Figures and figure supplements

Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq

Julie L Aspden, et al.
Figure 1. Poly-Ribo-Seq of small and large polysomes. (A) Venn diagram categorising annotated Drosophila smORFs as corroborated or uncorroborated based on evidence (FlyBase) from two out of three of: GO molecular function term assignment (green), peptidomic evidence (blue), and conservation outside of insects (red). Based on this, out of the total of 829 annotated smORFs, 665 are uncorroborated, and 494 have no evidence of translation. (B) Schematic of Poly-Ribo-Seq with representative UV absorbance profile for sucrose density gradient. Small (purple) and large (blue) polysomes are separated and subject to ribosome footprinting. (C) Composite plot from all FlyBase protein-coding genes of Poly-Ribo-Seq read counts across mRNAs in the vicinity of start (upper) and stop codons (lower) in small polysomes. (D) Median translational efficiencies of CDS, 5' and 3'-UTR regions for all protein-coding genes, error bars represent SE. DOI: 10.7554/eLife.03528.004
Figure 1—figure supplement 1. Poly-Ribo-Seq of small and large polysomes. (A) RT-PCR of RNA recovered from sucrose gradient fractions for one standard ORF mRNA (heph), three annotated smORF mRNAs (CG14818, CG9032, and CG43194) and one long non-coding RNA (roX1), with -RT control. Fractions corresponding to small (purple, 2-6 ribosomes) and large (blue, 7 or more ribosomes) polysomes are indicated. (B) Read densities (RPKM) from two biological replicates of the total cytoplasmic mRNA control exhibit very high correlation ($R^2 = 0.96$). (C and D) Read density plots showing phasing of ribosome footprinting reads in triplets corresponding to codons in CDS (C) and an absence of triplet phasing in 3′-UTRs (D) (small polysome data). DOI: 10.7554/eLife.03528.005
<table>
<thead>
<tr>
<th>Ribosome-mRNA complex</th>
<th>Polysomes</th>
<th>Translational Metrics</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Large polysomes (7+)</td>
<td>High RPKM (median 40.2)</td>
<td>-Canonical ORFs covered by many ribosomes -highly translated</td>
</tr>
<tr>
<td>B</td>
<td>Small polysomes (2-6)</td>
<td>Medium RPKM (median 27.2)</td>
<td>-Canonical ORF bound by a few ribosomes but distributed across ORF -lowly translated</td>
</tr>
<tr>
<td>C</td>
<td>Small polysomes (2-6)</td>
<td>High RPKM (median 82.7)</td>
<td>-Small ORF in own mRNA -highly translated</td>
</tr>
<tr>
<td>D</td>
<td>Small polysomes (2-6)</td>
<td>Low Coverage (&lt;0.40)</td>
<td>-Canonical ORF containing theoretical small ORF</td>
</tr>
</tbody>
</table>

**Figure 1—figure supplement 2.** Schematic interpretation of Poly-Ribo-Seq. Schematic summary of characterised (A–C) and theoretical (D) translation scenarios. Diagrams of ribosome–mRNA complexes are shown along with the polysome fraction in which it is detected, translational metrics and interpretation of this information, for (A–C) long canonical ORFs, (C) smORFs, and (D) canonical ORF containing a theoretical small ORF. DOI: 10.7554/eLife.03528.006
Figure 2. Poly-Ribo-Seq reveals translation of smORFs. (A) Ribosome footprinting densities (RPKM) from small polysomes correlate poorly with large polysomes (whereas two replicates of total cytoplasmic mRNA controls do, see Figure 1—figure supplement 1B). (B) Ribosome footprinting densities (RPKM) from small polysomes correlate highly between two biological replicates ($R^2 = 0.83$). (C) All 106 smORFs detected in large polysomes (blue) were...
also present in the 191 detected in small polysomes (purple). smORF footprints are much more abundant in small polysomes, as indicated by a higher TE value. (D) High coincidence of annotated smORFs detected as translated in three different Poly-Ribo-Seq experiments. Small polysome extensive experiment probes most deeply with 224 smORFs detected as translated (small polysomes: purple, small polysomes extensive: yellow, rRNA: turquoise). (E) Numbers and proportions of transcribed ORFs, which are translated, according to Poly-Ribo-Seq data (translated: green, untranslated: blue). The proportion of annotated smORFs translated is similar to that of standard CDSs. 121 annotated smORFs are newly detected as translated, plus 2708 uORFs and 313 smORFs from ncRNAs. (F) Venn diagram showing overlap between Poly-Ribo-Seq (dark green), our mass spectrometry experiments (purple) and Peptide Atlas proteomic data (red).

DOI: 10.7554/eLife.03528.007
Figure 3. Validation of smORF translation by tagging assay. (A–D) Ribosome footprints from small polysomes (pink) and mRNA reads (grey) mapped to smORFs, along with transcript and ORF models of (A) CG7630, (B) CG33774, (C) CR30055 (ncRNA), and (D) FBtr0072084_1 (uORF). Corresponding transfection assays in S2 cells are shown (FLAG antibody: green, F-actin stained with phalloidin: red, scale bars = 5 μm) together with Poly-Ribo-Seq metrics (RPKM, coverage and TE). Distribution of each peptide (reticular, other cytoplasmic or limited) is indicated.

DOI: 10.7554/eLife.03528.010
Figure 3—figure supplement 1. Validation of smORF translation by tagging assay. (A) Schematic of the transfection construct into which smORF 5′-UTRs and ORFs (no stop codon) were cloned under the Actin promoter, such as to be fused in frame to a C-terminal FLAG tag, with its own AUG start codon mutated to GCG. (B) Transfection negative controls, plasmid with no ORF (nor AUG), plasmid with the full-length tal transcript (minus 3′-UTR) with ORF-B tagged with FLAG, which has previously been shown not to be translated (Galindo et al., 2007), and a plasmid containing a putative smORF that is transcribed but not translated according to our Poly-Ribo-Seq (Uhg2-ORF1). (C) Immunoblot showing translation of FLAG-tagged smORFs. Figure 3—figure supplement 1. Continued on next page
(Table 3) corresponding to predicted sizes, along with β-tubulin loading control. (D) Different subcellular localisations of FLAG-tagged smORFs (green) corroborated by double staining with Mitotracker Red (red): “mitochondrial”, “other cytoplasmic” and “limited” (scale bar = 5 µm). (E) Correlation analysis of colocalisation between FLAG-tagged smORF peptides and Mitotracker Red, error bars represent SD from three experiments. (F) 50% of S2-cell translated smORFs show function in previous RNAi screens (Flymine). (G) Translation of FLAG-tagged pncr009:3L (ncRNA) ORFs 1, 2, and 3 in transfection assay with translational metric values shown (FLAG antibody: green, F-actin stained with phalloidin: red, scale bars = 5 µm). (H) Immunoblot showing detection of FLAG-tagged ORFs from pncr009:3L and CR30055 with predicted sizes (Table 4), along with β-tubulin loading control. (I) Translation of FLAG-tagged uORFs FBtr0072210_1 and FBtr0081720_1 in transfection assays with translational metric values shown (FLAG antibody: green, F-actin stained with phalloidin: red, scale bars = 5 µm).

DOI: 10.7554/eLife.03528.011

Figure 3—figure supplement 1. Continued

Figure 3—figure supplement 2. Poly-Ribo-Seq reveals translation of ORFs in ncRNAs. (A) Read density plot showing phasing of ribosome footprinting reads in the frame of smORFs within CR30055 and pncr009:3L detected as translated and confirmed by FLAG immunofluorescence translation assay. (B) Correlation of reads obtained by ORFs after Poly-Ribo-Seq (y axis) with reads obtained by sequencing of polysomal fractions before ribosome footprinting (x axis). The correlation is much stronger for canonical long ORFs and putative smORFs (grey) than for ncRNA ORFs (red). Many ncRNA ORFs below the 11.8 RPKM cut-off used to ascertain translation (green dotted line) can show association with polysomes (high Polysomal RNA RPKM), thus translation of ORFs in ncRNAs does not simply stem from non-coding association with polysomes.

DOI: 10.7554/eLife.03528.012
Figure 4. Bioinformatic indicators of smORFs. (A) Distribution of phastCons scores for intergenic regions, standard length protein-coding CDSs (longer than 100 aa), S2 cell-translated annotated smORFs, and all annotated smORFs, with fitted normal curves. Green dotted lines indicate the 90th percentile. Figure 4. Continued on next page.
of intergenic phastCons scores (0.55). (B) Relative abundance of particular amino acids in proteins (random expected: black, all CDSs: purple, all annotated smORFs: yellow, and translated smORFs: red). (C and D) Proportion of (C) S2-cell translated (32%) and (D) all smORFs (32%) predicted to contain transmembrane α helices (TMHMM). (E and F) Frequency distribution of smORF peptide lengths for (E) translated and (F) all annotated smORFs with medians shown by red dotted line.

DOI: 10.7554/eLife.03528.015
Figure 4—figure supplement 1. Bioinformatic indicators of smORFs. (A) Relative abundance of all amino acids in ORFs, (random: grey, all CDS: purple, all annotated smORFs: yellow, and translated annotated smORFs: red). (B) Enrichment of GO molecular function terms (GOrilla) within translated annotated smORFs in S2 cells when compared to translated standard protein-coding ORFs. Main overrepresented terms are structural constituents of ribosome (p = 3.28E-4), oxidoreductase activity and transmembrane transporter activity (p = 2.77E-5). (C–D) Frequency distribution of peptide lengths, phastCons, and relative abundance of particular amino acids of translated (C) uORFs and (D) ncRNA ORFs. Red dotted lines indicate the median amino acid lengths and green dotted lines indicate the 90th percentile cut-off from phastCons of intergenic regions, 0.55 (Figure 4A).

DOI: 10.7554/eLife.03528.016