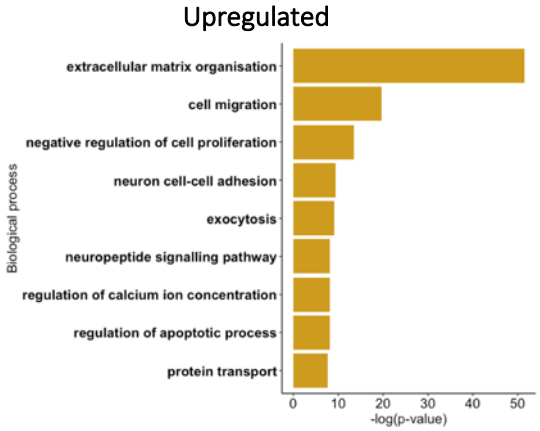
Sup Figure 1: Differentiation of SH-SY5Y results in a reduction in the level of active translation.

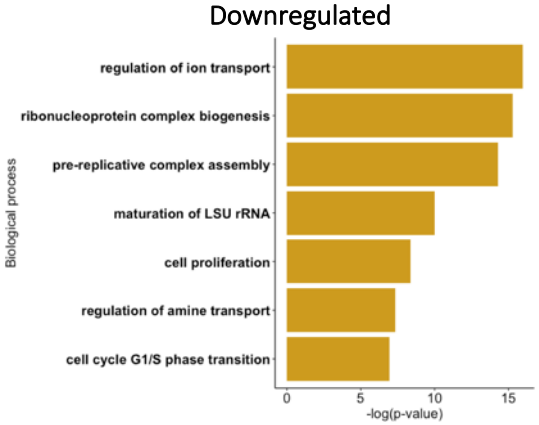
(A) Light microscopy reveals that undifferentiated SH-SY5Y cells grow in clumps and upon neural induction with RA they extend their neurites. (B) Immunofluorescence indicates c-Fos differentiation marker is expressed at higher levels in differentiated cells. (C) Quantification of c-Fos+ cells in Control 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, $p<0.05$).

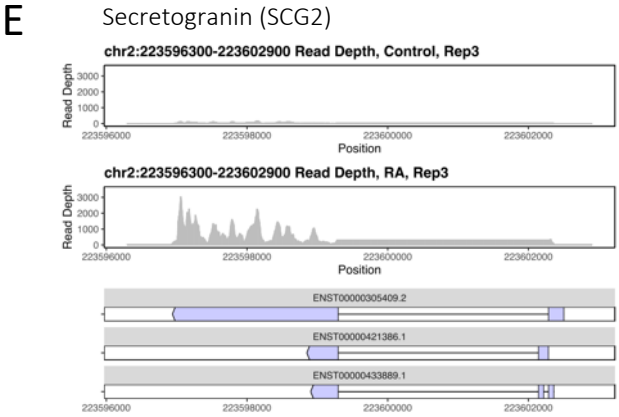
Figure 2

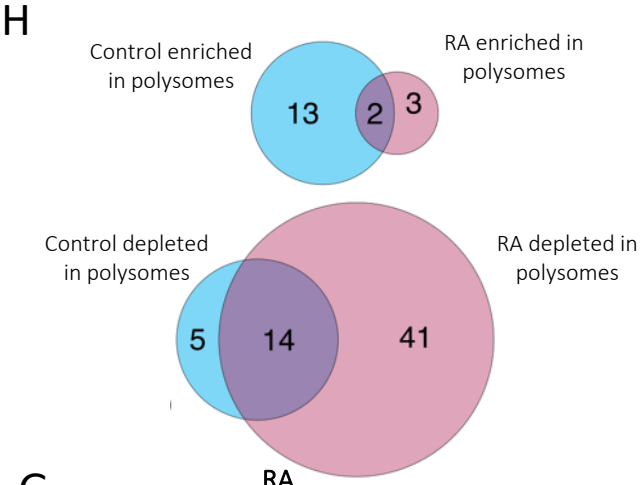

A
Control
RA
cytoplasm
20% of cytoplasm
80% of cytoplasm:
polysome fractionation
25% of polysomes:
polysome associated RNAs
Ribosome
foot-printing
AAAA
A) All cytoplasmic
poly-A (RNA-Seq)
B) Polysome associated
poly-A (RNA-Seq)
C) Actively translated
mRNAs & lncRNAs (Ribo-Seq)

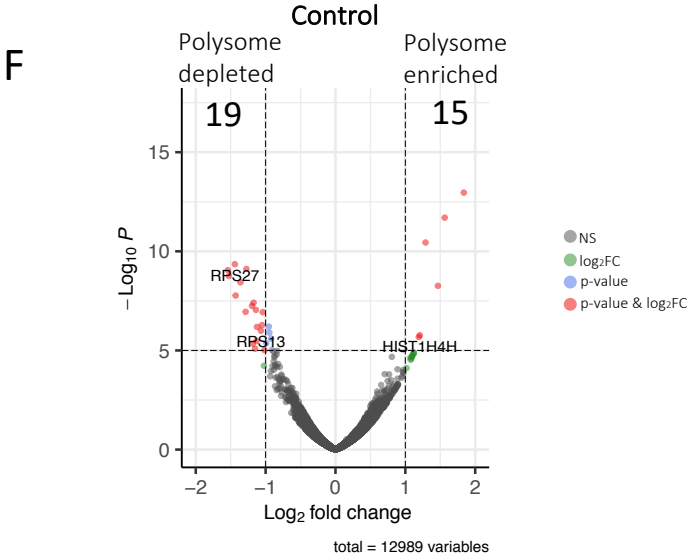

B
NS
Log2FC
p-value
p-value & Log2FC
Downregulated
Upregulated
732
897
CYP26B1
SLIT2
SCG2
MYC
CRABP2
CAMK2N1
−Log10 P
Log2 fold change
total = 14076 variables


C
Upregulated
extracellular matrix organisation
cell migration
negative regulation of cell proliferation
neuron cell-cell adhesion
exocytosis
neuropeptide signalling pathway
regulation of calcium ion concentration
regulation of apoptotic process
protein transport
Biological process
-log(p-value)


D
Downregulated
regulation of ion transport
ribonucleoprotein complex biogenesis
pre-replicative complex assembly
maturation of LSU rRNA
cell proliferation
regulation of amine transport
cell cycle G1/S phase transition
Biological process
-log(p-value)


E
Secretogranin (SCG2)
chr2:223596300-223602900 Read Depth, Control, Rep3
chr2:223596300-223602900 Read Depth, RA, Rep3
Read Depth
Position
ENST00000305409.2
ENST00000421386.1
ENST00000433889.1


H
Control enriched
in polysomes
RA enriched in
polysomes
13
2
3
Control depleted
in polysomes
RA depleted in
polysomes
5
14
41


F
Control
Polysome
depleted
Polysome
enriched
19
15
RPS27
RPS13
HIST1H4H
NS
log2FC
p-value
p-value & log2FC
−Log10 P
Log2 fold change
total = 12989 variables

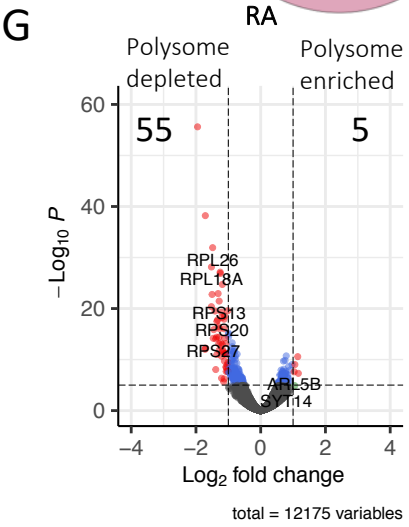
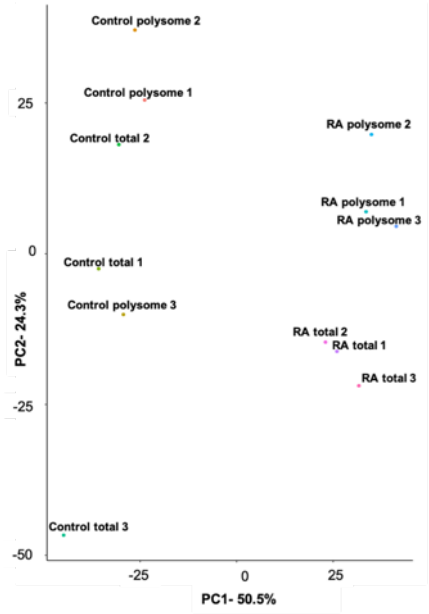

G
RA
Polysome
depleted
Polysome
enriched
55
5
RPL26
RPL18A
RPS13
RPS20
RPS27
ARL5B
SYT14
−Log10 P
Log2 fold change
total = 12175 variables

Figure 2: Differentiation results in global RNA changes

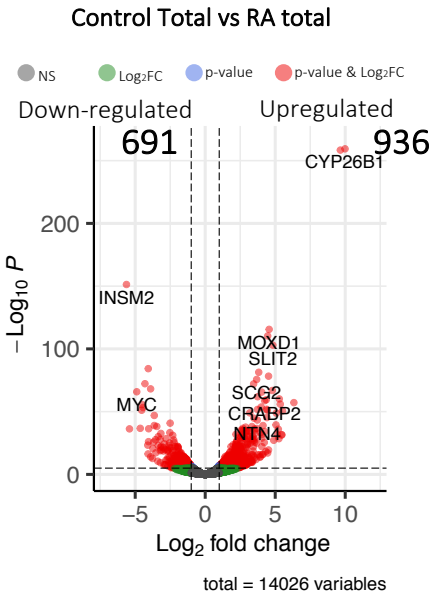
(A) Schematic of Poly-Ribo-Seq. (B) Volcano plot of protein-coding gene level changes in response to differentiation (polysome-associated); with 897 protein-coding genes significantly upregulated and 732 significantly downregulated (log_2 fold-change cut-off=1, $p^{adj}<0.05$). GO terms enriched in polysome-associated 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

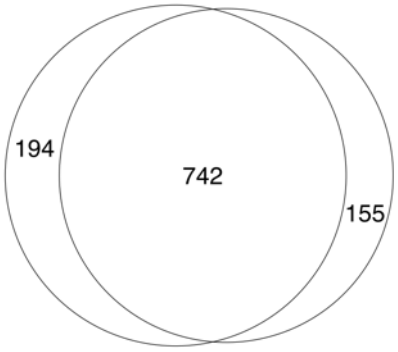


B



C

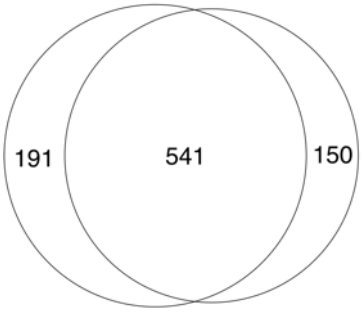
Upregulated upon differentiation in Total



Upregulated upon differentiation in Polysomes

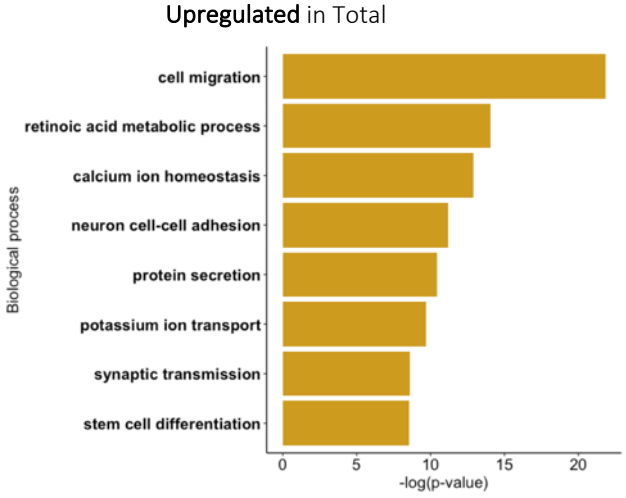
D

Downregulated upon differentiation in Total

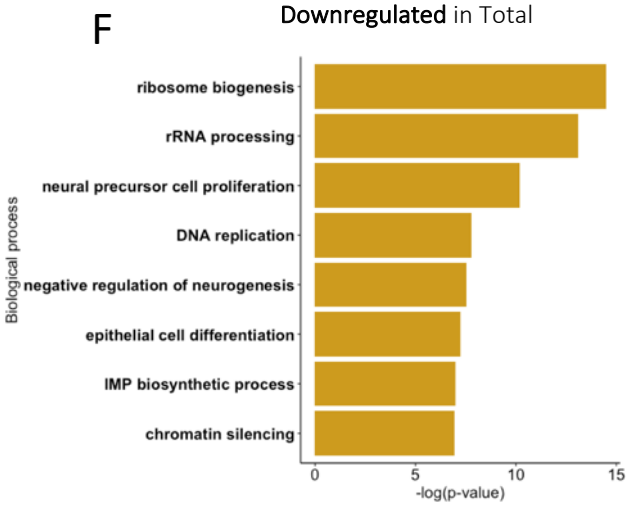


Downregulated upon differentiation in Polysomes

E



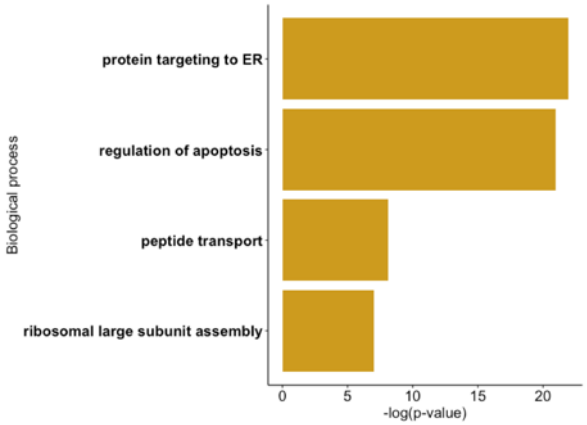
F



Sup 2 cont

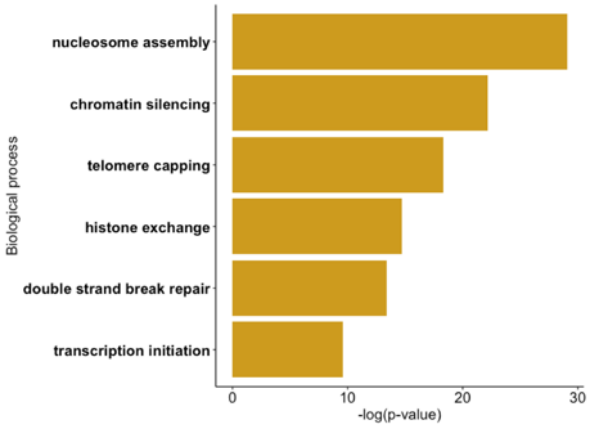
G

GO analysis of polysome depleted genes in Control



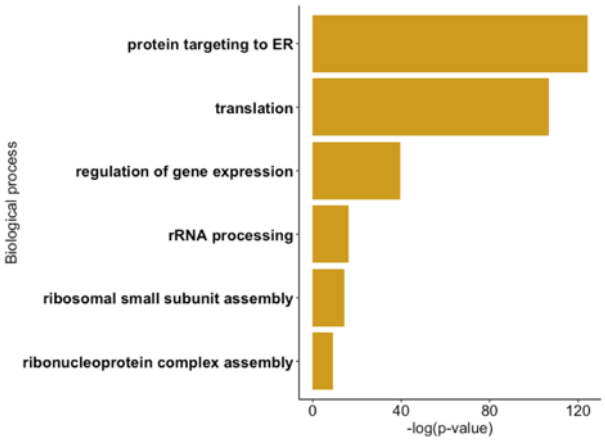
H

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, $p^{adj}<0.05$). 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

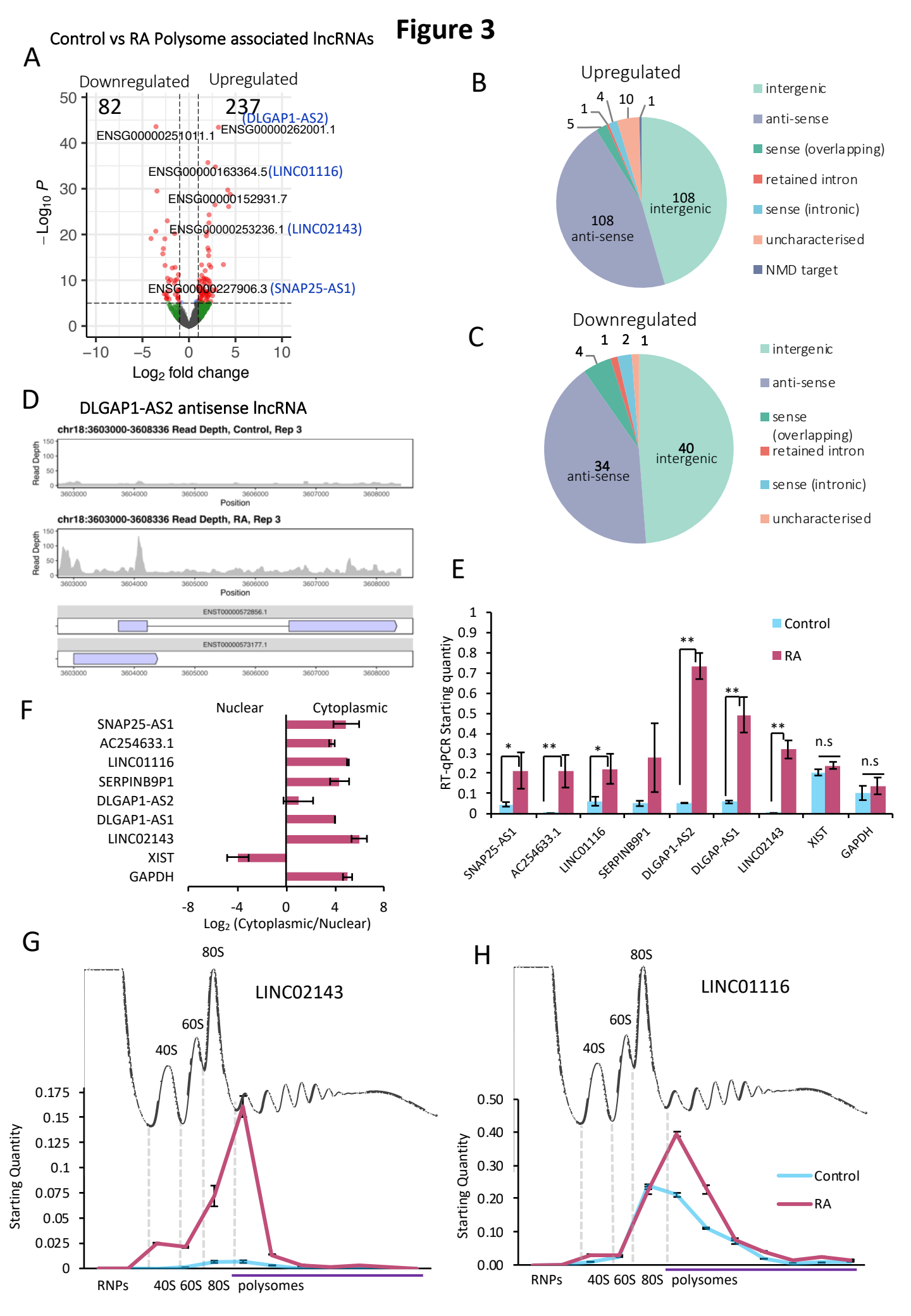

A
Control vs RA Polysome associated lncRNAs
Downregulated
Upregulated
82
237
ENSG00000251011.1
ENSG00000262001.1 (DLGAP1-AS2)
ENSG00000163364.5 (LINC01116)
ENSG00000152931.7
ENSG00000253236.1 (LINC02143)
ENSG00000227906.3 (SNAP25-AS1)
−Log10 P
Log2 fold change
B
Upregulated
108 intergenic
108 anti-sense
5
1
4
10
1
intergenic
anti-sense
sense (overlapping)
retained intron
sense (intronic)
uncharacterised
NMD target
C
Downregulated
4
1
2
1
40 intergenic
34 anti-sense
intergenic
anti-sense
sense (overlapping)
retained intron
sense (intronic)
uncharacterised
D
DLGAP1-AS2 antisense lncRNA
chr18:3603000-3608336 Read Depth, Control, Rep 3
chr18:3603000-3608336 Read Depth, RA, Rep 3
Read Depth
Position
ENST00000572856.1
ENST00000573177.1
E
Control
RA
RT-qPCR Starting quantiy
SNAP25-AS1
AC254633.1
LINC01116
SERPINB9P1
DLGAP1-AS2
DLGAP-AS1
LINC02143
XIST
GAPDH
n.s
F
Nuclear
Cytoplasmic
SNAP25-AS1
AC254633.1
LINC01116
SERPINB9P1
DLGAP1-AS2
DLGAP1-AS1
LINC02143
XIST
GAPDH
Log2 (Cytoplasmic/Nuclear)
G
LINC02143
40S
60S
80S
Starting Quantity
RNPs
40S 60S 80S
polysomes
H
LINC01116
40S
60S
80S
Starting Quantity
Control
RA
RNPs
40S 60S 80S
polysomes

Figure 3: LncRNAs regulated and associated with polysomes

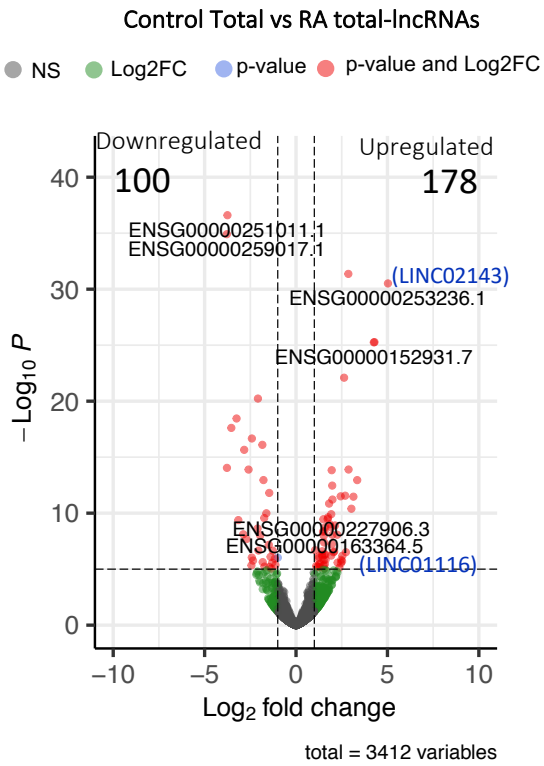
(A) Volcano plot displaying the differentially expressed polysome-associated lncRNAs (labelled by their geneIDs) between Control and RA populations. 237 lncRNAs are upregulated during differentiation and 82 downregulated (log_2 fold-change cutoff=1, $p^{adj}<0.05$). Pie charts showing types of lncRNAs (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).

Sup 3

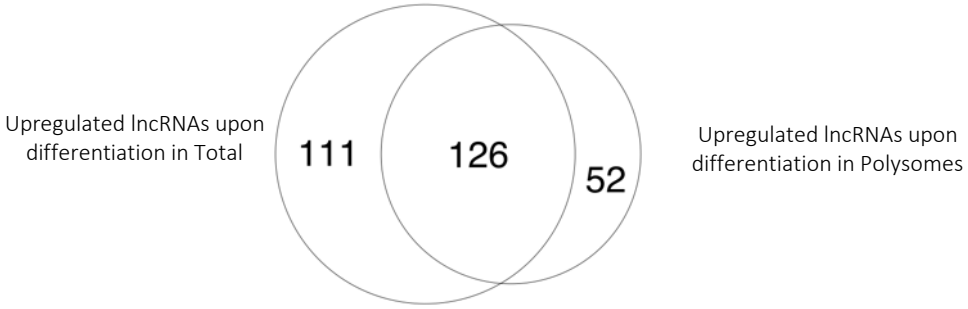
A

Raw data
Reads quality check
FastQC
Adapter removal
CutAdapt
Quality filter
Fastq (FASTX-Toolkit)
Reads quality check
FastQC
Reference genome (hg19)
Genome alignment
Rsubread
Obtaining read counts
featureCounts
Differential expression analysis (DESeq2)
-Differential expression analysis between Control and RA total and polysome populations
-Identification of differentially expressed genes

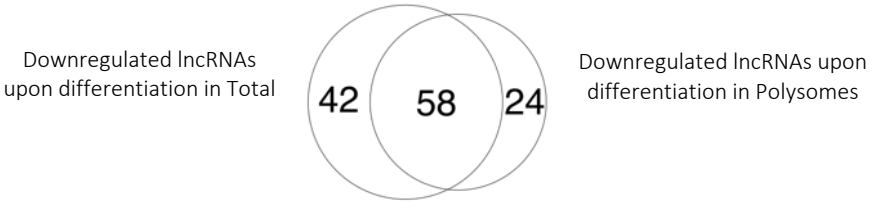
B


Control Total vs RA total-lncRNAs
NS
Log2FC
p-value
p-value and Log2FC
Downregulated
100
Upregulated
178
ENSG00000251011.1
ENSG00000259017.1
(LINC02143)
ENSG00000253236.1
ENSG00000152931.7
ENSG00000227906.3
ENSG00000163364.5
(LINC01116)
−Log10 P
Log2 fold change
total = 3412 variables

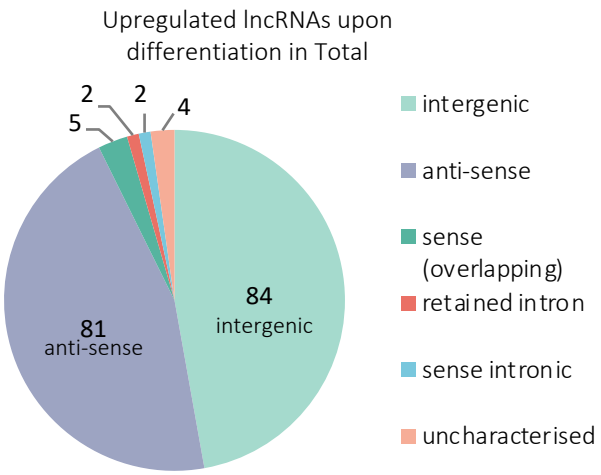
C


Upregulated lncRNAs upon differentiation in Total
111
126
52
Upregulated lncRNAs upon differentiation in Polysomes

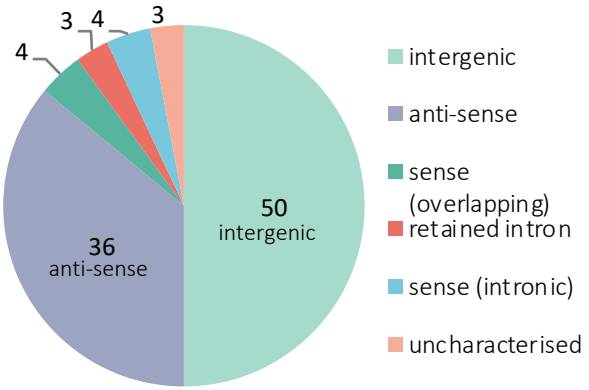
D


Downregulated lncRNAs upon differentiation in Total
42
58
24
Downregulated lncRNAs upon differentiation in Polysomes

E


Upregulated lncRNAs upon differentiation in Total
2
2
5
4
84
intergenic
81
anti-sense
intergenic
anti-sense
sense (overlapping)
retained intron
sense intronic
uncharacterised

F

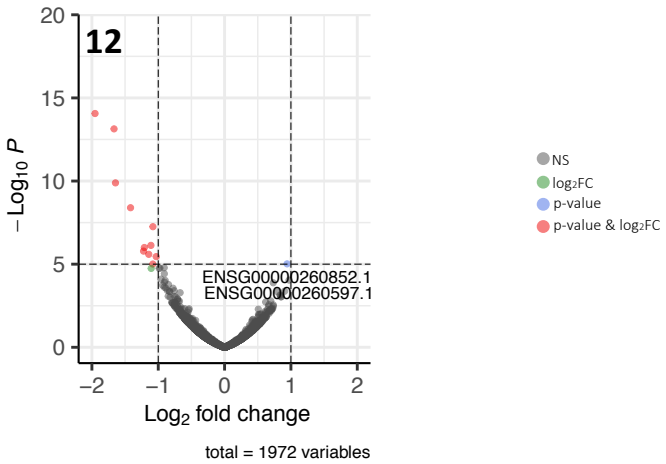

Downregulated lncRNAs upon differentiation in Total
3
4
3
4
50
intergenic
36
anti-sense
intergenic
anti-sense
sense (overlapping)
retained intron
sense (intronic)
uncharacterised

Sup Figure 3: Regulation of lncRNA expression and polysome association

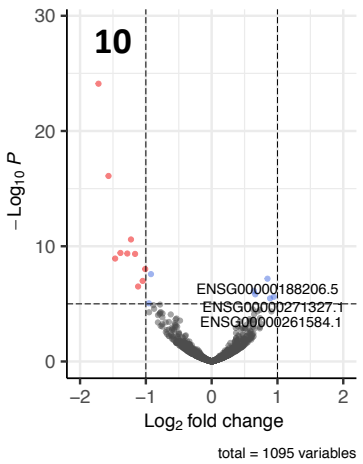
(A) Pipeline of lncRNA differential expression analysis. (B) Volcano plot displaying the significantly differentially expressed lncRNAs (labelled by their geneIDs) between Control and RA populations (Total) with $\log_2$ fold-change cutoff=1, $p^{adj}<0.05$. Venn diagrams of overlap in differentially 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).

Sup 4

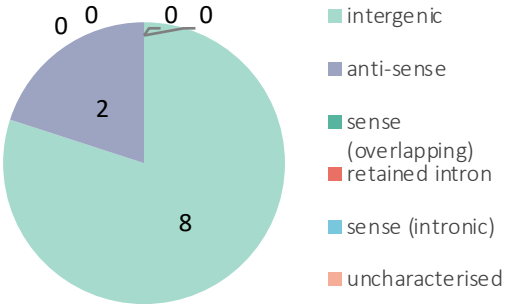
A



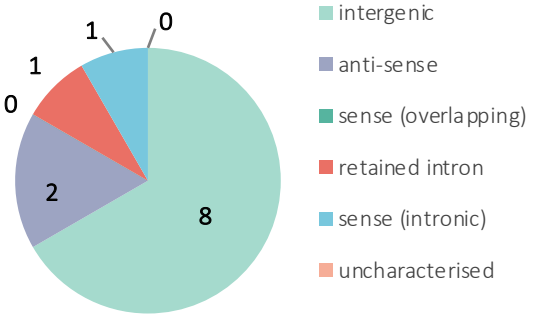
B



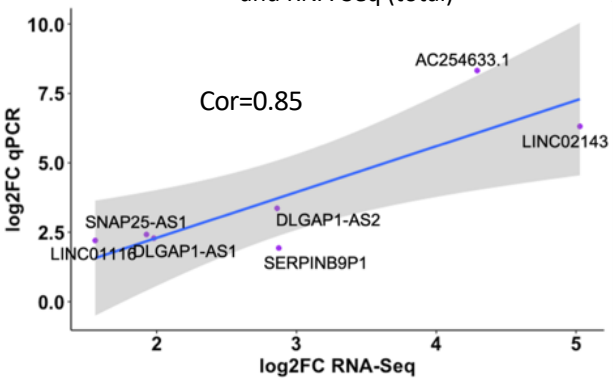
C



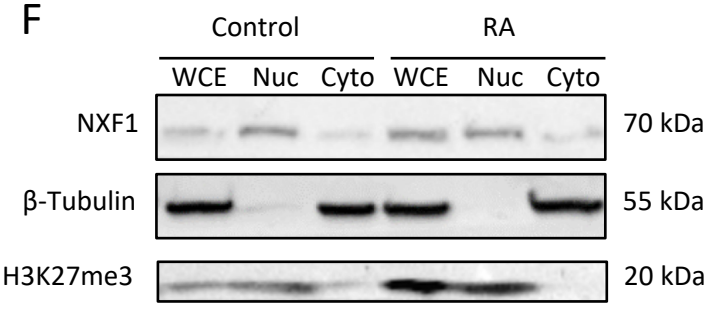
D



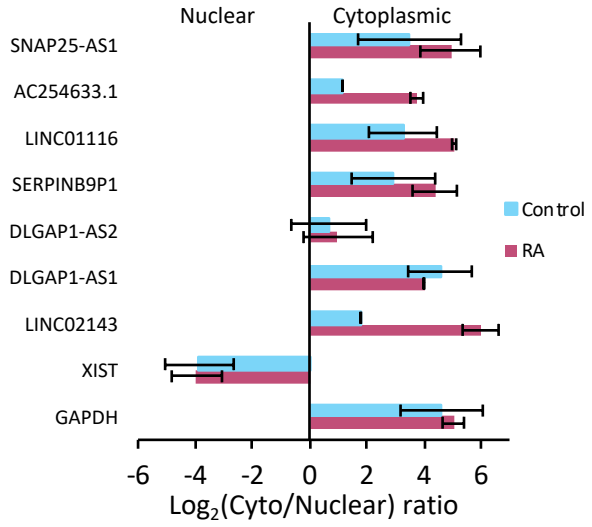
E



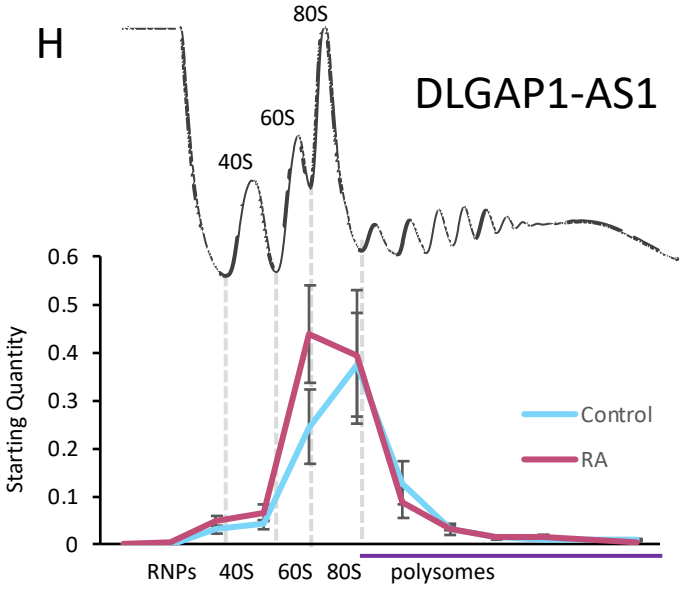
F



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, $p^{adj}<0.05$. 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

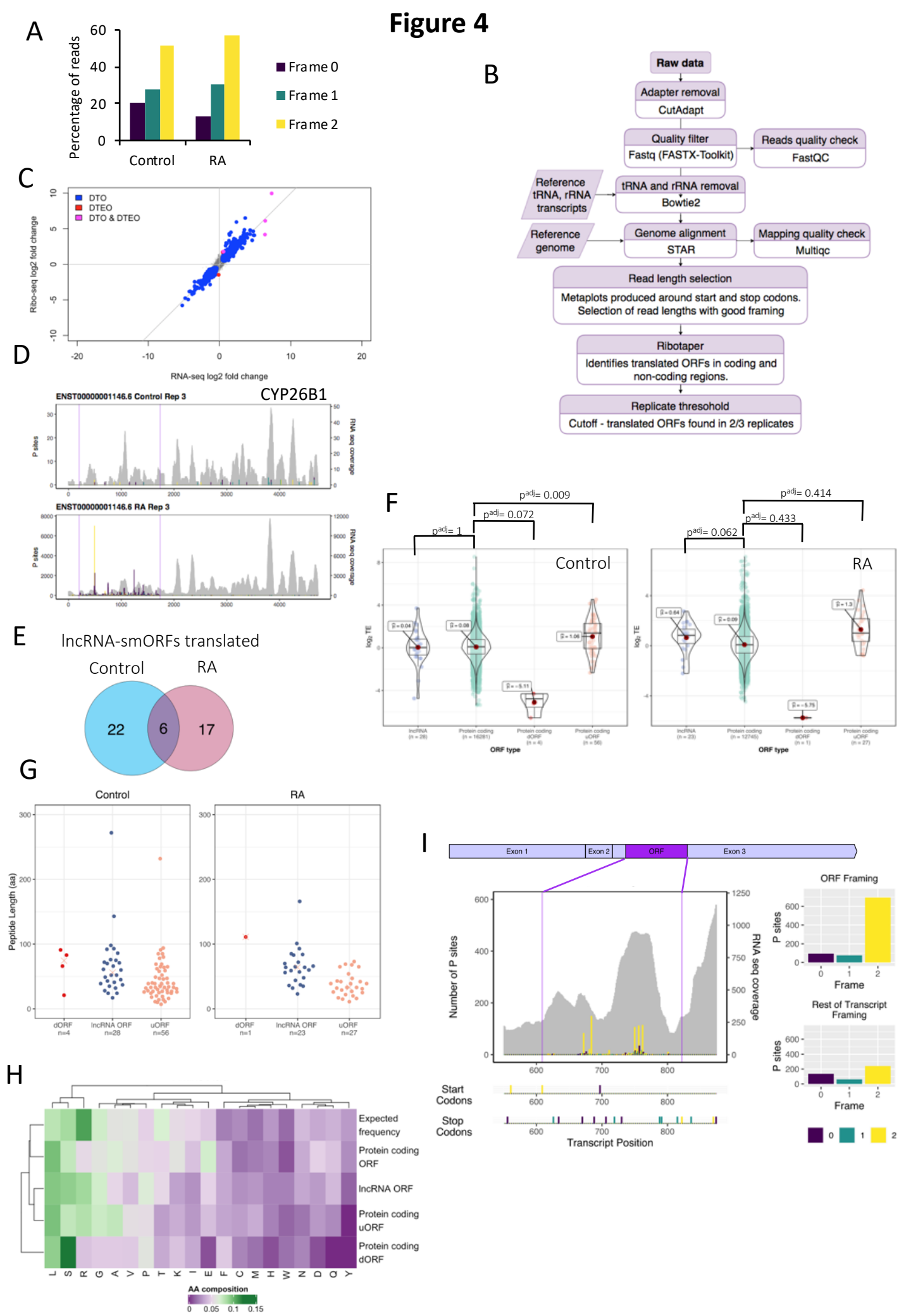

A
Percentage of reads
60
40
20
0
Control
RA
Frame 0
Frame 1
Frame 2
B
Raw data
Adapter removal
CutAdapt
Quality filter
Fastq (FASTX-Toolkit)
Reads quality check
FastQC
Reference tRNA, rRNA transcripts
tRNA and rRNA removal
Bowtie2
Reference genome
Genome alignment
STAR
Mapping quality check
Multiqc
Read length selection
Metaplots produced around start and stop codons. Selection of read lengths with good framing
Ribotaper
Identifies translated ORFs in coding and non-coding regions.
Replicate thresohold
Cutoff - translated ORFs found in 2/3 replicates
C
DTO
DTEO
DTO & DTEO
Ribo-seq log2 fold change
RNA-seq log2 fold change
D
CYP26B1
ENST00000001146.6 Control Rep 3
ENST00000001146.6 RA Rep 3
P sites
RNA seq coverage
E
lncRNA-smORFs translated
Control
RA
22
6
17
F
Control
RA
p^{adj}= 1
p^{adj}= 0.072
p^{adj}= 0.009
p^{adj}= 0.062
p^{adj}= 0.433
p^{adj}= 0.414
log2 TE
ORF type
G
Control
RA
Peptide Length (aa)
dORF n=4
lncRNA ORF n=28
uORF n=56
dORF n=1
lncRNA ORF n=23
uORF n=27
H
Expected frequency
Protein coding ORF
lncRNA ORF
Protein coding uORF
Protein coding dORF
L S R G A V P T K I E F C M H W N D Q Y
AA composition
0 0.05 0.1 0.15
I
Exon 1
Exon 2
ORF
Exon 3
Number of P sites
RNA seq coverage
Start Codons
Stop Codons
Transcript Position
ORF Framing
Rest of Transcript Framing
P sites
Frame

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 forwarded ORFs are driven by transcriptional regulation, with significant ΔRPF and ΔRNA but no significant ΔTE (blue). DTEOs (differential translation efficiency ORFs) or exclusive ORFs are driven by translational regulation, with significant ΔRPF and ΔTE without any significant ΔRNA (red). DTO & DTEOs (ORFs undergoing differential transciption and translation efficiency) are either intensified; a significant ΔTE acts with a significant ΔRNA, or buffered; a significant ΔTE is completely counteracting a ΔRNA, leading to no significant ΔRPF (pink). (D) View of Poly-Ribo-Seq data for CYP26B1. This is DTG ORF and exhibits significant ΔRNA and ΔRPF. (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.

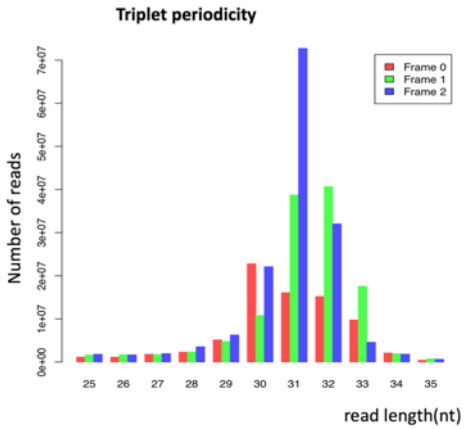
Translated ORFs	Control	RA	Overlap	Total
Protein-Coding ORFs	16,282	12,745	10,014	19,013
uORFs	56	27	12	71
dORFs	4	1	0	5
lncRNA-smORFs	28	23	6	45

Table 1: Translation of small ORFs

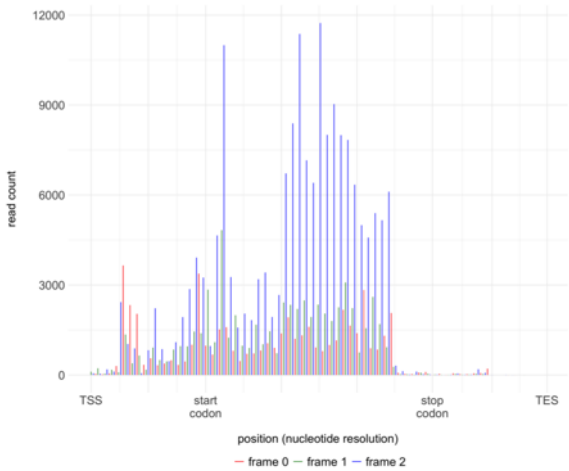
Number of ORFs detected as translated in Poly-Ribo-Seq

Sup Fig 5

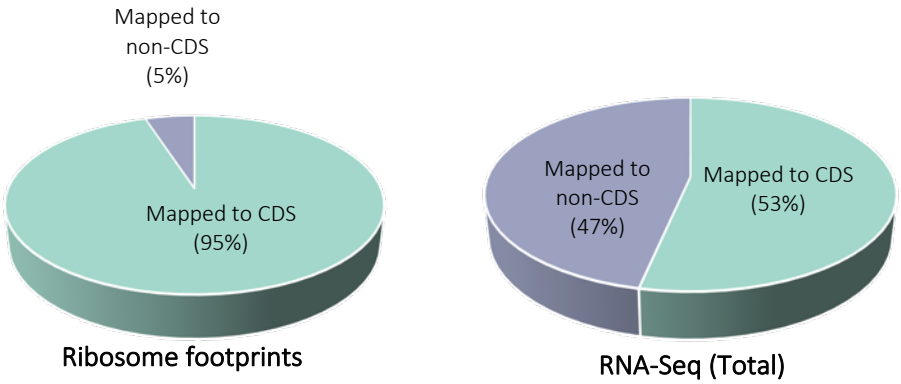
A



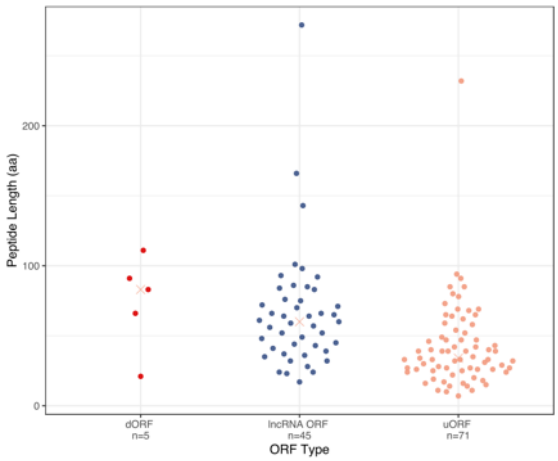
C



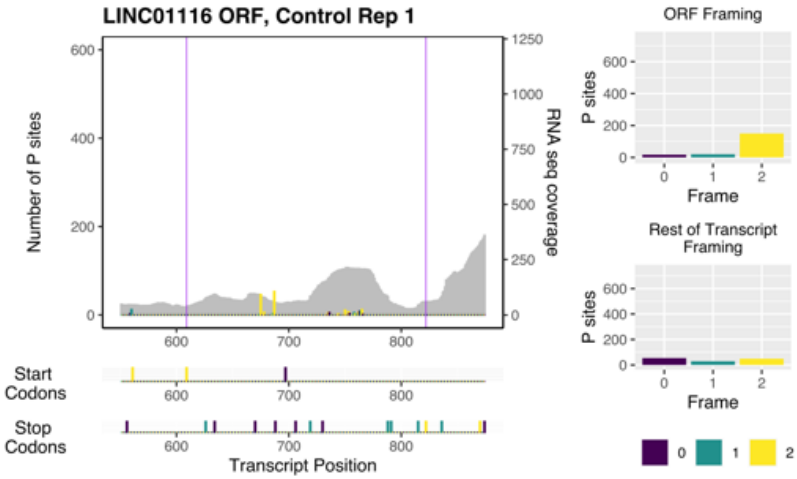
B



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

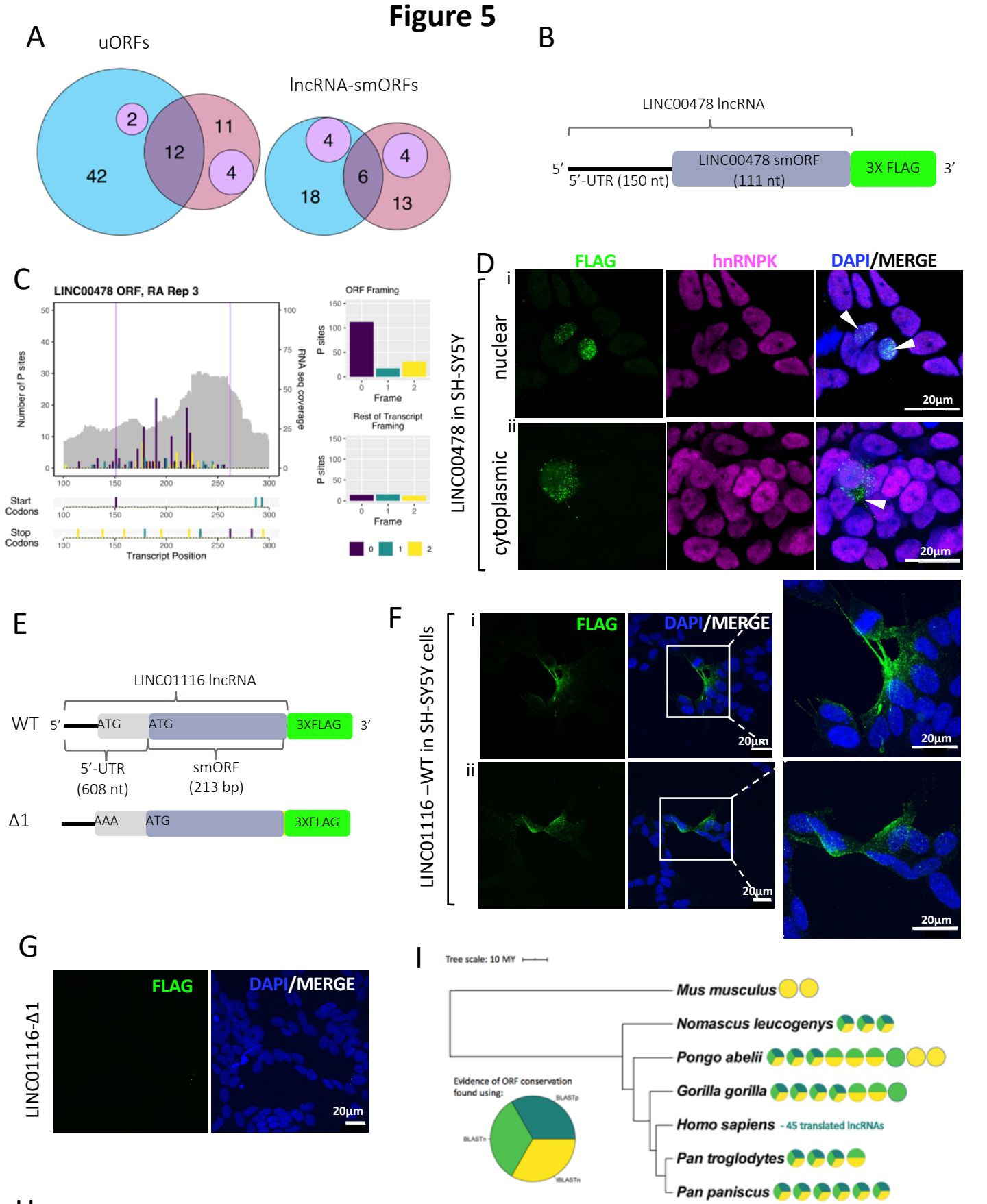

A
uORFs
42
2
12
11
4
lncRNA-smORFs
18
4
6
4
13
B
LINC00478 lncRNA
5'
5'-UTR (150 nt)
LINC00478 smORF
(111 nt)
3X FLAG
3'
C
LINC00478 ORF, RA Rep 3
Number of P sites
RNA seq coverage
Start Codons
Stop Codons
Transcript Position
ORF Framing
P sites
Frame
Rest of Transcript Framing
D
FLAG
hnRNPK
DAPI/MERGE
LINC00478 in SH-SY5Y
nuclear
cytoplasmic
20µm
E
LINC01116 lncRNA
WT
5'
ATG
ATG
3XFLAG
3'
5'-UTR
(608 nt)
smORF
(213 bp)
Δ1
AAA
ATG
3XFLAG
F
LINC01116 –WT in SH-SY5Y cells
FLAG
DAPI/MERGE
20µm
G
LINC01116-Δ1
FLAG
DAPI/MERGE
20µm
I
Tree scale: 10 MY
Mus musculus
Nomascus leucogenys
Pongo abelii
Gorilla gorilla
Homo sapiens - 45 translated lncRNAs
Pan troglodytes
Pan paniscus
Evidence of ORF conservation found using:
BLASTp
BLASTn
tBLASTn
H
Start codon
Stop codon
Consensus

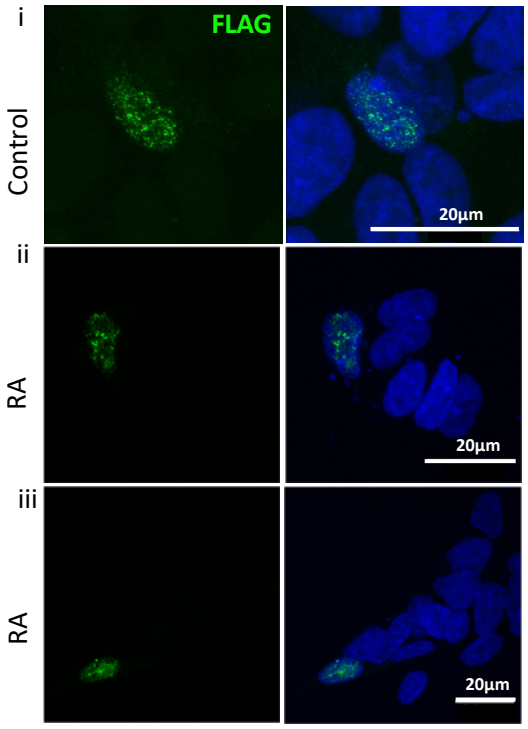
Figure 5: Peptide production from smORFs in lncRNAs

(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.

Sup 6

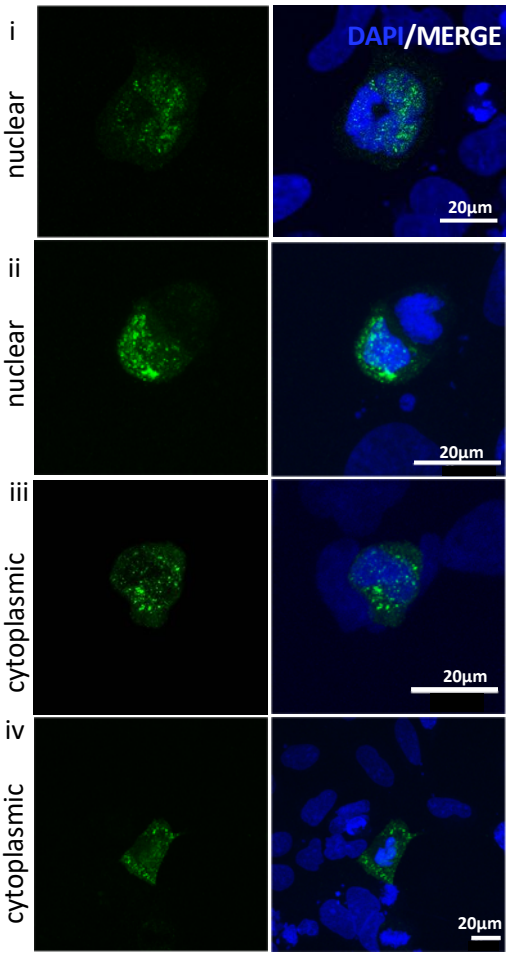
A

LINC00478 in SH-SY5Y



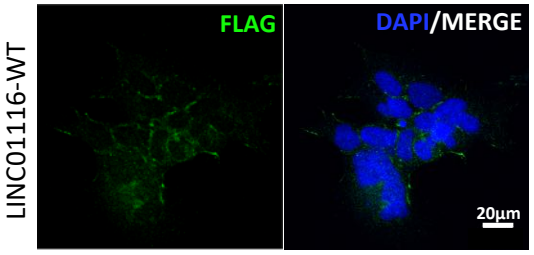
B

LINC00478 in HEK293

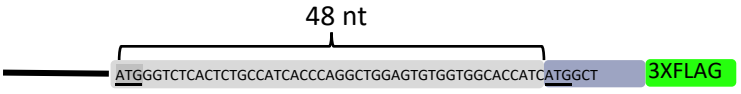


C

LINC01116 in HEK293

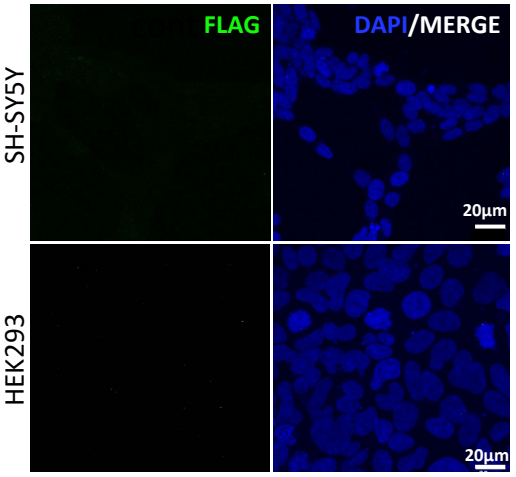


D

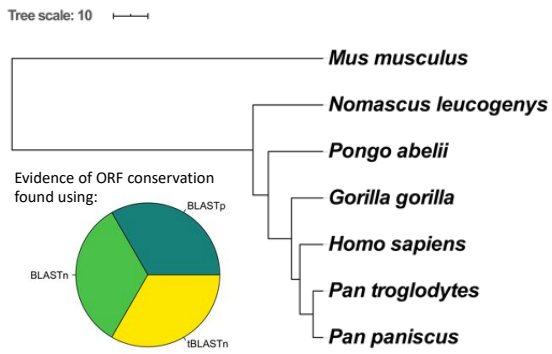


E

No transfection



F



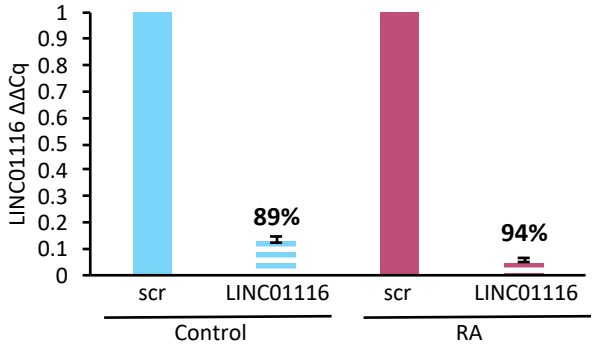
ENST00000526036.1_1226_2042	ENST00000501177.7_136_388	ENST00000590750.1_25_283	ENST00000453915.5_151_262	ENST00000504792.6_74_230	ENST00000504869.1_132_360	ENST00000449419.5_342_636	ENST00000454935.1_477_633	ENST00000557660.5_42_186	ENST00000566220.2_141_570	ENST00000502941.5_6_222	ENST00000510221.5_6_285	ENST00000645869.1_2_212	ENST00000651163.1_67_343	ENST00000441257.1_75_243	ENST00000429946.6_97_595	ENST00000442428.1_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

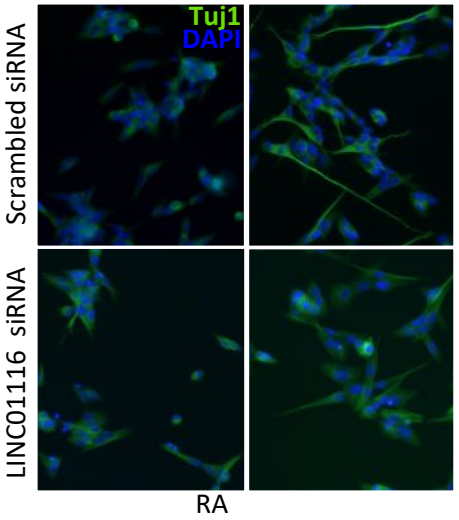
(A) Representative confocal images of FLAG-tagged LINC00478 peptide in (i) Control and (ii-iii) RA treated SH-SY5Y cells. LINC00478 exhibits nuclear distribution. (B) Representative confocal images of transfections of FLAG-tagged LINC00478 smORF in HEK293 cells; showing (i-ii) nuclear and (iii) cytoplasmic 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

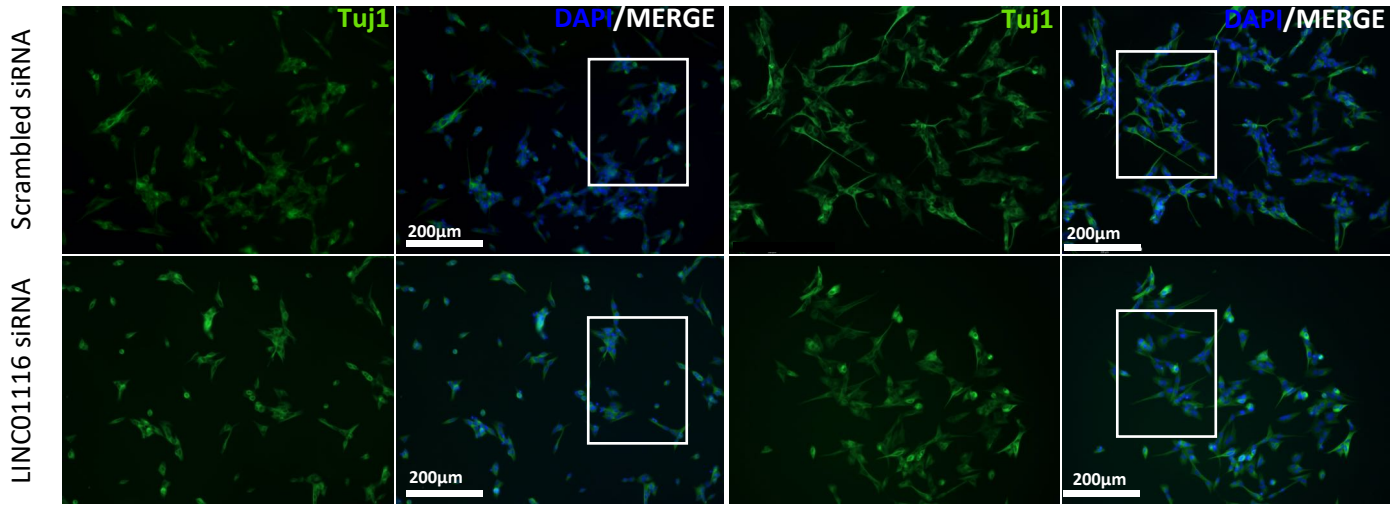
A



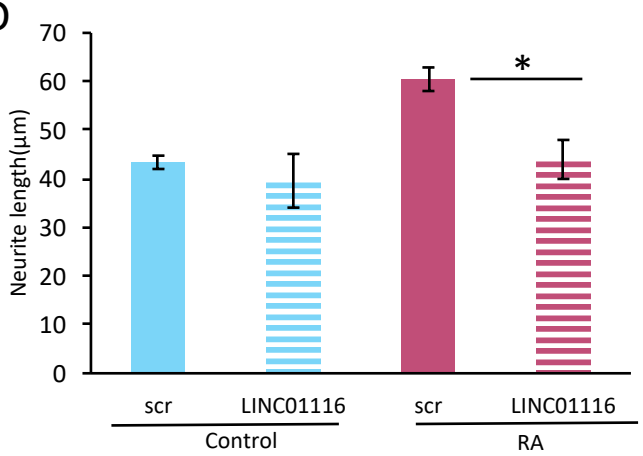
C



B

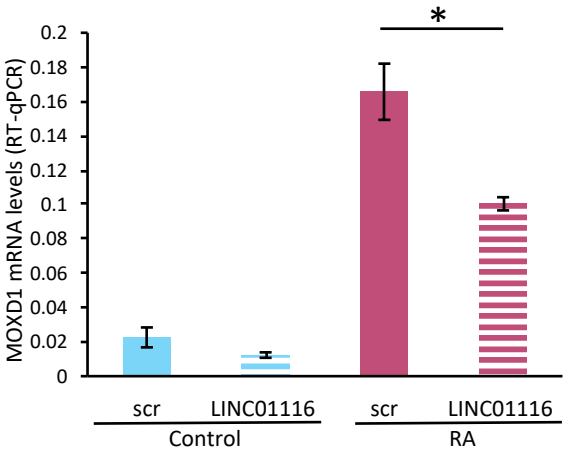


D



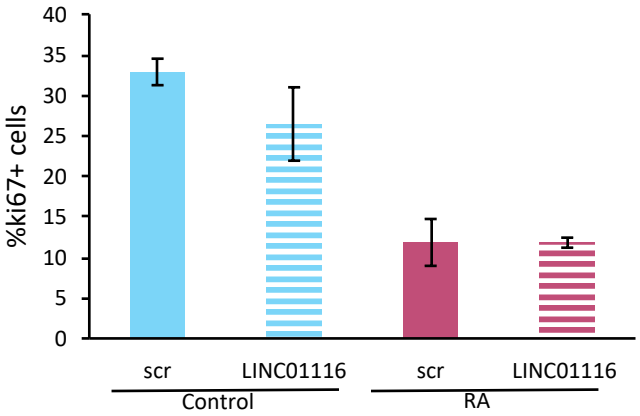
SE is plotted, Student's t-test, N=3, n>100, p<0.05

E



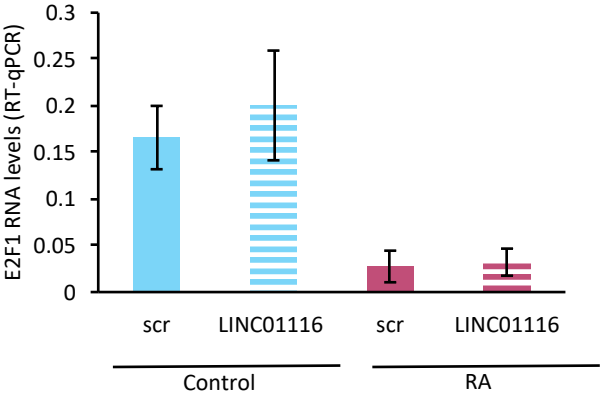
SE is plotted, Student's t-test, N=3, p<0.05

F



SE is plotted, Student's t-test, N=3, n>100, p>0.05

G



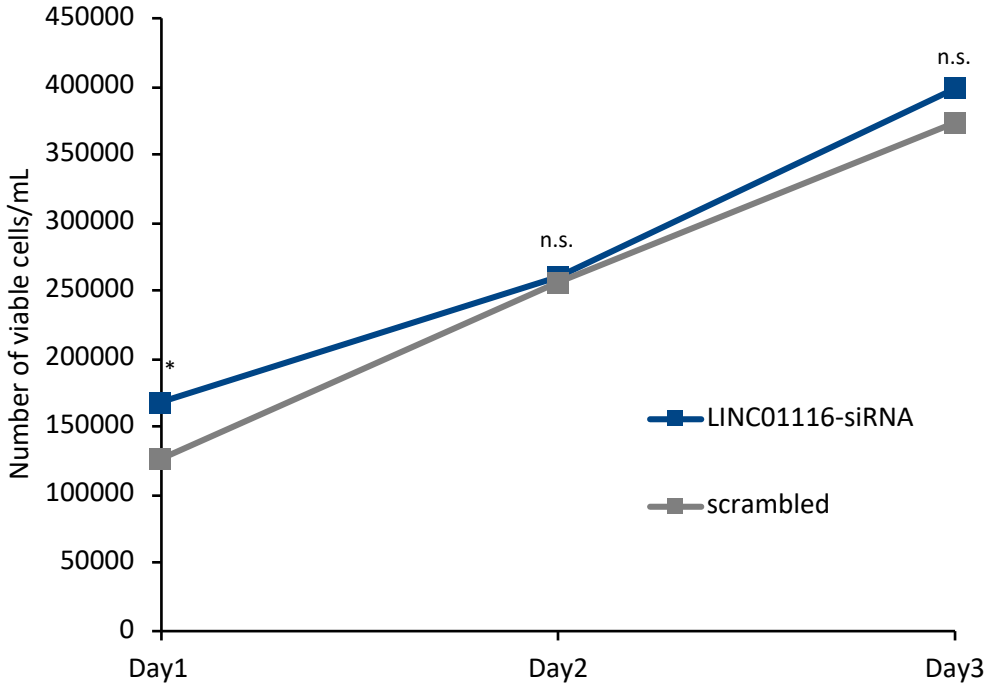
SE is plotted, Student's t-test, N=3, p>0.05

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

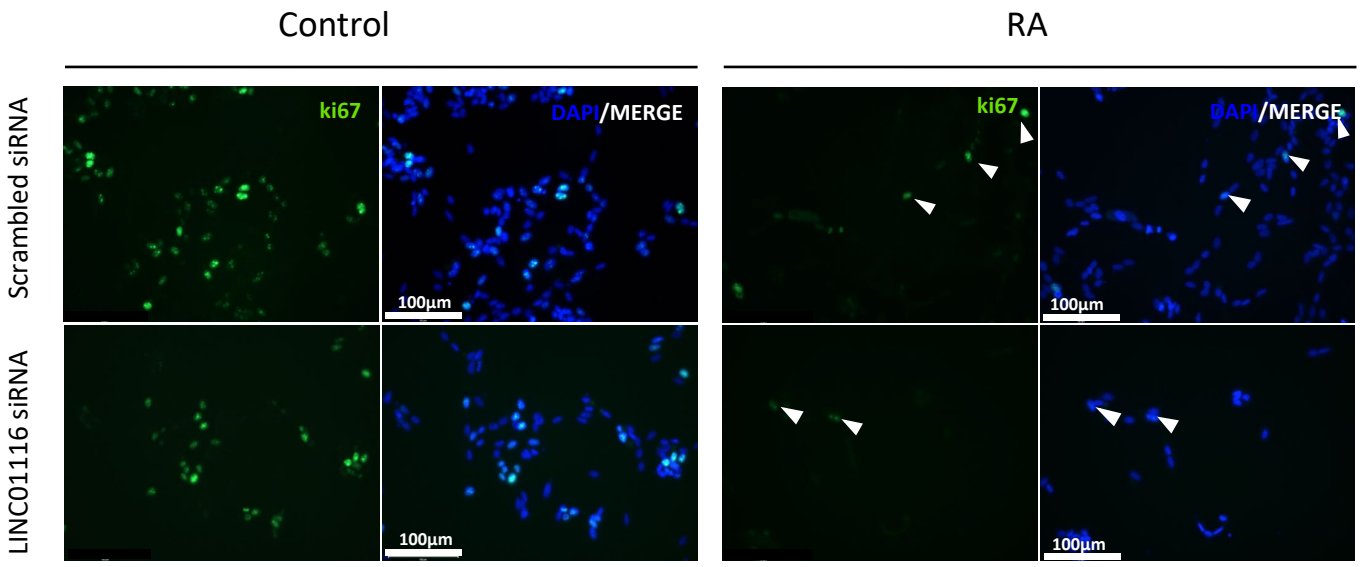
(A) LINC01116 was efficiently knocked down prior to differentiation and the knockdown is persistent throughout differentiation. (B) Representative immunofluorescence image of Control and RA SH-SY5Y cells, transfected with siRNA targeting LINC01116 and scrambled control, after staining for Tuj (βIII-tubulin) at day 3 post-differentiation (scale bar=200µm). 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

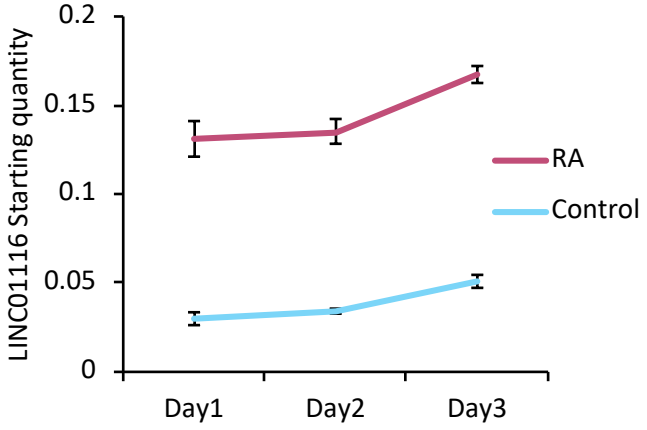
A



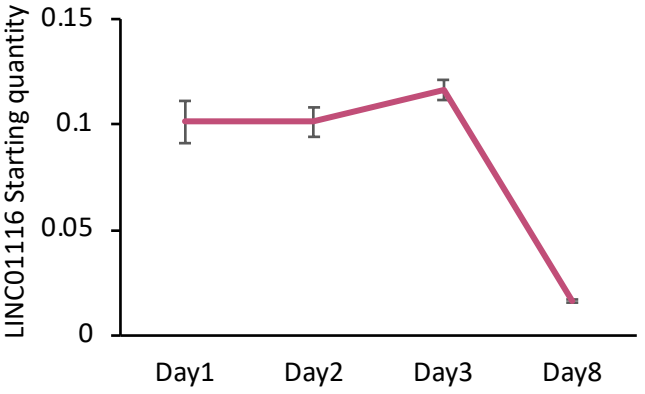
B



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=100µm (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).