



**Figure S7. Alternative methods to define coding regions or transcripts based on ribosome profiling.**

Alternative methods or features could have been used to define the translated ORF. For example, we observed a very high peak of RPFs one nucleotide off (out of frame) at some stop codons, likely corresponding to a structural change of the ribosome during translation termination. This strong peak was observed in all the time points and could potentially be used as a feature to define coding potential (Supplementary Fig. 1).

However, not all highly-translated annotated coding ORFs defined by ORFscore had this characteristic read, so it was not used to define translated ORFs. (Supplementary Fig. 7).

Translation Efficiency (TE) could also be used to define coding ORFs, however, selecting the ORF with highest TE per transcript did not capture the majority of annotated coding ORFs (green) (Supplementary Fig. 7).

Ribosome Release Score (RRS) is method that is designed to compare different classes of transcripts, and dependence on an intact, properly-annotated 3'UTR precludes the analysis of many genes, therefore a direct comparison to the ORFscore is not appropriate. We note that we have found several transcripts with annotated ORFs that could not be scored by the RRS but presented high ORFscore (Supplementary Fig. 7).

Others methods such as Translated ORF Classifier (TOC) (Chew et al., 2013) or Periodicity Transition Score (PTS) (Michel et al., 2012) were not evaluated because, similarly to RRS, TOC is not used to define the translated ORF, but rather classify whole transcripts as coding, leader- or trailer-like, and PTS was designed to detect transitions between reading frames that occur during programmed ribosomal frameshifting or in dual coding regions where the same nucleotide sequence codes for multiple proteins in different reading frames. Compared to the TOC coding predictions, we found that some transcripts previously classified as leader-like contain ORFs that are actively translated based on ORFscore. A set of transcripts were defined as coding by TOC but not by ORFscore. This might be due in part to the fact that TOC takes in account intrinsic ORF characteristics (such as relative size) while the ORFscore exclusively depends on experimental ribosome footprints.

ORFscore defines translated ORFs based on experimental data, however as we designate ORFs from each stop codon to the most distal Methionine, it is possible that the translation start site for a given coding region might not be accurate. Alternative Methionines or non-canonical start codons (e.g. CUG) could serve as translation initiation sites.

In this study the detection of smORFs depends on the quality of transcript annotation, and in some cases it is clear from the RNA input data that the transcript annotation is not fully supported by RNA-Seq. Therefore, additional translated smORFs could be identified by these methods once new transcript models are assembled. Regarding the dynamic of the annotation releases, we have found one gene that was not annotated as coding in the dataset used, but has since been annotated as protein-coding (NM\_001288641.1/ENSDART00000131802/linc-carm1).

- (A) Scatterplot of ORFscore compared to RNA expression for the best-scoring ORF in each transcript. Diagram of the RefSeq ORF color code is shown.
- (B) Scatterplot of translation efficiency (TE) versus RNA expression for the ORF with highest TE in each transcript.
- (C) Scatterplot of ORFscore versus RRS (Ribosome Release Score, Guttman et al. 2013) for all RefSeq transcripts. A notable portion of ORFs have a RRS of -99 (-6.62 adjusted, log scaled), indicating they were not able to be analyzed.
- (D) Scatterplot of ORFscore versus number of RPFs one nucleotide out of frame at the annotated stop. Note that many highly coding genes lack the stop peak (x-axis = 0).
- (E) Scatterplot of ORFscore versus coverage for all the genes analyzed by Chew et al. 2013. Each gene is colored based on the TOC coding prediction.
- (F) Venn diagram showing the number of genes per category identified by Chew et al. 2013 and by ORFscore.