nature portfolio

Peer Review File

Ribo-ITP enables identification of translons from limited input samples

Corresponding Author: Professor Can Cenik

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript by Ghatpande and Paul et al. presents a compelling proof-of-concept study applying their previously developed Ribo-ITP technique to microdissected hippocampal slices and single pre-implantation embryos. Originally designed for low-input samples, the successful application of Ribo-ITP to brain tissue further demonstrates its versatility and broadens its potential utility across diverse biological contexts. However, several aspects of the study would benefit from clarification and refinement to strengthen the manuscript.

• The application of Ribo-ITP to single preimplantation embryos is compelling. Have the authors attempted to validate translon translation using (if available) other proteomic datasets or riboseq datasets - done using conventional approach?
• In Figure 1F, transcriptome data shows “NA” for detected lncRNAs, while genomic analysis identifies 69 novel lncRNAs. Could the authors clarify why these lncRNAs were not detected in the transcriptome-mapped data? Possible explanations such as low expression levels, technical limitations, or filtering criteria should be discussed.
• The statement “Using reporter assays Wu et al., showed that the presence of downstream translons but not the number or sequence positively modulates translation efficiency of the CDS” is unclear. Please elaborate on how the number of translons was tested and include the missing citation for Wu et al.
• The statement “Four out of the seven tested embryonic translons drove at least some GFP expression” is vague. Please quantify GFP expression levels. Additionally, were Rps12, Olfr1217, and Top1 previously studied in this context? If not, did the authors validate the peptides using MS support their translation? Could stress conditions or developmental timing influence translon translation and explain discrepancies?
• While loss-of-function data are presented, and actually the screen approach could also become a versatile tool in the future, gain-of-function experiments would strengthen the conclusions of such assays. If the gene encoding the translon is overexpressed, does this rescue the observed phenotype?
• The use of 16 cycles in cDNA amplification is relatively high and may introduce bias, especially given the variability of input. Please discuss how this might affect data interpretation and whether alternative amplification strategies were considered.
• The manuscript mentions that only 70% of conventional ribo-seq studies achieve similar or better 3-nt periodicity. This implies that a significant proportion of conventional methods outperform Ribo-ITP. Could the authors clarify this comparison and discuss whether the translons identified by Ribo-ITP overlap with those found using conventional ribo-seq?
• Please ensure consistent formatting of figure and table references throughout the manuscript (e.g., change “Fig1F” and “Table S2” to “Figure 1F” and “Table S2”).

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
Ghatpande et al. applied the previously reported Ribo-ITP for difficult-to-collect samples such as micro-dissected hippocampal tissues and 16-cell and 32-cell stages of mouse embryonic development. The authors confirmed the quality of the Ribo-ITP data (3nt-periodicity, for example), and identified potential translons in the 5'and 3'UTR regions. Although the

authors did not confirm expression of the translons in experiments (by the methods other than Ribo-seq) by themselves, the identification of some peptides agreed with the previously reported proteomic data. Furthermore, the authors claimed a positive correlation in expression between translons and CDS, which was in part suggested by the reporter assay. Finally, the authors showed that clonal populations harboring mutations in both of the Phgdh upstream translon alleles showed subtle effects on cell growth.
The results presented in the manuscript might be intriguing, but there are concerns in the novel ty and the experimental design of the manuscript.
1) Novelty: the authors applied the previously reported Ribo-ITP to hippocampal tissues and 16-cell and 32-cell stages of embryos. Although the authors identified many translons, their identification had been difficult and had not provided novel functions of translons. Therefore, the results are mostly observations from the non-novel, previously published technique. The identification of translons from the two tissues were intriguing. However, since the technique has already been reported and many translons have already been identified by new techniques such as Ribo-seq methods for monitoring translation initiation sites, more functional analyses would be required and strengthen the manuscript.
2) Experimental design: The authors claimed that they used a method to induce LTP in CA1 pyramidal interneurons in acute hippocampal slices. It is not clear what types neurons are activated, and the authors did not provide any results of immunostaining to show what types of neurons are activated and what populations of neurons are activated under the condition the authors applied. Furthermore, the authors did not address activity dependence in translation in the manuscript. Therefore, the results are mostly observations and did not provide insights into stimulation-dependent translation upregulation associated with LTP, which should be a quite important theme in the field.
3) Inclusion of the data in Fig. 4 would be highly expected for hippocampal slices.
4) Fig. 5: It will be better to include an endogenous 5'UTR sequence and an endogenous linker between translon and GFP in each reporter construct, as this will be more natural. Alternatively, the authors could explain the design of the reporter constructs.
5) In the growth defect experiments (Fig. 5N), it is unclear whether the growth defects are caused by reduced expression of translon or CDS. In order to address a function of the Phgdh translon in 5UTR, a deletion of the specific translon in the endogenous Phgdh locus will be preferred.

Minor:
6) Figure 3 could be shown as a Supplementary figure. Alternatively, the authors could provide more analytical data such as mRNA structure, codon optimality, and conservation of some major micro-peptides across species.
7) There was no information of the legends of Fig. 5L-N.

(Remarks on code availability)
Details of the code are provided.

Reviewer #3

(Remarks to the Author)
Ribo-ITP expands the translatome of limited input samples

Here, Ghatpande et al perform Ribo-ITP on mouse hippocampal and embryonic samples, with the aim to explore translation in cell types that have remained harder to study despite general advances in the field of non-canonical translation. As well as generating large amounts of data, they identify certain expression patterns, and move on to functional biology work for a number of candidates.

For the positives, I find the technological advances to be compelling, i.e. the demonstration that this work can be done on low-input samples. Also, the generation of hippocampal and embryonic data is in itself important (and it looks to be good quality), with the potential to benefit the work of others, including mouse annotation efforts.

On the negative side, I found myself wanting more of a biological 'story.' Essentially, the manuscript stops short of connecting the discovered translons to concrete biological mechanisms, despite some work in this direction. I do agree with the author's point that the '...functional capacity [of translons] remains challenging to establish,' while the Discussion reasonably points towards downstream work that will inevitably come in later studies. Nonetheless, I certainly think they could do more interpretative work based on the data that they have produced. I'll also suggest how the study might be expanded by additional experimental work, while trying not to set unrealistic expectations as to how far the scope can be extended.

As an overarching point, most of the results and discussion consider translons in aggregate. I'd like to learn more about some individual cases, and it isn't until page 10 that any specific genes are even mentioned. There are brief comments on Eif5 and Clpx uORFs (these driving GFP expression), and then Phgdh in the CRISPR work, and I think that's it. As a general point, I - and probably the reader - would be interested to know more about how their impressive data generation work leads to improved biological understanding of specific genes, ideally tied to specific developmental pathways relating to embryonic / hippocampus development.

Here are some more specific points, in no particular order.

1. 'We selected seven embryonic upstream translons and cloned the sequence…' [p10]. What was the rationale behind these choices? And are these more interesting things to be said or tested about the 4 / 7 that drive GFP? Do any sequence

features predict success (start context, length, codon optimality, peptide charge)?

2. ‘we mined two proteomic datasets focusing on non-canonical translation in mice brains and identified peptide evidence for a total of 47 translons.’ There is no further discussion here, so it reads like something of a loose thread. Are these not interesting candidate microproteins that could be looked at in more detail? How confident are these identifications at the technical level? I appreciate that this is not a ‘proteomics manuscript’. (That said - if it is feasible - targeted PRM/SRM validation of certain ‘best’ candidates would be a nice way to add strong evidence for their existence, and perhaps to test the proposed ‘unstable peptide’ model.)

3. The CRISPR work is a rather short section at the end of the Results. A difficulty with this sort of work is disentangling peptide vs RNA-level effects. The pooled CRISPR screen seems to identify phenotypes, but mutations may alter RNA features rather than peptide function. In other words, phenotypes may result from alterations to 5’ UTR regulatory features, including RNA structures. Base-editing approaches that abolish translation without disrupting sequence context (ideally start codon editing), combined with rescue constructs, would clarify this. Is this feasible, even for the single Phghd case?

4. With these functional assays, the authors are - I think - being led by their data in the choices of candidates to take forward. Could they also subject candidate uORFs found in genes that already have concrete involvements in neuronal or developmental pathways to similar targeted assays? This could broaden the scope of the narrative, taking us more into hypothesis testing. Not essential - and lots more work - but perhaps worth considering.

5. The authors calculate correlations between uORF translation and main CDS translation for many genes. Some genes show positive coupling, others negative, and most are neutral. These results are summarized statistically, but - again - not really linked to biology. Some fairly quick gene set enrichment work could provide illumination here. For example, it would be interesting if positively coupled uORFs were found in genes enriched in specific functional categories (i.e. RNA-binding proteins, transcription factors, etc), and negative couplings enriched in others.

6. The production of embryonic data is a particularly important outcome of this work, but again I feel like there is a lot more that could be leveraged here for novel insight. The only real scientific conclusion is the rather broad statement that uORFs in embryonic genes tend to be widely expressed. Something that could transform this part of the work is an effort to integrate publicly available scRNA datasets from embryonic cells (I believe mouse has good data across different lineages). This could provide better grounding for efforts to examine uORF - CDS coupling (i.e. genes with positive or negative coupling might be seen to be more or less lineage specific as per RNA expression), and it might also be worth examining whether specific uORF features (length, initiation codon etc) show bias to different lineages. Really, any kind of novel insights into the role of translation in embryo biology could be important here.

7. A couple of technical points on the Ribo-seq work. A major challenge we find is how to deal with overlapping ORF calls, which can be due to usage of alternative initiation codons, alternative splicing, etc, but also poor quality calls. The authors are clearly aware of this problem, and have taken steps for disambiguation. I’d just like a few more details to help understand this better. You collapse in-frame ORFs sharing a stop codon by keeping those with ≥75% of the max read support, correct? Then you apply a bias toward canonical ATGs: non-canonical starts are dropped unless the group has only non-canonical starts. Correct? What about overlapping ORFs that don’t share the same stop? These are common in 5’ UTRs. I’d also be interested to learn more about the filtering outcomes here. How often does the 75% rule prune out longer overlapping ORFs? What fraction of predicted ORFs are discarded at this stage? How many near-cognate-only groups remain after this filtering? Do near-cognate-only ORFs pass the same periodicity and reproducibility QC at similar rates as ATG-initiated ORFs?

8. Finally, integrating both transcriptome and genome mapping is an interesting strategy, and probably a good idea given that each has its own strengths and weaknesses. But the fact that they produce different output sets raises the question of what is due to biology and what is down to pipeline behaviour. In fact, the differences between these sets are not really reconciled in the text. More information would be useful here. How many ORFs are shared, and how many are unique to each, and what features distinguish them (e.g. start codon type, ORF length, periodicity score)? Above all, I’d like to understand how many calls are pipeline specific and why.

To conclude, none of these suggestions for additional work are individually essential. My overarching point is that - for the most part - the article reads more like a technical report, or perhaps a resource publication. In that sense it stands up pretty well as it is; the authors have produced some excellent data, and they also note in the abstract that it is a ‘proof of concept study’. This is fair; the methodological advances and data generation are clearly the major part of the work. Still, I would encourage them to go further in the interpretative side of things, and I only labour the point about finding new biological insights because I think they are in a good position to do this.

(Remarks on code availability)
I'm not a coder or developer, so this is going beyond my expertise. There is a readme file, which seems pretty useful.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thanks to the autors for their detailed and conscientious answers to our questions and concerns.
Regarding the question about gain-of-functions studies to strenghten the conclusions, the discussion could be expanded to point these concerns out.
Nevertheless, I think that the Ribo-ITP technique provides a powerful tool for low-input samples and as shown in this study, also provides new possibilities in tissue-related experiments.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
1.Novelty:
The authors acknowledge that functional characterization of the translocons would further strengthen the manuscript. However, they were unable to perform additional functional analyses and instead focused on validation experiments using the GFP reporter screen. In addition, the revised manuscript removed the cell growth data due to concerns regarding reproducibility. To address whether translocon expression can regulate translation of its CDS, the authors newly included systematic mutational prediction data.
While this reviewer understands that regulation of CDS translation may represent one aspect of translocon function, this concept is already well established. Therefore, more specific functional data in cells for at least one novel translocon would be highly desirable to establish sufficient novelty and improve the overall quality of the manuscript. Importantly, several newer Ribo-seq–based approaches, such as Ribo-TISH (which the authors cite) and translation initiation–mediated Ribo-seq methods, have already enabled the identification of numerous novel translocons, some of which have been shown to possess specific cellular functions.
In light of these considerations, the current manuscript primarily presents observational findings using a previously established Ribo-seq approach and appears to provide only a modest incremental advance over prior studies.

2.Technique:
This reviewer appreciates that optimization of RNase I was a key step in improving the Ribo-ITP method, particularly in enabling its application to small sample sizes. However, it has already been well established that digestion with RNase I is critical for Ribo-seq library preparation because it improves the detection of 3-nt periodicity, which is essential for rigorous ORF identification.
Therefore, given that the core framework of Ribo-ITP has already been established, the use and optimization of RNase I instead of MNase in this study—while technically useful—do not appear to constitute a substantial methodological advance suitable for publication in this journal.

3.Biological interpretation:
The authors suggest that the lack of detection of immediate-early genes in the hippocampal dataset following LTP induction may be due to sample heterogeneity arising from difficulties in consistent microdissection. However, translation of immediate-early genes would be expected to increase not only in neurons but also in astrocytes. Therefore, this explanation does not appear fully convincing.
Rather, these observations raise concerns that the physiological state of the brain samples in the culture system or microdissected preparations used in this study may differ from normal conditions. Under such circumstances, there may be significant concerns regarding the reliability of the identified translocons, and the dataset presented in the manuscript may therefore have substantial limitations.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have done good work to improve the manuscript, performing additional analysis and making key changes to the text. I am now satisfied that this work is suitable for publication.

(Remarks on code availability)

Reply to Reviewers Comments:

We thank all three reviewers for their constructive feedback and commenting on the high quality, and potential impact of our work. We have performed new experiments and analyses to address the reviewers comments. The manuscript has also been edited substantially to incorporate all the reviewers comments. We provide a detailed point-by-point response below.

Reviewer 1:

The manuscript by Ghatpande and Paul et al. presents a compelling proof-of-concept study applying their previously developed Ribo-ITP technique to microdissected hippocampal slices and single pre-implantation embryos. Originally designed for low-input samples, the successful application of Ribo-ITP to brain tissue further demonstrates its versatility and broadens its potential utility across diverse biological contexts. However, several aspects of the study would benefit from clarification and refinement to strengthen the manuscript.

We thank the reviewer for their positive assessment and constructive suggestions of our paper. We have addressed all specific points below.

• The application of Ribo-ITP to single preimplantation embryos is compelling. Have the authors attempted to validate translon translation using (if available) other proteomic datasets or riboseq datasets - done using conventional approach?

The only proteomics dataset we are aware of covering these developmental stages is from 2017[1], in which hundreds pooled embryos were collected per stage to quantify protein abundance. We interrogated this dataset for peptides encoded by the translons identified in our study. Despite the large number of embryos analyzed, the reported proteome coverage was limited to approximately 3767 proteins per stage, and we did not detect peptides corresponding to translon encoded products.

While it is possible that these translon derived peptides are expressed at low abundance, it is likely that they were missed due to technical limitations of the proteomic workflow used in this study. Detection of micropeptides typically requires dedicated enrichment and optimized sample preparation strategies, which have been shown to substantially improve recovery of small and low abundance proteins[2,3]. Such approaches were not employed in the 2017 study.

We now explicitly discuss these limitations and the lack of suitable orthogonal proteomic datasets in the revised manuscript (lines 543-547).

• In Figure 1F, transcriptome data shows “NA” for detected lncRNAs, while genomic analysis identifies 69 novel lncRNAs. Could the authors clarify why these lncRNAs were not detected in the transcriptome-mapped data? Possible explanations such as low expression levels, technical limitations, or filtering criteria should be discussed.

The transcriptome reference used for mapping ribosome footprints includes only annotated protein-coding transcripts. As a result, lncRNAs were excluded at the mapping stage, and their absence in the transcriptome-based analysis reflects this filtering criterion rather than lack of expression. We have now clarified this point in the results section of the revised manuscript (lines 151-153).

• The statement “Using reporter assays Wu et al., showed that the presence of downstream translons but not the number or sequence positively modulates translation efficiency of the CDS” is unclear. Please elaborate on how the number of translons was tested and include the missing citation for Wu et al.

We apologize for this unclear sentence. We have now rephrased it as “In *Wu et al.* (2020), the authors showed that the presence of downstream translation events (which they refer to as translated dORFs) enhances translation efficiency of the canonical coding sequence by using fluorescent reporter assays. They further demonstrated that increasing the number of dORFs correlates with progressively higher main ORF translation by comparing reporters with one, two, or multiple translated dORFs and observing additive effects on reporter output[4].” and added the associated citation. (lines 356-363)

• The statement “Four out of the seven tested embryonic translons drove at least some GFP expression” is vague. Please quantify GFP expression levels.

We agree with the reviewer that the current images are not quantitative evidence of translation. Very few cells express GFP as well as many translons show dim GFP expression, this makes quantitative analysis unreliable. Therefore we decided to present this data qualitatively. For better quantitative assessment, we have now performed a high throughput assay. We expanded our initial set of 7 (in the original submission) to a library of 443 translons. Of these, 378 were detected with at least 1 read count in the GFP positive population, indicating that the majority of translons exhibit at least some translational activity. This observation is consistent with their detectability by ribosome profiling. To identify translons with robust translational output, we compared normalized read counts between the GFP positive and Input populations. Using mean log2 counts greater than 2 and mean log2 fold change greater than 2 as thresholds, we identified 51 translons that were enriched in the GFP positive population. This enabled us to analyze

sequence features associated within that population as requested by the reviewer. We found that translons enriched for GFP expression were significantly more likely to initiate with an ATG start codon and not overlap the annotated CDS region.
We have now moved the original GFP images to supplemental figure 4 and the GFP screen data is presented in figure 4 in the revised manuscript. (lines 274-308)

Additionally, were Rps12, Olfr1217, and Top1 previously studied in this context? If not, did the authors validate the peptides using MS support their translation?

To the best of our knowledge, there were no previous studies regarding the upstream translons in Rps12, Olfr1217 and Top1 genes. We have not performed peptide validation by MS.

Could stress conditions or developmental timing influence translon translation and explain discrepancies?

We thank the reviewer for raising this important point. It is definitely possible that translons might be translated in specific conditions or developmental stages. We have now mentioned this in the discussion section of the paper (lines 532-536).

• While loss-of-function data are presented, and actually the screen approach could also become a versatile tool in the future, gain-of-function experiments would strengthen the conclusions of such assays. If the gene encoding the translon is overexpressed, does this rescue the observed phenotype?

We agree that gain of function experiments could, in principle, strengthen these conclusions. We performed overexpression of the Phgdh upstream translon using a T2A GFP reporter construct in Phgdh mutant translon clonal mESC lines. However, in these experiments we were unable to reproducibly observe the growth defect seen in the original mutant clone.

We suspect that this reflects phenotypic adaptation of the clonal lines following prolonged culture after translon disruption, which may mask short term growth effects. Given the lack of long term reproducibility across passages, we felt that inclusion of these overexpression data would be difficult to interpret. We have therefore removed the single clone growth curve data from the revised manuscript.

• The use of 16 cycles in cDNA amplification is relatively high and may introduce bias, especially given the variability of input. Please discuss how this might affect data interpretation and whether alternative amplification strategies were considered.

Thanks for raising this issue. We empirically tested multiple PCR cycle numbers and selected the minimum number of cycles required to obtain sufficient library material, which varied depending on the input source. This rationale and optimization are now explicitly described in the Methods section.

Importantly, all libraries incorporated unique molecular identifiers during library preparation. This enables differentiation of individual ribosome footprints from PCR duplicates, allowing amplification bias to be mitigated computationally. Across libraries, deduplication rates ranged from approximately 10 to 60%. We have now provided these details in the results section (lines 649-650).

• The manuscript mentions that only 70% of conventional ribo-seq studies achieve similar or better 3-nt periodicity. This implies that a significant proportion of conventional methods outperform Ribo-ITP. Could the authors clarify this comparison and discuss whether the translons identified by Ribo-ITP overlap with those found using conventional ribo-seq?

We appreciate the opportunity to clarify this point. We agree that our original wording was ambiguous and may have implied that a substantial fraction of conventional ribosome profiling datasets outperform Ribo ITP. This was not our intent. We have now added a new Supplementary Figure 3 that explicitly places the 3-nt periodicity of our Ribo-ITP libraries within the empirical distribution observed across a large collection of conventional ribosome profiling studies. This figure indicates that our data fall well within the typically high periodicity range achieved by standard approaches.

Our goal is to emphasize that Ribo-ITP achieves periodicity comparable to bulk ribosome profiling despite using orders of magnitude less input material. For example in single cell RNA sequencing, reduced input generally leads to reduced coverage. The fact that we can recover high quality ribosome footprints from such limited material supports the robustness of the method.

Regarding overlap with translons identified by conventional ribosome profiling, we analyzed a large compendium of bulk datasets using RiboBase and found that the translons detected by Ribo-ITP are also expressed in many conventional ribosome profiling experiments (Figure 5). We have added these points to the Discussion section of the revised manuscript lines 481-483.

• Please ensure consistent formatting of figure and table references throughout the manuscript (e.g., change “Fig1F” and “Table S2” to “Figure 1F” and “Table S2”).

Thanks for pointing out these, we have ensured consistent formatting in the revised manuscript.

Reviewer 2:

Ghatpande et al. applied the previously reported Ribo-ITP for difficult-to-collect samples such as micro-dissected hippocampal tissues and 16-cell and 32-cell stages of mouse embryonic development. The authors confirmed the quality of the Ribo-ITP data (3nt-periodicity, for example), and identified potential translons in the 5'and 3'UTR regions. Although the authors did not confirm expression of the translons in experiments (by the methods other than Ribo-seq) by themselves, the identification of some peptides agreed with the previously reported proteomic data. Furthermore, the authors claimed a positive correlation in expression between translons and CDS, which was in part suggested by the reporter assay. Finally, the authors showed that clonal populations harboring mutations in both of the Phgdh upstream translon alleles showed subtle effects on cell growth.

The results presented in the manuscript might be intriguing, but there are concerns in the novelty and the experimental design of the manuscript.

We thank the reviewer for their comments that we address in detail below. Specifically, we have now added additional experimental evidence (transcriptome-wide reporter screen) to confirm the expression of translons, we revised the text to more clearly articulate the novelty, and provided additional support regarding the experimental design.

1) Novelty: the authors applied the previously reported Ribo-ITP to hippocampal tissues and 16-cell and 32-cell stages of embryos. Although the authors identified many translons, their identification had been difficult and had not provided novel functions of translons. Therefore, the results are mostly observations from the non-novel, previously published technique. The identification of translons from the two tissues were intriguing. However, since the technique has already been reported and many translons have already been identified by new techniques such as Ribo-seq methods for monitoring translation initiation sites, more functional analyses would be required and strengthen the manuscript.

We agree that functional characterization of translons would further strengthen the manuscript. To address this, we have substantially expanded the scope of functional analysis in the revised version by adding two new components. First, we performed a high throughput GFP reporter screen to quantitatively assess the translational capacity of hundreds of translons identified. Second, we implemented a systematic mutational prediction framework to test whether subsets of translons act as regulatory elements modulating CDS translation.

We respectfully disagree with the characterization of this study as observations from a "non-novel, previously published technique". While Ribo-ITP was developed in Ozadam et al. 2023 *Nature*, that work focused largely on MNase based ribosome footprints, which are not suitable for translon discovery due to reduced periodicity and imprecise footprint boundaries. Similarly, the only other ribosome profiling method demonstrated in single cell contexts (VanInsberghe et al. 2021 *Nature*) is also MNase based and therefore not compatible with translon identification.

In the present study, we optimized Ribo-ITP using RNase I, which is a critical technical improvement that yields the 3-nt periodicity required for reliable translon discovery. To the best of our knowledge, translon identification by ribosome profiling from ultra low input samples such as preimplantation embryos or hippocampal tissue has not been previously achieved. Existing methods for translon detection, including initiation site profiling approaches, require substantially higher input material and have not been applicable to these biological contexts.

More broadly, this work serves as a proof of concept demonstrating that high quality ribosome profiling and translon discovery are feasible in cellular contexts that were previously inaccessible. Analogous to the widespread adoption of scRNA-seq after its initial introduction, applications of enabling technologies to new biological systems often uncover novel findings even when the underlying methodology has been previously established. We have revised the framing of the manuscript to articulate this point more clearly and to emphasize that the primary novelty lies in enabling translon discovery in low input systems while providing initial functional evidence to guide future mechanistic studies.

2) Experimental design: The authors claimed that they used a method to induce LTP in CA1 pyramidal interneurons in acute hippocampal slices. It is not clear what types neurons are activated, and the authors did not provide any results of immunostaining to show what types of neurons are activated and what populations of neurons are activated under the condition the authors applied. Furthermore, the authors did not address activity dependence in translation in the manuscript. Therefore, the results are mostly observations and did not provide insights into stimulation-dependent translation upregulation associated with LTP, which should be a quite important theme in the field.

The type of stimulation used (theta burst stimulation) activates two major types of neurons, CA1 pyramidal neurons (~89%) and GABAergic interneurons (~11%) in the CA1 region of the hippocampus. CA1 pyramidal neurons are relatively homogeneous; however, multiple subtypes of GABAergic interneurons are present in this region. Our

dissection methods further increase the proportion of pyramidal neurons represented in our data but as some interneuron cell bodies lie in the stratum pyramidale layer (in particular parvalbumin+ interneurons), they can not be fully excluded[5]. Because we cut the axon tracts prior to stimulation, these are the only neuron types activated. Additionally, up to 4-8% of the volume of stratum radiatum is astrocytes[6] which will be activated by the stimulation and may contribute to the data. Again, our dissection will reduce the proportional contribution of astrocytes by removing many of the cell bodies, but we cannot discount them entirely. We routinely carry out similar LTP induction in the lab and published multiple electrophysiological studies using this experimental design in the past [7–9].

Activity-dependent translational regulation in neurons in general as well as in the context of LTP and long term memory has been well-studied for a very long time [10–20] However, activity dependent translon translation is not well understood. Therefore instead of focusing on stimulation- dependent translational changes associated with LTP, we wanted to identify if there are any translons that are expressed in this context. However our data did not show consistent changes in global gene expression associated with LTP (the reasons for which are discussed in the results section).

We have now added the details of stimulation-dependent translational changes in the revised manuscript in lines 87-90 and lines 100-106.

3) Inclusion of the data in Fig. 4 would be highly expected for hippocampal slices.

The Ribobase based analysis of hippocampal translon expression has been presented in the revised manuscript as Fig 5.

4) Fig. 5: It will be better to include an endogenous 5'UTR sequence and an endogenous linker between translon and GFP in each reporter construct, as this will be more natural. Alternatively, the authors could explain the design of the reporter constructs.

We thank the reviewer for this question. We agree with the reviewer that using endogenous 5'UTR sequence is the better experimental design choice. Hence, we have constructed our reporter to use endogenous 5'UTR sequences that we have found to contain a translon. We then fused the GFP in frame with a translon whose stop codon was removed such that it would only be expressed if we were able to identify the correct coding frame of the translon. Note that there is no linker sequence between the translon and GFP in the reporter construct. To further generalize the findings from these experiments, we have now expanded our initial set of 7 (in the original submission) to a library of 443. Of these, 378 were detected with at least 1 read count in the GFP positive population, indicating that the majority of translons exhibit at least some translational activity. This observation is consistent with their detectability by ribosome profiling. To

identify translons with robust translational output, we compared normalized read counts between the GFP positive and Input populations. Using mean log2 counts greater than 2 and mean log2 fold change greater than 2 as thresholds, we identified 51 translons that were enriched in the GFP positive population. This enabled us to analyze sequence features associated with that population as requested by the reviewer. We found that translons enriched for GFP expression were significantly more likely to initiate with an ATG start codon and not overlap annotated CDS regions.

These data are now incorporated into revised Figure 4 and Figure S4 and directly address which translons robustly drive translation and which sequence features predict success in the reporter assay. (lines 274-308)

5) In the growth defect experiments (Fig. 5N), it is unclear whether the growth defects are caused by reduced expression of translon or CDS. In order to address a function of the Phgdh translon in 5UTR, a deletion of the specific translon in the endogenous Phgdh locus will be preferred.

Thanks for the suggestion and we agree that further characterization would be necessary to establish the impact of this particular translon. We believe that the main strength of our work is in establishing a powerful approach to translon discovery in rare populations rather than focusing on a particular translon. As such, we conducted additional experiments as described above to strengthen the main focus of the study and decided to de-emphasized Phgdh specific experiments in the revised manuscript. We appreciate the reviewer's comments, which helped us sharpen the core message of the study.

Minor:
6) Figure 3 could be shown as a Supplementary figure. Alternatively, the authors could provide more analytical data such as mRNA structure, codon optimality, and conservation of some major micro-peptides across species.

We believe Figure 3 to be an essential component of our study as it provides an overview of sequence features of all the translons. We appreciate the suggestion about providing analytical data on these translons. There is growing evidence that conservation of functional translons can be highly varied [21] i.e. some translons can be conserved from mouse to humans[22] while others are found only in a single species[23]. Furthermore, although codon optimality and mRNA structure provide support for regulatory translons as ribosomes might stall at suboptimal codons and affect translation of the CDS, they do not inform about function of the encoded peptides. Hence, we decided not to add these additional metrics to avoid diluting the core focus of the manuscript.

7) There was no information of the legends of Fig. 5L-N.

We apologize for this oversight, this has been corrected in the revised manuscript.

(Remarks on code availability)

Details of the code are provided.

Reviewer 3:
Ribo-ITP expands the translatome of limited input samples

Here, Ghatpande et al perform Ribo-ITP on mouse hippocampal and embryonic samples, with the aim to explore translation in cell types that have remained harder to study despite general advances in the field of non-canonical translation. As well as generating large amounts of data, they identify certain expression patterns, and move on to functional biology work for a number of candidates. For the positives, I find the technological advances to be compelling, i.e. the demonstration that this work can be done on low-input samples. Also, the generation of hippocampal and embryonic data is in itself important (and it looks to be good quality), with the potential to benefit the work of others, including mouse annotation efforts.

We thank the reviewer for appreciating the technical advances that our study presents and its impact on the field of non-canonical translation in general.

On the negative side, I found myself wanting more of a biological 'story.' Essentially, the manuscript stops short of connecting the discovered translons to concrete biological mechanisms, despite some work in this direction. I do agree with the author's point that the '...functional capacity [of translons] remains challenging to establish,' while the Discussion reasonably points towards downstream work that will inevitably come in later studies. Nonetheless, I certainly think they could do more interpretative work based on the data that they have produced. I'll also suggest how the study might be expanded by additional experimental work, while trying not to set unrealistic expectations as to how far the scope can be extended.

We thank the reviewer for their constructive suggestions that have strengthened the manuscript. We have carried out additional experiments as well as a major rewrite to address these concerns as detailed below and in the revised manuscript.

As an overarching point, most of the results and discussion consider translons in aggregate. I'd like to learn more about some individual cases, and it isn't until page 10 that any specific genes are even mentioned. There are brief comments on Eif5 and Clpx uORFs (these driving GFP expression), and then Phgdh in the CRISPR work, and I think that's it. As a general point, I - and probably the reader - would be interested to know more about how their impressive data generation work leads to improved biological understanding of specific genes, ideally tied to specific developmental pathways relating to embryonic / hippocampus development.

We thank the reviewer for this helpful suggestion. We agree that the first version of the manuscript emphasized translons in aggregate and did not sufficiently connect our data to specific genes and biological pathways. In the revised manuscript, we have restructured the Results and Discussion to highlight representative individual cases.

Specifically, we now highlight translons associated with genes involved in key developmental pathways relevant to the stages analyzed, including examples related to epilepsy, intellectual disability and frontotemporal dementia. These include expanded discussion of translons that drive GFP reporter expression in lines 299-308, translons showing high translon/CDS partial correlations in lines 408-416 and translons with peptidomic evidence in lines 429-435. We also discuss the regulatory functions of translons in specific genes in lines 461-474.

Here are some more specific points, in no particular order.

1. 'We selected seven embryonic upstream translons and cloned the sequence…' [p10]. What was the rationale behind these choices? And are these more interesting things to be said or tested about the 4 / 7 that drive GFP? Do any sequence features predict success (start context, length, codon optimality, peptide charge)?

We thank the reviewer for raising this point. The initial seven embryonic upstream translons were selected without prior filtering, largely to demonstrate feasibility of the reporter assay. More importantly, the small number of tested translons preclude the types of sequence feature analyses that would be highly informative as indicated by the reviewer.

To directly address this limitation, we performed a high throughput GFP reporter screen encompassing the vast majority of translons identified in our dataset (the missing ones are due to technical limitations of DNA synthesis). Following our reporter design, we cloned the library of translons into a lentiviral vector under the EF1 promoter with a puromycin selectable marker and expressed the library as a pooled population in mESCs. After puromycin selection and recovery, we collected an Input population and sorted the remaining cells to isolate the GFP positive ones.

We detected 443 translons reproducibly in the Input population across replicates. Of these, 378 were detected with at least 2 read counts in the GFP positive population, indicating that the majority of translons exhibit at least some translational activity. This observation is consistent with their detectability by ribosome profiling. To identify

translons with strong translational output, we compared normalized read counts between the GFP positive and Input populations. Using mean log2 counts greater than 2 and mean log2 fold change greater than 2 as thresholds, we identified 51 translons that were enriched in the GFP positive population. This enabled us to analyze sequence features associated with that population as requested by the reviewer. We found that translons enriched for GFP expression were significantly more likely to initiate with an ATG start codon and not overlap annotated CDS regions.

These data are now incorporated into revised Figure 4 and Figure S4 and directly address which translons robustly drive translation and which sequence features predict success in the reporter assay. (lines 274-308)

2. 'we mined two proteomic datasets focusing on non-canonical translation in mice brains and identified peptide evidence for a total of 47 translons.' There is no further discussion here, so it reads like something of a loose thread. Are these not interesting candidate microproteins that could be looked at in more detail? How confident are these identifications at the technical level? I appreciate that this is not a 'proteomics manuscript'. (That said - if it is feasible - targeted PRM/SRM validation of certain 'best' candidates would be a nice way to add strong evidence for their existence, and perhaps to test the proposed 'unstable peptide' model.)

Thanks for the suggestion. We have expanded our discussion of the 47 translons that have proteomic evidence. We identified multiple micropeptides originating from translons on mRNAs encoding kinases such as *Map3k9*, *Map3k19*, *Wnk1* and *Bmpr2* (21 mRNAs) and phosphatases including *Pten*, *Ppm1b*, *Dusp3*, and *Ppp1r9b* (9 mRNAs). Synaptic signalling is fundamentally driven by kinase- and phosphatase-dependent phosphorylation cycles, setting the stage for rapid activity dependent modulations[24–26]. Micropeptides have been previously identified to participate in protein phosphorylation[27–30]. As such, these micropeptides are potential candidates for further characterization in future work. We have added this in the results section of the revised manuscript (lines 4429-435).

3. The CRISPR work is a rather short section at the end of the Results. A difficulty with this sort of work is disentangling peptide vs RNA-level effects. The pooled CRISPR screen seems to identify phenotypes, but mutations may alter RNA features rather than peptide function. In other words, phenotypes may result from alterations to 5' UTR regulatory features, including RNA structures. Base-editing approaches that abolish translation without disrupting sequence context (ideally start codon editing), combined with rescue constructs, would clarify this. Is this feasible, even for the single Phghd case?

We agree with the reviewer that the pooled CRISPR screen cannot, on its own, distinguish peptide mediated effects from RNA level regulatory effects such as altered 5′ UTR structure. To address this experimentally, we attempted to isolate targeted mutations in the Phgdh upstream translon. However, growth curve analysis of independently derived mutant clones did not reproducibly recapitulate the phenotype observed in the pooled screen in Supplementary Figure 4S. We suspect this reflects adaptation following prolonged culture of clonal lines, which may mask translon associated effects. Due to this lack of reproducibility, we have removed the single clone growth curve data from the revised manuscript.

We thank the reviewer for the suggestion of start codon specific editing to abrogate translation while preserving RNA context. We agree this is a very promising experimental strategy, however systematic base editing across many translons is not feasible within the scope of this study. Instead, we implemented a complementary computational approach to directly test this idea at scale.

Specifically, we used RiboNN, a sequence based deep learning model that can accurately predict translation efficiency[31]. For each translon, we mutated the start codon to a stop codon in silico and computed the resulting change in CDS translation efficiency (delta TE). As a control, we introduced stop codon mutations at five random positions in the same 5′ UTR but outside the translon. This allowed us to nominate effects that are likely attributable to loss of translon translation rather than UTR sequence disruption.

We observed that mutating the start codon of upstream translons results in a significant change in TE_{CDS} while mutating the downstream translons do not affect TE_{CDS}. Moreover, upstream translons that suppress TE_{CDS} tend to initiate at ATG codons while those that promote TE_{CDS} tend to initiate at near cognate start codons.

To test our predictions experimentally, we selected the upstream translon present in the Chornicon receptor *Cnih2* gene, which is an auxiliary subunit of the ionotropic glutamate receptor of the AMPA subtype that is widely expressed in the CA1 region of the hippocampus and helps in receptor trafficking and regulates its gating properties[32,33]. When we experimentally mutated the upstream translon in the Cnih2 transcript which was one of the top suppressors according to our predictions, we saw an increase in Cnih2 CDS expression in HEK293T cells. This confirms that some of the translons identified in our experiments might act as regulatory translons and control CDS translation. We have now included these new results in the revised manuscript (Fig 5, lines 437-474).

4. With these functional assays, the authors are - I think - being led by their data in the choices of candidates to take forward. Could they also subject candidate uORFs found in genes that already have concrete involvements in neuronal or developmental pathways to similar targeted assays? This could broaden the scope of the narrative, taking us more into hypothesis testing. Not essential - and lots more work - but perhaps worth considering.

As the reviewer notes, functional interrogation of individual translons would require substantial additional experimental work and would constitute a focused follow up. We have therefore revised the framing of the manuscript to highlight our primary goal of establishing a robust approach for identifying translons from ultra low input samples. We view this methodological advance as the core contribution of the study. We now discuss such hypothesis driven translon analyses as an important future direction in the Discussion section lines 520-521 and 530-536.

5. The authors calculate correlations between uORF translation and main CDS translation for many genes. Some genes show positive coupling, others negative, and most are neutral. These results are summarized statistically, but - again - not really linked to biology. Some fairly quick gene set enrichment work could provide illumination here. For example, it would be interesting if positively coupled uORFs were found in genes enriched in specific functional categories (i.e. RNA-binding proteins, transcription factors, etc), and negative couplings enriched in others.

Thanks for this suggestion. We analyzed the functional annotations for the transcripts showing high and low partial correlations using Gene Ontology terms. We observed that the translons showing positive coupling encoded proteins were involved in synaptic transmission and signalling, and transmitter gated ion channel activity. While the translon showing negative coupling encodes proteins involved in neurodegenerative disorders like spinocerebellar ataxia and amylotropic lateral sclerosis/rontotemporal dementia. Suggesting that these mRNAs might have distinct roles in neuronal processes. We have discussed this in the revised manuscript in lines 408-416.

6. The production of embryonic data is a particularly important outcome of this work, but again I feel like there is a lot more that could be leveraged here for novel insight. The only real scientific conclusion is the rather broad statement that uORFs in embryonic genes tend to be widely expressed. Something that could transform this part of the work is an effort to integrate publicly available scRNA datasets from embryonic cells (I believe mouse has good data across different lineages). This could provide better grounding for efforts to examine uORF - CDS coupling (i.e. genes with positive or negative coupling might be seen to be more or less lineage specific as per RNA expression), and it might

also be worth examining whether specific uORF features (length, initiation codon etc) show bias to different lineages. Really, any kind of novel insights into the role of translation in embryo biology could be important here.

To gain insights into the role of translons in embryo development, we reanalysed data from two mouse preimplantation embryo scRNAseq datasets and a publically available dataset (CellXGene)[34,35]. Deng et al. covers preimplantation stages from zygote until late blastula while Mohammed et al. used preimplantation stage (E3.5), implantation stage (E4.5-5.5) and post implantation embryos (E6.5). We also used the CellXgene database which houses scRNAseq data for 2-6 hour intervals from gastrula to birth. We hypothesized that these data may facilitate identifying translons with lineage specific expression. However, we did not identify strong enrichment for lineage specific translon expression. Nevertheless, we agree with the reviewer that an integrative analysis between the identified translons with appropriate scRNAseq datasets is a promising approach to inform about translon functions. We have discussed this point in the revised discussion section (lines 559-562).

7. A couple of technical points on the Ribo-seq work. A major challenge we find is how to deal with overlapping ORF calls, which can be due to usage of alternative initiation codons, alternative splicing, etc, but also poor quality calls. The authors are clearly aware of this problem, and have taken steps for disambiguation. I'd just like a few more details to help understand this better.

We have broken down the comment into individual questions and answered them individually below.

You collapse in-frame ORFs sharing a stop codon by keeping those with ≥75% of the max read support, correct?

The reviewer's description of the individual filtering steps is correct.

Then you apply a bias toward canonical ATGs: non-canonical starts are dropped unless the group has only non-canonical starts. Correct?

The canonical start codon preference is applied first, followed by the 75% maximum reads filter. We have updated the methods section to reflect the workflow. (lines 677-679)

What about overlapping ORFs that don't share the same stop? These are common in 5' UTRs.

Overlapping ORFs that do not share the same stop codon belong are retained as distinct translons. (lines 671-676)

I'd also be interested to learn more about the filtering outcomes here. How often does the 75% rule prune out longer overlapping ORFs? What fraction of predicted ORFs are discarded at this stage? How many near-cognate-only groups remain after this filtering? Do near-cognate-only ORFs pass the same periodicity and reproducibility QC at similar rates as ATG-initiated ORFs?

We have now added the details in supplemental figure S3 and discussed them in the relevant results sections. The figures display the number and start codon composition of the ORFs remaining after each filter.

8. Finally, integrating both transcriptome and genome mapping is an interesting strategy, and probably a good idea given that each has its own strengths and weaknesses. But the fact that they produce different output sets raises the question of what is due to biology and what is down to pipeline behaviour. In fact, the differences between these sets are not really reconciled in the text. More information would be useful here. How many ORFs are shared, and how many are unique to each, and what features distinguish them (e.g. start codon type, ORF length, periodicity score)? Above all, I'd like to understand how many calls are pipeline specific and why.

Based on the reviewer's comment, we assessed the overlap between the hippocampal translons identified by the two methods. As we had few embryonic translons to begin with we focused on the neuronal translons. The translon naming by RiboTISH and our pipeline is different therefore we looked at the overlap between the genomic coordinates of the translons identified by both the pipelines.We kept a 10bp leeway (translons within 10bp distance were considered the same) when looking at the genomic coordinate positions. There was a 34% overlap between the upstream and downstream translons identified by the two methods. There was no clear bias towards the amino acid length but ATG initiating and upstream non-overlapping translons were overrepresented suggesting that the method used to identify translons has an impact on translon calling. We have added a line describing this in the results section (lines 149-151) of the revised manuscript.

To conclude, none of these suggestions for additional work are individually essential. My overarching point is that - for the most part - the article reads more like a technical report, or perhaps a resource publication. In that sense it stands up pretty well as it is; the authors have produced some excellent data, and they also note in the abstract that it is a 'proof of concept study'. This is fair; the methodological advances and data generation are

clearly the major part of the work. Still, I would encourage them to go further in the interpretative side of things, and I only labour the point about finding new biological insights because I think they are in a good position to do this.

References:

1. Protein Expression Landscape of Mouse Embryos during Pre-implantation Development. *Cell Reports* **21**, 3957–3969 (2017).
2. Leong, A. Z.-X. *et al.* Short open reading frames (sORFs) and microproteins: an update on their identification and validation measures. *J. Biomed. Sci.* **29**, 19 (2022).
3. Mohsen, J. J., Martel, A. A. & Slavoff, S. A. Microproteins-Discovery, structure, and function. *Proteomics* **23**, e2100211 (2023).
4. Wu, Q. *et al.* Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J.* **39**, e104763 (2020).
5. Bezaire, M. J. & Soltesz, I. Quantitative assessment of CA1 local circuits: knowledge base for interneuron-pyramidal cell connectivity. *Hippocampus* **23**, 751–785 (2013).
6. Refaeli, R. *et al.* Features of hippocampal astrocytic domains and their spatial relation to excitatory and inhibitory neurons. *Glia* **69**, 2378–2390 (2021).
7. Chancey, J. H., Ahmed, A. A., Guillén, F. I., Ghatpande, V. & Howard, M. A. Complex synaptic and intrinsic interactions disrupt input/output functions in the hippocampus of Scn1b knock-out mice. *J. Neurosci.* **43**, 8562–8577 (2023).
8. Chancey, J. H. & Howard, M. A. Synaptic integration in CA1 pyramidal neurons is intact despite deficits in GABAergic transmission in the Scn1a haploinsufficiency mouse model of Dravet syndrome. *eNeuro* **9**, ENEURO.0080–22.2022 (2022).
9. Howard, M. A., Rubenstein, J. L. R. & Baraban, S. C. Bidirectional homeostatic plasticity induced by interneuron cell death and transplantation in vivo. *Proc. Natl.*

Acad. Sci. U. S. A. **111**, 492–497 (2014).

10. Buffington, S. A., Huang, W. & Costa-Mattioli, M. Translational control in synaptic plasticity and cognitive dysfunction. *Annu Rev Neurosci* **37**, 17–38 (2014).
11. McGaugh, J. L. Memory--a Century of Consolidation. *Science* (2000) doi:10.1126/science.287.5451.248.
12. Science. *AAAS* https://www.science.org/doi/pdf/10.1126/science.1067020.
13. Sun, C. *et al.* The prevalence and specificity of local protein synthesis during neuronal synaptic plasticity. *Sci. Adv.* **7**, eabj0790 (2021).
14. Glock, C. *et al.* The translatome of neuronal cell bodies, dendrites, and axons. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2113929118 (2021).
15. Perez, J. D., Fusco, C. M. & Schuman, E. M. A functional dissection of the mRNA and locally synthesized protein population in neuronal dendrites and axons. *Annu. Rev. Genet.* **55**, 183–207 (2021).
16. Sun, C. & Schuman, E. A multi-omics view of neuronal subcellular protein synthesis. *Curr. Opin. Neurobiol.* **80**, 102705 (2023).
17. Santini, E., Huynh, T. N. & Klann, E. Mechanisms of translation control underlying long-lasting synaptic plasticity and the consolidation of long-term memory. *Prog. Mol. Biol. Transl. Sci.* **122**, 131–167 (2014).
18. Shrestha, P. & Klann, E. Spatiotemporally resolved protein synthesis as a molecular framework for memory consolidation. *Trends Neurosci.* **45**, 297–311 (2022).
19. Borovok, N. *et al.* Dynamics of hippocampal protein expression during long-term spatial memory formation. *Mol. Cell. Proteomics* **15**, 523–541 (2016).

20. Effect of acetoxycycloheximide and of an acetoxycycloheximide-puromycin mixture on cerebral protein synthesis and memory in mice. http://dx.doi.org/10.1073/pnas.55.2.369 doi:10.1073/pnas.55.2.369.
21. Pai, V. J. *et al.* Microprotein-encoding RNA regulation in cells treated with pro-inflammatory and pro-fibrotic stimuli. *BMC Genomics* **25**, 1034 (2024).
22. Zhang, S. *et al.* Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly. *Nat. Commun.* **11**, 1312 (2020).
23. Wacholder, A. *et al.* A vast evolutionarily transient translatome contributes to phenotype and fitness. *Cell Syst.* **14**, 363–381.e8 (2023).
24. Malinow, R., Schulman, H. & Tsien, R. W. Inhibition of postsynaptic PKC or CaMKII blocks induction but not expression of LTP. *Science* **245**, 862–866 (1989).
25. Genoux, D. *et al.* Protein phosphatase 1 is a molecular constraint on learning and memory. *Nature* **418**, 970–975 (2002).
26. Chang, J.-Y. *et al.* CaMKII Autophosphorylation Is Necessary for Optimal Integration of Ca Signals during LTP Induction, but Not Maintenance. *Neuron* **94**, 800–808.e4 (2017).
27. Wang, Y. *et al.* LncRNA-encoded polypeptide ASRPS inhibits triple-negative breast cancer angiogenesis. *J Exp Med* **217**, (2020).
28. Guo, B. *et al.* Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. *EMBO J* **39**, e102190 (2020).
29. Pang, Y. *et al.* Peptide SMIM30 promotes HCC development by inducing SRC/YES1 membrane anchoring and MAPK pathway activation. *J Hepatol* **73**, 1155–1169 (2020).

30. Meng, K. *et al.* LINC00493-encoded microprotein SMIM26 exerts anti-metastatic activity in renal cell carcinoma. *EMBO Rep* **24**, e56282 (2023).
31. Zheng, D. *et al.* Predicting the translation efficiency of messenger RNA in mammalian cells. *Nat. Biotechnol.* (2025) doi:10.1038/s41587-025-02712-x.
32. Shi, Y. *et al.* Functional comparison of the effects of TARPs and cornichons on AMPA receptor trafficking and gating. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16315–16319 (2010).
33. Kato, A. S. *et al.* Hippocampal AMPA receptor gating controlled by both TARP and cornichon proteins. *Neuron* **68**, 1082–1096 (2010).
34. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
35. Mohammed, H. *et al.* Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. *Cell Rep.* **20**, 1215–1228 (2017).

REVIEWERS' COMMENTS

We thank all three reviewers for their constructive feedback and commenting on the high quality, and potential impact of our work. The manuscript has been edited to incorporate all the reviewers comments. We provide a detailed point-by-point response below.

Reviewer #1 (Remarks to the Author):

Thanks to the authors for their detailed and conscientious answers to our questions and concerns.
Regarding the question about gain-of-functions studies to strenghten the conclusions, the discussion could be expanded to point these concerns out.
Nevertheless, I think that the Ribo-ITP technique provides a powerful tool for low-input samples and as shown in this study, also provides new possibilities in tissue-related experiments.

Thanks. We have added this as a future direction in the discussion.

Reviewer #2 (Remarks to the Author):

1.Novelty:
The authors acknowledge that functional characterization of the translocons would further strengthen the manuscript. However, they were unable to perform additional functional analyses and instead focused on validation experiments using the GFP reporter screen. In addition, the revised manuscript removed the cell growth data due to concerns regarding reproducibility. To address whether translocon expression can regulate translation of its CDS, the authors newly included systematic mutational prediction data.
While this reviewer understands that regulation of CDS translation may represent one aspect of translocon function, this concept is already well established. Therefore, more specific functional data in cells for at least one novel translocon would be highly desirable to establish sufficient novelty and improve the overall quality of the manuscript. Importantly, several newer Ribo-seq–based approaches, such as Ribo-TISH (which the authors cite) and translation initiation–mediated Ribo-seq methods, have already enabled the identification of numerous novel translocons, some of which have been shown to possess specific cellular functions.
In light of these considerations, the current manuscript primarily presents observational findings using a previously established Ribo-seq approach and appears to provide only a modest incremental advance over prior studies.

2.Technique:
This reviewer appreciates that optimization of RNase I was a key step in improving the Ribo-ITP method, particularly in enabling its application to small sample sizes. However, it has already been well established that digestion with RNase I is critical for Ribo-seq library preparation because it improves the detection of 3-nt periodicity, which is essential for rigorous ORF identification.

Therefore, given that the core framework of Ribo-ITP has already been established, the use and optimization of RNase I instead of MNase in this study—while technically useful—do not appear to constitute a substantial methodological advance suitable for publication in this journal.

3.Biological interpretation:
The authors suggest that the lack of detection of immediate-early genes in the hippocampal dataset following LTP induction may be due to sample heterogeneity arising from difficulties in consistent microdissection. However, translation of immediate-early genes would be expected to increase not only in neurons but also in astrocytes. Therefore, this explanation does not appear fully convincing.
Rather, these observations raise concerns that the physiological state of the brain samples in the culture system or microdissected preparations used in this study may differ from normal conditions. Under such circumstances, there may be significant concerns regarding the reliability of the identified translocons, and the dataset presented in the manuscript may therefore have substantial limitations.

We agree with the reviewer regarding the possibility of cell types other than neurons contributing to the signal. We have now mentioned this explicitly in the main text and added this as a caveat of the study.

Reviewer #3 (Remarks to the Author):

The authors have done good work to improve the manuscript, performing additional analysis and making key changes to the text. I am now satisfied that this work is suitable for publication.

We thank the reviewer for their positive assessment of our paper.