Charting DENR-dependent translation reinitiation uncovers predictive uORF features and links to circadian timekeeping via *Clock*

Violeta Castelo-Szekely[1], Mara De Matos[1], and David Gatfield[1,*]

[1]Center for Integrative Genomics. University of Lausanne. Genopode. 1015 Lausanne, Switzerland.
[*]Corresponding author. David.Gatfield@unil.ch

Abstract

The non-canonical initiation factor DENR (Density regulated protein) promotes translation reinitiation on mRNAs harbouring upstream open reading frames (uORFs). Moreover, DENR depletion shortens circadian period in mouse fibroblasts, suggesting involvement of uORF usage and reinitiation in clock regulation. To identify DENR-regulated translation events transcriptome-wide and, in particular, specific core clock transcripts affected by this mechanism, we have used ribosome profiling in DENR-deficient NIH3T3

cells. We uncovered 240 transcripts with altered translation rate, and used linear regression analysis to extract uORF features predictive of DENR dependence. Among core clock genes, we identified *Clock* as a DENR target. Using *Clock* 5' UTR mutants, we mapped the specific uORF through which DENR acts to regulate CLOCK protein biosynthesis. Notably, these experiments revealed an alternative downstream start codon, likely representing the bona fide CLOCK N-terminus. Our findings provide insights into uORF-mediated translational regulation that can regulate the mammalian circadian clock and gene expression at large.

Introduction

Initiation is considered the rate-limiting step for eukaryotic translation (Jackson et al., 2010) and can be regulated by multiple mechanisms, including through upstream open reading frames (uORFs), which are short translated sequences defined by a start codon within the 5' untranslated region (5' UTR) of the transcript and an in-frame stop codon. Canonical translation initiation involves association of the small ribosomal subunit with an mRNA's 5' cap structure and scanning of the 5' UTR until a start codon is

recognised; in cases when this translation event is on a uORF, the ribosome can resume scanning after termination, allowing reinitiation at the downstream protein coding sequence (CDS). Alternatively, a scanning ribosome might skip certain uORFs, in a process known as leaky scanning. Translation reinitiation and leaky scanning are poorly understood mechanisms that may involve canonical and non-canonical initiation factors and that are likely regulated by specific features of the uORF and within the transcript (reviewed in (Barbosa et al., 2013; Wethmar, 2014)).

The heterodimeric non-canonical intiation factor consisting of Density regulated protein (DENR) and Malignant T-cell amplified sequence 1 (MCTS1) has been described to bind to the small ribosomal subunit and to promote the recruitment of the initiator tRNA, Met-tRNA$_i^{Met}$, to viral mRNAs in an eIF2-independent fashion (Skabkin et al., 2010). In *Drosophila*, DENR-MCTS1 was identified as a reinitiation factor required for efficient translation on transcripts containing short and strong uORFs (Schleich et al., 2014). Furthermore, using the degree of DENR regulation experimentally determined for synthetic reporters that contained uORFs of varying lengths and Kozak strengths, endogenous DENR targets have recently been predicted in human cells (Schleich et al., 2017). However, an unbiased identification of DENR

targets *in vivo* and the direct, transcriptome-wide quantification of the effects exerted by DENR on translational regulation are still lacking.

uORF usage is pervasive in the mammalian transcriptome (with up to 40% of transcripts predicted to contain at least one uORF (Calvo et al., 2009)), and generally leads to reduced translation efficiency (TE) on the main CDS due to the inefficiency of reinitiation and leaky scanning (Wethmar, 2014). We have previously reported that several transcripts encoding components of the circadian clock machinery contained translated uORFs (Janich et al., 2015). Moreover, DENR loss-of-function - as a means of globally interfering with uORF regulation - shortened the period of free-running circadian oscillations in mouse fibroblasts by ∼1.5-hours (Janich et al., 2015). However, it had remained elusive on which circadian clock transcripts and through which uORFs DENR was able to act.

Here, we have used ribosome profiling in mouse NIH3T3 fibroblasts to detect and quantify DENR-mediated translational regulation transcriptome-wide, and in particular within core clock genes. We have used comprehensive regression analysis to characterise uORF features associated with translational regulation exerted by DENR, and present experimental validation of targets arising from our translatome-wide approaches. Finally, we show that

DENR is necessary for efficient CLOCK biosynthesis from an unannotated CDS after strong uORF translation.

Results

Ribosome occupancy is increased on 5' UTRs upon *Denr* knockdown

We hypothesized that if DENR is required for efficient protein production from the main CDS of mRNAs that undergo reinitiation after uORF translation, we would be able to identify DENR-regulated transcripts by comparing ribosome occupancies transcriptome-wide in DENR loss-of-function vs. control cells by ribosome profiling. Conceivably, DENR target mRNAs should exhibit decreased translation efficiencies (TEs) on their CDS when DENR is absent (Fig. 1A). We therefore performed shRNA-mediated knockdown of *Denr* in mouse NIH3T3 fibroblasts, a cell line in which we had previously observed that *Denr* depletion led to a shortening of the period of circadian oscillations (Janich et al., 2015). Cells were lentivirally transduced with two different *Denr*-targeting shRNAs (shRNA1, shRNA2; in triplicate) or control RNAs (scramble shRNA, in triplicate; Gfp shRNA in duplicate), grown

to confluency and harvested in the presence of cycloheximide (CHX) to stabilize elongating ribosomes. After preparation and sequencing of ribosome-protected mRNA fragment (RPF-seq) and total RNA libraries (RNA-seq) as previously described (Janich et al., 2015), we obtained ribosome footprints that were predominantly 29-30 nt in length (Supplementary Fig. 1A), that were enriched for protein coding sequences (Supplementary Fig. 1B) and that showed the characteristic triplet periodicity expected from high-quality footprint data (Supplementary Fig. 1C). Across replicates, CDS-mapping normalised reads were highly correlated ($R^2 > 0.97$, Supplementary Fig. 1D), indicating good reproducibility. The RNA-seq and RPF-seq data also confirmed the efficiency of *Denr* knockdown, showing a 77-90% reduction in *Denr* transcript levels, whereas the abundance of *Mcts1* mRNA, encoding the DENR heterodimerization partner, or of *Eif2d*, encoding a structurally and functionally related homologue of the DENR-MCTS1 dimer (Skabkin et al., 2010), remained unchanged (Fig. 1B,C and Supplementary Fig. 1E).

As a first assessment of whether reduced *Denr* levels affected ribosome occupancies on the CDS and on uORFs – as would be expected if reinitiation efficiencies were altered – we calculated relative 5' UTR ribosome occupancies ("relative 5' UTR translation") as the ratio of 5' UTR-to-CDS-mapping

footprint reads for *Denr*-depleted and control cells in a transcriptome-wide fashion. Quantification of footprints from the full 5' UTR thus served as a proxy for uORF translation, consistent with the idea that translated uORFs would be the major source of 5' UTR-mapping reads, and analogous to published methods (Sendoel et al., 2017; Blair et al., 2017). Of note, 5148 genes, corresponding to 48.9% of all expressed genes and presumably enriched for translated uORFs (see later), had sufficient coverage on their 5' UTRs to be quantifiable in this analysis. For this gene set, we observed a significant increase in relative translation on 5' UTRs in *Denr*-deficient cells (Fig. 1D), reflecting a redistribution of translating ribosomes from CDS to 5' UTRs that was specific for the footprint data and absent from the RNA-seq data (Fig. 1E). Nevertheless, on pairwise comparisons, less than 500 genes showed a translational shift that was statistically significant (Wilcoxon signed-rank test, FDR < 0.1, n $= 493$, Supplementary Fig. 2A), consistent with the idea that DENR-MCTS1 regulate protein synthesis rates on a specific subset of transcripts rather than globally (Schleich et al., 2014).

In addition to the DENR-dependent change in footprint densities on 5' UTRs vs. CDS, the redistribution also showed a distinct pattern within the 5' UTR, where it was more pronounced on the distal than on the proximal

part of the UTR (Fig. 1F). Conceivably, these changes on the translational landscape may reflect complex alterations in the extent of leaky scanning and uORF usage, and are in line with a role of DENR in translation reinitiation (Schleich et al., 2014).

DENR targets show decreased CDS translation efficiency, are enriched for uORFs, and are associated with functional pathways

From the shift in the distribution of 5' UTR vs. CDS footprints that was affecting the transcriptome very broadly in *Denr*-deficient cells (Fig. 1D), we wished to narrow down to DENR targets that were potentially functionally relevant. We reasoned that changes in CDS translation efficiency (TE; defined as the ratio of CDS-mapping RPF-seq to RNA-seq reads) would represent an appropriate measure to identify such targets, as TE differences would alter the amount of protein produced from the transcript. We therefore calculated transcriptome-wide TEs and, given that there is currently no established consensus on the most suitable analytical tool for differential translation analysis, we then applied several available algorithms to detect

144 significant TE changes. Briefly, the four methods we employed, Xtail (Xiao
145 et al., 2016), RiboDiff (Zhong et al., 2017), Riborex (Li et al., 2017), and
146 Babel (Olshen et al., 2013), are all conceived for the quantification of trans-
147 lational changes that deviate from possible variations already occurring at
148 the level of mRNAs abundances. They differ in parameter estimation and
149 modeling approaches, as well as in sensitivity, false discovery rates and com-
150 putational speed. When applied using settings that maximize comparability
151 of results (see Materials and Methods), we identified 240, 78, 41 and 30 genes
152 that had a significant change in TE between *Denr*-deficient and control cells
153 with Xtail, RiboDiff, Riborex, and Babel, respectively (FDR < 0.1 for all
154 four algorithms). Importantly, differential TE genes detected by Riborex,
155 RiboDiff and Xtail completely overlapped, while only about half of the genes
156 detected by Babel were also found by any of the other methods (Supplemen-
157 tary Fig. 3A). Moreover, most genes detected by Xtail, RiboDiff and Riborex
158 showed decreased TE in *Denr*-depleted cells, as expected for the knockdown
159 of a reinitiation factor, and they spanned the full range of expression levels
160 (Fig. 2A, and Supplementary Fig. 3B). By contrast, Babel was biased to-
161 wards low abundance transcripts and identified similar numbers of up- and
162 down-regulated TE genes (Supplementary Fig. 3B). From these results, we

concluded that Xtail likely represented the most suitable method to identify DENR targets, and further estimated the false discovery rate of the algorithm by randomly shuffling the dataset labels, which eliminated nearly all differential TE genes (Supplementary Fig. 3C).

Among the 240 differential TE genes identified by Xtail (see Supplementary file 1) were several for which the presence of regulatory uORFs has previously been described, such as *Atf5* (Watatani et al., 2008), *Rnaseh1* (Suzuki et al., 2010; Liang et al., 2016), *Jak2* and *Map2k6* (Schulz et al., 2018). In order to investigate whether uORF translation was generally associated with TE decrease among DENR-regulated genes, we annotated uORFs transcriptome wide. Briefly, we included AUG and non-canonical, near-AUG (i.e., CUG, GUG and UUG) initiation codons located in 5' UTRs, and assessed the active translation of the corresponding ORFs by their footprint coverage, frame preference and triplet codon periodicity (Supplementary Fig. 4A). Using this approach, we detected translated uORFs in nearly half of all expressed genes (n = 4755 genes, Fig. 2B), consistent with previous predictions (Calvo et al., 2009). For the majority of genes, the uORFs were fully contained in the 5' UTRs (51.8%, n = 2464 genes), but genes with uORFs overlapping the CDS (19.5%, n = 926 genes), or that had both overlapping

182 and non-overlapping uORFs (28.7%, n = 1365 genes) were frequent, too (Sup-
183 plementary Fig. 4B). Notably, DENR targets were highly enriched for genes
184 with translated uORFs (72.7%, p = 5.73e-4, Fisher test) and, in particular,
185 with non-overlapping uORFs (Supplementary Fig. 4B). The latter obser-
186 vation is consistent with the idea that after terminating on an overlapping
187 uORF, a ribosome would presumably not be able to backtrack to reinitiate
188 on an upstream CDS start codon, explaining why overlapping uORFs ap-
189 pear insensitive to DENR regulation. Moreover, relative 5' UTR translation
190 in *Denr*-depleted vs. control cells calculated over the full 5' UTR – as a
191 proxy of uORF translation – was significantly higher for DENR targets than
192 for the global set of uORF-containing genes (Supplementary Fig. 4C,D). In
193 summary, these observations were fully consistent with an involvement of
194 DENR in reinitiation and in efficient CDS translation of a subset of trans-
195 lated uORF-containing mRNAs.

196 We next analysed whether the 240 differential TE genes showed functional
197 enrichment for specific pathways. GO term analysis revealed categories as-
198 sociated with translation and mRNA stability, as well as cellular functions
199 related to protein phosphorylation and catabolism, kinase binding and cell
200 cycle (Fig. 2C and Supplementary Fig. 4E), in line with previous observa-

tions in Drosophila S2 cells (Schleich et al., 2014) and indicating evolutionary conservation of DENR functions.

DENR-regulated genes included cancer-relevant genes, such as *Klhdc8a*, *Etaa1*, *Map2k5*, *Slc20a1* or *Vegfd* (Mukasa et al., 2010; Childs et al., 2015; Diao et al., 2016; Sato and Akimoto, 2017; Thelen et al., 2008), and individual inspection confirmed reduced CDS translation efficiency (Fig. 3A), as well as a shift of ribosome occupancy from the CDS to the 5' UTR in *Denr*-depleted cells (Fig. 3B). This signature was characteristic of many differential TE genes (see Supplementary Fig. 5A, B), and correlated with the presence of translated uORFs (Supplementary Fig. 5C).

In order to validate the DENR dependence of CDS translation on 5' UTR elements for selected identified targets, we cloned the 5' UTR and immediate start codon context of the transcript of interest in frame and upstream of *Firefly* luciferase (FL) in a lentiviral vector that also expresses *Renilla* luciferase (RL, for internal normalisation) from the same bidirectional promoter (Fig. 3C, top). *Denr* depletion led to a 18-44% reduction in FL/RL signals for five different control reporters that contained either no cloned 5' UTR, or 5' UTRs of genes that were not among the differential TE genes (of these, *Cry1* and *Fth1* harbours translated uORFs as well, whereas *Deaf1*, and

Rpl30 do not) (Fig. 3C, bottom). By contrast, we observed a much stronger reduction in FL/RL signals – by 49 - 90% – for four reporters containing differential TE gene UTRs (*Etaa1, Klhdc8a, Map2k5 and Lrrc28*), indicating that DENR was indeed involved in their regulation. Furthermore, mutation of the uORF start codon increased reporter signal and partially abolished DENR dependence (Fig. 3C, bottom).

Taken together, we concluded that DENR regulates protein synthesis rates of a set of $\sim$ 200 murine transcripts containing uORFs and involved in translation, RNA stability and kinase activity, by promoting uORF bypassing and/or efficient reinitiation.

Linear regression analysis correlates DENR-dependence with uORF number, start codon identity, and intercistronic distance

uORFs have been proposed to act through a bewildering diversity of mechanisms, including through effects on CDS translation (by affecting start site selection/leaky scanning/re-initiation/ribosome shunting), but also on transcript abundance (by eliciting nonsense-mediated mRNA decay), and in some

cases even by producing bioactive small peptides (reviewed in (Wethmar, 2014)). However, rules that would predict the specific mechanism from specific uORF features, have not yet been established. Our datasets could potentially allow identifying which signatures render certain uORF-containing transcripts DENR-dependent, as opposed to the majority of uORF-containing transcripts whose TE was not affected by DENR. Previous studies have suggested that, at least in the context of synthetic uORF reporters in Drosophila S2 (Schleich et al., 2014) and human HeLa cells (Schleich et al., 2017), very short uORFs (coding for only 1-2 amino acids) with favourable Kozak contexts around the start codon were most strongly dependent on DENR for efficient reinitiation of a downstream reporter CDS.

In order to determine whether DENR-dependent translation events were associated with specific transcript features in a comprehensive and unbiased fashion, we used a multiple regression analysis approach that modeled the CDS TE change observed upon *Denr* knockdown, considering nine features that could potentially regulate uORF translation and efficient translation reinitiation on the CDS. These features included the number of uORFs, the uORF start codon identity (AUG vs. near-cognate) and context, the uORF length and GC content, the relative location within the 5' UTR, the presence

of overlapping uORFs, and the CDS start codon context (Fig. 4A). The model was applied to all translated uORF-containing genes (n = 4755), and yielded several features with significant predictive value for regulation by DENR (Fig. 4B-J, adjusted R2 = 0.057, F-statistic = 33.21 on 4731 DF, p < 2.2e-16). In particular, the number of uORFs present on a transcript was significantly associated with TE regulation (Fig. 4B, b = -0.019, p < 2e-16), with transcripts containing several uORFs being increasingly DENR-dependent, suggesting an additive effect of the presence of multiple uORFs for the need of efficient reinitiation. On the other hand, the presence of CDS-overlapping uORFs was not associated with TE fold change upon *Denr* depletion (Fig. 4C, b = 0.005, p = 0.44), as expected for a reinitiation factor, given that overlapping uORFs would be reinitiation-incompetent unless the transcripts contained a downstream alternative CDS start codon.

A second feature predictive of TE regulation by DENR was the uORF start codon. Briefly, to account for the presence of multiple uORFs with different start codons, we computed the proportion of AUG uORFs on each transcript. Using this metric to quantify the effect of canonical (AUG) vs. non-canonical (non-AUG) uORF initiation codons, the model revealed that AUG uORF-containing transcripts showed the strongest CDS TE reduction

upon *Denr* knockdown (Fig. 4D, b = -0.05, p = 1.15e-9). Moreover, DENR targets were enriched for AUG uORFs as compared to the ensemble of translated uORF-containing transcripts (p = 0.0044, Fisher test, Supplementary Fig. 6A), and genes harboring only AUG uORFs showed the strongest effect on CDS TE in *Denr*-deficient cells (Supplementary Fig. 6B). Along the same lines, and as shown before (Schleich et al., 2014, 2017), stronger Kozak contexts around uORF start codons were associated with TE regulation as well (Fig. 4E, b = -0.0026, p = 0.013). By contrast, the Kozak context around the main ORF start codon bore no predictive value (Fig. 4F, b = -0.0006, p = 0.38). Taken together, these findings indicated that DENR was more likely to affect CDS translation when the uORFs had a favourable context for efficient translation. This is fully compatible with a role of DENR in reinitiating after termination at strong uORFs; nevertheless, we wish to point out that these observations are formally also compatible with an alternative (or additional) role in controlling the fidelity of uORF start codon selection. Of note, selectivity of translation initiation factors for uORF start codons has been previously reported; for example, eIF2A is involved in CUG uORF translation (Starck et al., 2016; Sendoel et al., 2017), whereas eIF1 overexpression promotes AUG initiation (Ivanov et al., 2010) and eIF1 knockdown

reduces AUG uORF translation in favour of non-canonical ones (Fijalkowska et al., 2017). The increased ribosome occupancy of 5' UTRs (Fig. 1d, f) and the association with AUG-initiated uORFs revealed by our model, may thus indicate that DENR, contrary to eIF1, acts to bypass AUG uORFs on its targets to ensure efficient CDS protein biosynthesis.

Already early studies using a model transcript had suggested that the location of a uORF with respect to the main ORF was able to dictate the efficiency of reinitiation, as longer spacing would generally allow the reacquisition of the necessary initiation factors to the scanning ribosome (Kozak, 1987). Our model provides an unbiased and transcriptome-wide validation of this notion for DENR-regulated transcripts, where larger distances from the 3'-most uORF stop codon to the main ORF start were associated with downregulation of TE (Fig. 4G, b = -0.0101, p = 7.89e-7). This finding indicated that uORFs located further upstream from the CDS were more likely to reinitiate in a DENR-dependent fashion. Importantly, this effect of uORF-CDS distance was not caused by longer 5' UTRs *per se*, given that the distance from the 5' cap to the first uORF was not associated with DENR regulation (Fig. 4H, b = -0.0018, p = 0.18). Moreover, the distance between the CDS and its closest uORF was only weakly correlated with 5' UTR length

(spearman rho = 0.16), and the correlation of the distance between the 5' cap and the first uORF was even stronger (spearman rho = 0.27, Supplementary Fig.6C, D).

No linear relationship was found between uORF length and CDS TE (Fig. 4I, b = 0.0006, p = 0.81), in agreement with previous results in human cells using synthetic uORFs encoding > 2 amino acids (9 nt) (Schleich et al., 2017). Of note, a length dependence for extremely short uORFs of 1-2 amino acids, as reported by Schleich et al. 2017 (Schleich et al., 2017), would go unnoticed in our analyses, as we excluded them due to the uncertainties associated with their annotation (see Materials and Methods).

Finally, the uORF GC content showed a significant relationship with TE fold change (Fig. 4J, b = 0.058, p = 0.005), suggesting that potential weaker secondary structures (low GC) favoured reinitiation involving DENR.

For the above model, we used the CDS TE fold change as response, which has the benefit of accounting for DENR regulation in a continuous manner. Of note, however, also a logistic regression model based on a binary response (i.e. DENR target vs. non-target) recapitulates the main predictions for the number of uORFs (b = 0.015, p = 1.01e-9), for the proportion of AUG uORFs (b= 0.02, p = 0.046) and for the spacing between the 3'-most uORF and the

CDS (b = 0.005, p = 0.034) (Supplementary Fig. 6E, whole model statistics: adjusted R2 = 0.0117, F-statistic = 7.278 on 4732 DF, p = 1.381e-10), further supporting our predictions for the strongest cases of DENR translational control.

In summary, our results showed that DENR acts to regulate efficient protein synthesis from transcripts containing AUG uORFs and those in strong Kozak contexts, and that the distance between the most downstream uORF and the CDS start is an important parameter for efficient reinitiation.

DENR-mediated reinitiation regulates CLOCK protein biosynthesis

We recently found that uORF translation was pervasive within circadian core clock transcripts, and that the depletion of *Denr* in NIH3T3 cells engendered robust circadian period shortening by 1-1.5 hours (Janich et al., 2015). We therefore sought to uncover which core clock transcripts were DENR-regulated, and to map the specific uORFs through which reinitiation occurred. Conceivably, these would represent likely contributors to the observed circadian phenotype.

The expression of several core clock genes was altered in *Denr*-depleted vs.

control cells. In most cases (*Per1/2/3*, *Cry1/2*, *Nr1d1*, *Ror*α) the observed changes occurred already at the level of mRNA abundance and correspondingly led to comparable changes at the footprint level as well (Fig. 5A, and Supplementary Fig. 7A). None of these transcripts (nor *Bmal1*, whose lower RNA abundance was partially compensated at the footprint level) was associated with marked TE reduction (Fig. 5B, and Supplementary Fig. 7B), and we thus considered them unlikely candidates for direct DENR targets. For *Clock*, however, the reduction in expression was mostly translational, resulting in a ∼20% decrease in protein biosynthesis, as quantified from *Clock* RPFs (Figs. 5A, B and Supplementary Figs. 7A, B). In our initial analysis of differential TE, *Clock* had, however, not passed the significance threshold (Fig. 2A and Supplementary File 1). The reason was that the applied algorithm, Xtail, evaluates statistical significance of TE differences through two separate tests – one regarding differences between the RNA and the RPF fold-change distributions, and a second one evaluating differences in TE distributions directly – and then derives the final p-value from the most conservative of the two tests (Xiao et al., 2016). While the additional test might prevent false positives, it can discard differential TE genes close to the statistical threshold. Importantly, for one of the tests, the TE fold change for

369 *Clock* was significant (p = 0.025 in test 1; p = 0.19 for test 2; Supplementary
370 File 1); *Clock* thus appears to represent a DENR target whose detection is
371 at the borderline of statistical significance.

372 We annotated two translated AUG-initiated uORFs in *Clock* (Fig. 5C
373 and Supplementary Fig.7C, D). The second uORF overlapped with the main
374 CDS in a different frame – a configuration that would render the two trans-
375 lation events (uORF-2 and annotated CDS) mutually exclusive. However,
376 downstream of the uORF-2 stop codon and 10 amino acids into the anno-
377 tated *Clock* CDS, we noted an additional AUG in the *Clock* coding frame;
378 we thus hypothesised that, when uORF-2 was translated, initiation at this
379 alternative AUG could still allow for the biosynthesis of a CLOCK protein
380 variant that was N-terminally truncated by 9 amino acids (Fig. 5C and
381 Supplementary Fig. 7E). In *Denr*-deficient cells, translation was increased
382 on both uORFs (Fig. 5C) and, in particular, uORF-2 showed ∼23% higher
383 translation rate upon *Denr* depletion (Fig. 5D), suggesting that DENR-
384 mediated leaky scanning of, or reinitiation after, uORF-2 was crucial for the
385 translation of *Clock*. We looked for further evidence of uORF-2 and alter-
386 native CDS translation using GWIPS-viz, an online genome browser for the
387 visualisation of ribosome profiling data from published studies (Michel et al.,

388 2014). Interestingly, initiating ribosomes can be detected on both uORF-2
389 and alternative CDS start codons (but not the annotated *Clock* start codon)
390 across human and mouse datasets, and elongating ribosome profiling studies
391 showed good footprint coverage on uORF-2 and the characteristic accumu-
392 lation of ribosomes at the start and stop codons (Supplementary Fig. 8).
393 Together, our and published data thus suggested abundant translation of
394 uORF-2 and of *Clock* from the alternative AUG, indicating that reinitiation
395 has an important role in ensuring CLOCK protein biosynthesis.

396 In order to investigate uORF usage, effect, and DENR-dependence for
397 CLOCK biosynthesis from the annotated and alternative CDS, we cloned
398 the *Clock* 5' UTR and first 14 codons (containing both the annotated and
399 alternative AUGs) in frame and upstream of *Firefly* luciferase in our dual
400 luciferase vector (Fig. 5F, top). We then systematically mutated the uORF-
401 1, uORF-2, and CDS annotated/alternative start codons, or combinations
402 thereof (Supplementary Fig. 7F), and compared the effects on luciferase
403 output in *Denr*-depleted vs. control cells. Of note – and in contrast to
404 Fig. 3C – we also adjusted for the effect of *Denr* depletion on the control
405 reporter, whose levels were set to 100, to obtain the isolated effects of the
406 *Clock* variants.

407 *Denr* depletion resulted in an almost 30% reduction in expression for the
408 *Clock Wt* construct (Fig. 5F, bottom), in line with the decrease in rela-
409 tive translation quantified *in vivo* (Fig. 5D), and also with reduced levels
410 of endogenous CLOCK protein (Supplementary Fig. 9A). Surprisingly, mu-
411 tation of the annotated CDS initiation codon did not substantially affect
412 CLOCK synthesis or DENR dependence when compared to *Clock Wt*. How-
413 ever, mutation of the alternative AUG decreased luciferase signal by ∼95%
414 and abolished DENR dependence (Fig. 5F, bottom), suggesting that most
415 CLOCK biosynthesis in NIH3T3 cells involves the alternative AUG. Muta-
416 tion of uORF-1 led to a 19% increase in luciferase signal and fully retained
417 DENR dependence. By contrast, uORF-2 mutation led to a ∼50% increase in
418 protein synthesis and eliminated DENR regulation. These findings indicated
419 that uORF-2 translation acts to significantly reduce CLOCK biosynthesis
420 and to render it DENR-dependent. This notion was further supported by
421 the combinatorial mutations. Mutation of both uORF-1 + 2 start codons
422 thus resulted in an additive, >90% increase in protein synthesis, and in a
423 loss of DENR-dependence. In line with minor levels of synthesis of the long
424 CLOCK isoform, mutation of the annotated CDS start codon in combination
425 with uORF mutations led to similar results as mutation of the uORFs alone.

426 Furthermore, mutation of the alternative AUG in combination with uORF
427 mutations led to strong decrease in luciferase signals, consistent with the
428 predominant synthesis of the short CLOCK isoform. In particular, the case
429 where both the alternative CDS and uORF-2 (or uORF-1+2) start codons
430 were mutated – a scenario that would allow unhindered CLOCK synthesis
431 from the annotated AUG – resulted in a >80% reduction in CLOCK protein
432 production (Fig. 5F, bottom). Taken together, we concluded that *Clock*
433 required DENR for efficient protein synthesis from a CDS that is 9 codons
434 shorter than the annotated one, and that uORF-2 plays a key role in co-
435 ordinating the complex interactions that occur at the translational level on
436 the *Clock* transcript (Fig. 5G). Moreover, we propose that the CLOCK N-
437 terminus is likely not only wrongly annotated in NIH3T3 cells, but based on
438 our analysis of published ribosome profiling data, also in other cell types and
439 in tissues (Supplementary Fig. 8).

440 Discussion

441 For a long time, the annotation of uORFs remained a particular challenge,
442 in part because they are short, frequently initiate from non-AUG codons,

and do not obey the same rules of evolutionary conservation as most protein coding sequences (Wethmar, 2014). This has changed with the availability of ribosome profiling data, which has made it possible to systematically identify translated uORFs from footprints. The resulting burst in interest in uORFs and the rising appreciation of how prevalent their usage is in mammals, is, however, not yet matched by an understanding at the mechanistic level. In this study, we have investigated one particularly intriguing mechanism that is implicated in the regulation through uORFs, namely translation reinitiation. As an entry point, we have used the loss-of-function of DENR, a non-canonical initiation factor recently associated with reinitiation both in *Drosophila* (Schleich et al., 2014) and in HeLa cells(Schleich et al., 2017), with the aim of identifying rules that explain why certain uORFs will trigger DENR-dependent reinitiation while other uORFs do not. The conclusions we were able to draw from experiments in mouse NIH3T3 fibroblasts showed similarities and striking differences to the previous studies in fly and in human cells (Schleich et al., 2014, 2017). Species-specificity in uORF biology may explain some of the disparities, but differences in experimental and computational approaches likely have an important influence as well.

We view our quantitative modeling approach of DENR-dependent changes

in CDS TE as particularly unbiased as it relies on translational changes detected on endogenous mRNAs transcriptome-wide. Using all uORF-containing transcripts annotated in our analyses, we have identified four prominent features with predictive value for the magnitude of DENR-mediated translation efficiency regulation on the CDS. Three of the features – i.e, number of uORFs, start codon identity, start codon sequence context – may reflect a straightforward logic between a stronger requirement for reinitiation when the usage of uORFs is efficient. To this end, our study thus independently validates (and, notably, links to DENR) the idea that the presence of several uORFs is related to the modulation of CDS TE through successive reinitiation events (Calvo et al., 2009). Moreover, it is in line with the human (Schleich et al., 2017) and fly data (Schleich et al., 2014) that showed the strongest DENR dependence for AUG-initiated uORFs and strong Kozak contexts, suggesting mechanistic conservation.

Two other findings of the model are noteworthy. First, the fourth predictive feature – the uORF-to-CDS distance – i.e. the notion that uORFs located further away from the main ORF would be more prone to trigger DENR-dependent reinitiation. Second, the absence of predictive value of uORF length. It should be noted that intercistronic distance effects were re-

ported early on in the uORF field, typically using distinct uORF-containing model transcripts (Kozak, 1987, 2001; Luukkonen et al., 1995), and they had generally been assumed to reflect the kinetics of association and dissociation of initiation and termination factors. Our data validate this notion for the distance between uORF and CDS, supporting in a transcriptome-wide fashion the hypothesis that longer distances allow for extended re-scanning times, which would be favourable for the loss of termination factors and/or the re-acquisition of the initiation machinery, on time for CDS translation. By contrast, uORF length did not figure among the predictors in our model. As a caveat, the nature of our approach (transcriptome-wide from endogenous transcripts) did not allow to confidently annotate uORFs down to only 1-2 amino acids. We may thus be missing DENR effects that have recently been observed in reporters carrying such ultra-short synthetic uORFs in human cells (Schleich et al., 2017) (see Results). However, at least for uORFs regulated by DENR, the hypothesis that increased uORF length impairs reinitiation efficiency, does not appear valid. There may still be a role for the time required for uORF translation in determining scanning resumption, as linear length *per se* may simply not be a good proxy for translation time (Poyry et al., 2004; Calvo et al., 2009; Barbosa et al., 2013).

Overall, our model explains a relatively modest proportion of the total variance (adjusted R^2 of 0.057). This is reminiscent of analogous models that have been developed to predict, for example, miRNA target sites, whose adjusted R^2s range from 5-13%, but that have nevertheless proven biologically useful (Agarwal et al., 2015). In our case, the presence of multiple uORFs on a transcript, often overlapping to each other, and each with its own, distinct features (start codon, length, distance to the CDS, etc), and the resulting combinatorial effects, clearly complicated their scoring and likely masked some of the origins of translational regulation. We deliberately wished to keep our model simple, aiming at identifying the most significant predictors, but at the expense of overall explanatory power. In the future, the use of other regression types and models that include interaction terms could potentially explain higher biological variability, although at the cost of becoming considerably more difficult to interpret. Overall, our study constitutes a first attempt to analyse signatures associated with reinitiation through DENR, and it could already recapitulate previous observations (Schleich et al., 2014, 2017); as we continue to identify more molecular features that can regulate non-canonical translation initiation and that can be integrated into the model in the future, a larger proportion of biological variance will be explained.

We have identified 240 transcripts as likely direct targets of DENR, many of which are involved in proliferation and associated with cancer. Of note, DENR-MCTS1 have themselves been identified as oncogenes (Reinert et al., 2006) that are upregulated, for example, in prostate cancer (Cerami et al., 2012). Moreover, translational reprogramming, especially with regard to upstream initiation sites, appears to represent an important cancer driver (Sendoel et al., 2017), and understanding how the translation of cancer-relevant DENR targets is regulated may thus lead to important insights of diagnostic or therapeutic value. We found, for example, very strong DENR requirement for the translation of Kelch Domain Containing 8A (KLHDC8A). It has been reported that KLHDC8A overexpression in human gliomas that become resistant to epidermal growth factor receptor (EGFR) silencing (a commonly used anticancer therapeutic approach), allows tumours to maintain aggressiveness through an unknown EGFR-independent pathway (Mukasa et al., 2010). It is conceivable that upon loss of EGFR sensitivity, KLHDC8A levels increase through DENR-mediated translational control. Similarly, other candidates from our DENR target list (Supplementary file 1) are upregulated in (and possibly causally contributing to) different cancers, such as Vascular endothelial growth factor D (VEGF-D), or Mitogen-Activated Protein Ki-

538 nase 5 (MAP2K5) (Thelen et al., 2008; Diao et al., 2016). By understanding
539 how DENR regulates its targets, it may thus be possible to reduce its activity
540 and thus the oncogenic levels of the modulated proteins in proliferating, but
541 not in quiescent cells.

542 How do DENR and circadian rhythms mechanistically connect? The
543 DENR target list did not reveal core clock genes that would explain the cir-
544 cadian clock phenotype, i.e. the ∼1.5-hour period shortening of free-running
545 oscillations that we previously reported for *Denr*-deficient cells (Janich et al.,
546 2015). Individual inspection of core clock gene transcripts, however, sug-
547 gested that *Clock* was a DENR target just below the significance threshold.
548 Follow-up experiments confirmed regulation *Clock* regulation through DENR
549 and allowed mapping the responsible uORF. These findings establish uORF
550 translation and re-initiation as novel, potentially regulated mechanisms act-
551 ing on core clock gene expression, and they open several intriguing questions.
552 First, is it possible that reduced biosynthesis of *Clock* would be responsible
553 for the observed phenotype? Loss-of-function of components of the gene
554 expression machinery typically go along with sizeable distortions in global
555 RNA and protein profiles, and *Denr* knockdown is no exception in this re-
556 spect. Associated phenotypes are thus likely to represent composite effects,

557 rather than the result of single misregulated gene expression events. We
558 therefore do not view reduced CLOCK biosynthesis as a satisfactory hypoth-
559 esis that alone would be able to explain the substantial period shortening.
560 Nevertheless, it is reassuring that *Clock*$^{-}/^{-}$ fibroblasts have recently been
561 shown to have shorter periods as well (Landgraf et al., 2016), which sug-
562 gests that downregulation of *Clock* translation upon *Denr* depletion may be
563 a contributing factor to the phenotype.

564 A second intriguing question relates to the alternative initiation site that
565 would generate an N-terminally truncated variant of CLOCK. We propose
566 that the 9 amino acid shorter CLOCK is the main variant produced in
567 NIH3T3 cells, and likely other cell types and species – so do such CLOCK iso-
568 forms have function implications? The bHLH domain responsible for DNA
569 binding of CLOCK starts at amino acid 26, and the very N-terminus has
570 not yet been associated with any function of the protein, and it is also not
571 included in the crystal structure of the CLOCK-BMAL1 complex (Huang
572 et al., 2012). However, it is noteworthy that for several myogenic bHLH
573 transcription factors, N-terminal motifs are critical for cooperative DNA
574 binding with cofactors (Knoepfler et al., 1999). Much of our current knowl-
575 edge on CLOCK stems from *in vivo* and *in vitro* experiments that have used

576 *Clock* cDNA constructs devoid of the endogenous uORF regulation, and thus
577 defining CLOCK's N-terminus to the annotated one. It will be exciting to
578 re-evaluate selected DNA-binding and activity assays using both CLOCK
579 isoforms side by side.

580 Finally, a speculative question to ask is whether DENR-mediated control
581 is a regulated, rather than constitutive step in CLOCK biosynthesis, serving
582 for example to uncouple CLOCK protein abundance from mRNA levels under
583 certain physiological conditions. DENR was originally identified as a protein
584 whose expression increases in cultured cells at high density, but not during
585 growth arrest *per se*, which led to its name Density regulated protein (Deyo
586 et al., 1998). First links between cell density and circadian rhythmicity have
587 indeed been described in recent years (Noguchi et al., 2013; Guenthner et al.,
588 2014), and it would be interesting to assess whether CLOCK uORFs and
589 DENR play roles in density-dependent differences of clock parameters.

590 Methods

591 **Cell culture.** NIH3T3 and HEK293FT cells were cultured under standard
592 conditions (DMEM; 10% FCS, 1% penicillin/streptomycin, all from Invitro-

593 gen; 37C; 5% CO2).

594 Lentiviral particle production in HEK293FT cells using envelope pMD2.G
595 and packaging psPAX2 plasmids, and viral transduction of NIH3T3 cells,
596 were performed following published protocols (Salmon and Trono, 2007),
597 with puromycin selection at 5 μg/ml for 4 days.

598 **Cloning and plasmids.** For the generation of lentiviral shRNA expression
599 vectors, two different sequences targeting *Denr* were cloned into pLKO.1puro
600 backbone vector (Addgene no. 10878 (Moffat et al., 2006)): shRNA1: GTAC-
601 CACAGAAGGTCACGATA, corresponding to clone TRCN0000308443 of
602 the TRC shRNA Library from the Broad Institute; and shRNA2: GT-
603 GCCAAGTTAGATGCGGATT, corresponding to clone TRCN0000098826.
604 pLKO.1puro constructs containing Gfp, and Scramble (Addgene no. 1864)
605 shRNAs served as controls.

606 For the generation of dual luciferase (Firefly/Renilla) reporter plasmids,
607 fragments containing the 5'UTR and first 5-14 codons of the selected DENR
608 targets were amplified by PCR from genomic DNA or cDNA, cloned into the
609 BamHI site of the prLV1 dual luciferase reporter plasmid (Du et al., 2014),
610 and validated by sequencing. The following primers were used for PCR:

Etaa1, forward: aaaggatccGACTTGCAAAGATGGCGCTGCAC, reverse: tttggatccactagtGTCCTTCAGCTGCATTCACATTT; Map2k5, forward: aaaggatccGGTTCCGGAGTAACAGCGGTCTAAC, reverse: tttggatccactagtGGCCAGCCACAGCATTACAGGTTAA; Lrrc28, forward: aaaggatccGCCGCTGAGTGCCGGTCAGCGGGC, reverse: tttggatccactagtGATTTCGGATGCCATGACTGAACA; Klhdc8a, forward: aaaggatccGTCTCCGACCCTGTAGACACTGCAG, reverse: tttggatccactagtATTGGGCACTTCCATGGCAGCCCGG; Clock, forward: aaaggatccGGGGAGGAGCGCGGCGGTAGCGGTG, reverse: tttggatccactagtAACAATTGAGCTCATTTTACTACAGCTTACGG.

5'UTR mutants were generated by site-directed mutagenesis of the corresponding plasmids, using the following primers: Etaa1-Mut, forward: cgaaggatcCGACTTGCAAAGgtaccGCTGCACGGTGGCGCGCGGGCTC, reverse: GAGCCCGCGCGCCACCGTGCAGCggtacCTTTGCAAGTCGgatccttcg; Map2k5-Mut, forward: CTCTTACTCACAAGGACTACAGgctagcTTGTGGGTTCTGTTTTGTCCAC, reverse: GTGGACAAAACAGAACCCACAAgctagcCTGTAGTCCTTGTGAGTAAGAG; Klhdc8a-Mut, forward: GGCAGGGAGCCCGTGCAGCGCGgtaccGAGGCTGAGAGAGGGGACGCGCC, reverse: GGCGCGTCCCCTCTCTCAGCCTCggtacCG

CGCTGCACGGGCTCCCTGCC; Clock-uORF1-Mut, forward: CCCAAA ATCACCAGCAAGAGTTgctagcGTCAGTCACACAGAAGACGGCC, reverse: GGCCGTCTTCTGTGTGACTGACgctagcAACTCTTGCTGGTG ATTTTGGG; Clock-uORF2-Mut, forward: AAAGTGAAAGAGGAGAAG TACAggtacCTACCACAAGACGAAAACATAAT, reverse: ATTATGTT TTCGTCTTGTGGTAGgtaccTGTACTTCTCCTCTTTCACTTT; Clock-AnnotCDS-Mut, forward: CAAGACGAAAACATAATGTGTTtcGGTGT TTACCGTAAGCTGTAG, reverse: CTACAGCTTACGGTAAACACCga AACACATTATGTTTTCGTCTTG; Clock-AlterCDS-Mut, forward: GTT TACCGTAAGCTGTAGTAAAcaGAGCTCAATTGTTactagtg, reverse: c actagtAACAATTGAGCTCtgTTTACTACAGCTTACGGTAAAC.

Dual luciferase assay NIH3T3 cells were first lentivirally transduced with the dual luciferase reporters containing the 5'UTR construct of interest and then transduced with Denr-shRNA/scramble-shRNA. After puromycin selection, cells were lysed in passive lysis buffer and luciferase activities were quantified using DualGlo luciferase assay system and a GloMax 96 Microplate luminometer (all from Promega). Firefly/Renilla (F/R) luciferase activities were first normalized to that of the empty vector in scramble shRNA-

transduced cells. For *Denr* target experiments, F/R signals were also normalized to the scramble-transduced, non-mutated construct. For *Clock* mutational experiments, F/R signals were secondly normalized to the scramble-transduced *Clock*-Wt construct.

Western blotting. Total protein extracts from NIH3T3 cells were prepared according to the NUN procedure (Lavery and Schibler, 1993). SDS-PAGE and immunoblot were performed according to standard protocols, using the NuPage 4-12% Bis-Tris gels (Invitrogen) and iBlot2 transfer system (Invitrogen). Antibodies used were rabbit anti-CLOCK 1:3000 (ab3517), mouse anti-U2AF65 1:5000 (Sigma), rabbit anti-DENR 1:4000 (ab108221), and anti-rabbit-HRP or anti-mouse-HRP 1:10000 (Promega). Blots were visualised with the Fusion Fx7 system and quantified with ImageJ 1.48k.

Ribosome profiling. One 15-cm dish of confluent NIH3T3 cells was used per replicate. Ribosome profiling and parallel RNA-seq were performed in triplicate for *Denr* shRNA1, *Denr* shRNA2 and scr shRNA, and in duplicate for *Gfp* shRNA, following the ART-Seq protocol by Epicentre-Illumina (Cat.no. RPHMR12126) with minor modifications as described below. Cells were treated with cycloheximide (100 μg/ml) for 2 min at 37C to arrest trans-

lation elongation, and immediately harvested, followed by lysate preparation. For the generation of ribosome-protected fragments, cell lysates were treated with 5 units RNase I (ART-Seq) per OD260 and monosomes were purified in MicroSpin S-400 columns (GE Healthcare). For both RPF and RNA-seq, 5 μg of RNA were used for rRNA depletion following the Ribo-Zero Magnetic Kit intructions from Epicentre-Illumina (Cat.no. MRZH11124). Finally, libraries were amplified using 12 PCR cycles, and were sequenced on a HiSeq2500 platform.

Sequencing pre-processing, alignment and quantification. Initial processing, mapping and quantification of mRNA and footprint abundance were performed as previously (Castelo-Szekely et al., 2017). Briefly, adaptor sequence was trimmed with Cutadapt (Martin, 2011) and trimmed sequences were filtered using a custom python script to keep 26-35 nt and 21-60 nt reads for RPF and RNA-seq, respectively. Reads were then sequentially aligned to mouse rRNA, tRNA and cDNA (Ensembl mouse release 84) using Bowtie2 (Langmead and Salzberg, 2012) and mouse genome using Tophat v2.0.11 (Trapnell et al., 2009). Trimmed and filtered sequences from RNA-seq were also directly mapped against the mouse genome with Tophat to estimate

expressed transcript isoforms (using Cufflinks v2.2.1 (Trapnell et al., 2010)). Transcriptome-mapping reads from the sequential alignment were counted towards their location on 5'UTR, CDS or 3'UTR per gene, using only the set of expressed isoforms as estimated above. The location of the putative A-site of the footprint was used for RPF-seq counting (position +15 from the 5' for reads $\leq$30 nt and position +16 for reads >31 nt), and the 5'end was used for RNA-seq reads. For genes with multiple expressed isoforms, reads that did not map unambiguously to a single feature were assigned to one in preference order: CDS, 5'UTR, 3'UTR.

RPF and RNA-seq read counts were normalized separately with the upper quartile normalization from edgeR (Robinson and Oshlack, 2010) and then transformed to RPKM values as read count per gene kilobase per normalized library size. Genes with less than 10 reads in 2/3 of samples and mitochondrially-enconded genes were removed before normalization.

Translation efficiencies were then calculated per sample as the ratio of RPF RPKM to RNA RPKM of the region of interest (5'UTR, CDS or uORF), log2 transformed and averaged over all replicates.

Differential translation efficiency analysis. Significant changes in TE between knockdown and control cells were assessed with several available algoithms: Xtail (Xiao et al., 2016) with options 'bins=10000', RiboDiff (Zhong et al., 2017) with options '-d 1 -p 1', Riborex (Li et al., 2017) with options 'engine="DESeq2"', and Babel (Olshen et al., 2013) with options 'nreps=100000'.

Briefly, all four methods, with the options implemented in our study for maximal consistency among them, model raw read counts by following a negative binomial distribution, and use global normalization methods for sequencing depth that would not account for global changes in TE if exisiting. RiboDiff and Riborex use DESeq2 to estimate a single mean and dispersion for both RPF- and RNA-seq, and detect differentially translated genes by fitting RPF- and RNA-seq read counts to a single GLM, whereas Xtail uses DESeq2 for mean and dispersion estimation separately for RPF- and RNA-seq and fits two independent GLMs. Xtail then derives probability distributions for fold changes of mRNA or RPF, or for RPF-to-mRNA ratios, and evaluates statistical significance between conditions from one of the two (the most conservative) pipelines. Babel relies on edgeR for parameter estimation and uses an errors-in-variables regression to test for significant

720 changes in translation within and between conditions.

721 Differential TE was set at FDR < 0.1 for all four methods and perfor-
722 mance for false positive discovery rate was further assessed by randomly
723 label-shuffling knockdown and control datasets, which should not yield dif-
724 ferential TE genes.

725 **uORF annotation and translation analyses.** For uORF detection, a
726 set of transcripts was used for which reads can be unambiguously assigned
727 to the 5' UTR, i.e. transcripts that are the only protein-coding isoform
728 expressed (n = 4637). In order to include as many transcripts as possible
729 for the annotation, genes with multiple protein-coding isoforms were also
730 considered if 1) one isoform represented $> 65\%$ of all expressed (as derived
731 from our transcript expression estimation, see "Sequencing pre-processing,
732 alignment and quantification", n=3211) or 2) all expressed isoforms had the
733 same CDS start, in which case the transcript with the highest expression was
734 selected (n=1826). For the selected transcripts, uORFs were annotated and
735 considered as translated with the following criteria: 1) started with AUG,
736 CUG, GUG or UUG, 2) had an in-frame stop codon within the 5' UTR
737 or overlapping/within the CDS, 3) were at least 9 nt long (including stop

codon), 4) had a coverage > 25%, 5) showed a frame preference and 6) the preferred frame was the one corresponding to the uORF 5' end/start codon. Importantly, we considered only uORFs coding for at least 2 amino acids (i.e. 9 nt, including the stop codon), as for shorter uORFs frame bias assessment would be highly prone to false positives, and Kozak context scoring would be compromised (for 6 nt uORFs, the position +4 would always be the T of the stop codon, and thus cannot be a G of the strongest Kozak context).

Multiple regression analysis. For the modelling, all transcripts containing translated uORFs in *Denr*-knockdown or control cells were included (n = 4755) and the following uORF features were used as predictors of the fold change in CDS TE across conditions: 1) the number of uORFs, 2) the uORFs length, using the median length in transcripts with several uORFs, 3) the proportion of AUG uORFs, 4) the uORF Kozak context, scored relative to the consensus GccA/Gcc###G, where upper-case letters denote highly conserved bases (scored +3), lower-case letters indicate the most common nucleotides (scored +1) and ### represents the start codon (not scored). The median Kozak score was used when several uORFs were present on a transcript. 5) the CDS Kozak context, scored in the same manner, 6) the

distance of the first uORF start codon from the 5' cap, 7) the distance of the last uORF stop codon to the CDS start, 8) the uORF GC content, using the median for transcripts with several uORFs, 9) the presence of overlapping uORFs in the transcript (binary predictor). Predictors with skewed distributions (number of uORFs, uORF length, distance from the cap and distance to the CDS start) were log2 transformed. Partial regression plots were generated with the R package 'visreg'. Absence of multicollinearity between predictors was assessed with the variance inflation factor using R function vif (car package).

Functional enrichment analysis. Functional enrichment analysis was performed using g:Profiler (Reimand et al., 2016), including Gene Ontology Biological Process and Molecular Function terms and biological pathway information from Reactome. Functional categories of more than 2 and less than 1500 members were used and p-values were corrected by Benjamini-Hochberg FDR. Categories with FDR < 0.05 were considered as significant and were used for visualization with Cytoscape v.3.5.1 EnrichmentMap tool (Merico et al., 2010).

Data availability Ribosome profiling data was deposited to GEO, under accession number: GSE116221. Code for ribosome profiling pre-processsing, aligning and quantification, and for uORF annotation, is available (Castelo-Szekely, 2017).

Acknowledgements

We thank the Lausanne Genomics Technologies Facility and Vital-IT for high-throughput sequencing and computing infrastructure. We thank Angelica Liechti and Georgia Katsioudi who were involved in initial steps of reporter plasmid clonings. We are indebted to funding through: Swiss National Science Foundation grants 157528, 179190; National Centre of Competence in Research (NCCR) RNA & Disease; Olga Mayenfisch Stiftung; SystemsX.ch StoNets consortium; Foundation Herbette; University of Lausanne.

Author contributions

V.C.S. designed experiments, performed ribosome profiling, reporter assays and western blotting, analysed the data, and wrote the manuscript; M.D.M. performed molecular cloning experiments; D.G. designed experiments, su-

pervised the study, and wrote the manuscript.

Competing financial interests

The authors declare no competing interests.

References

Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4:e05005.

Barbosa, C., Peixeiro, I., and Romao, L. (2013). Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLOS Genetics*, 9(8):e1003529.

Blair, J. D., Hockemeyer, D., Doudna, J. A., Bateup, H. S., and Floor, S. N. (2017). Widespread Translational Remodeling during Human Neuronal Differentiation. *Cell Reports*, 21(7):2005–2016.

Calvo, S. E., Pagliarini, D. J., and Mootha, V. K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are

polymorphic among humans. *Proceedings of the National Academy of Sciences*, 106(18):7507–7512.

Castelo-Szekely, V. (2017). gatfieldlab/cross-organ_riboprof cross-organ_riboprof_v1 zenodo. https://zenodo.org/record/521199#.wzyyeby-mcg.

Castelo-Szekely, V., Arpat, A. B., Janich, P., and Gatfield, D. (2017). Translational contributions to tissue specificity in rhythmic and constitutive gene expression. 18:116.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*, 2(5):401–404.

Childs, E. J., Mocci, E., Campa, D., Bracci, P. M., Gallinger, S., Goggins, M., Li, D., Neale, R. E., Olson, S. H., Scelo, G., Amundadottir, L. T., Bamlet, W. R., Bijlsma, M. F., Blackford, A., Borges, M., Brennan, P., Brenner, H., Bueno-de Mesquita, H. B., Canzian, F., Capurso, G., Cavestro, G. M., Chaffee, K. G., Chanock, S. J., Cleary, S. P., Cotterchio,

821 M., Foretova, L., Fuchs, C., Funel, N., Gazouli, M., Hassan, M., Herman,
822 J. M., Holcatova, I., Holly, E. A., Hoover, R. N., Hung, R. J., Janout,
823 V., Key, T. J., Kupcinskas, J., Kurtz, R. C., Landi, S., Lu, L., Malecka-
824 Panas, E., Mambrini, A., Mohelnikova-Duchonova, B., Neoptolemos, J. P.,
825 Oberg, A. L., Orlow, I., Pasquali, C., Pezzilli, R., Rizzato, C., Saldia, A.,
826 Scarpa, A., Stolzenberg-Solomon, R. Z., Strobel, O., Tavano, F., Vashist,
827 Y. K., Vodicka, P., Wolpin, B. M., Yu, H., Petersen, G. M., Risch, H. A.,
828 and Klein, A. P. (2015). Common variation at 2p13.3, 3q29, 7p13 and
829 17q25.1 associated with susceptibility to pancreatic cancer. *Nature Genet-*
830 *ics*, 47(8):911–916.

831 Deyo, J. E., Chiao, P. J., and Tainsky, M. A. (1998). drp, a novel protein
832 expressed at high cell density but not during growth arrest. *DNA and cell*
833 *biology*, 17(5):437–447.

834 Diao, D., Wang, L., Wan, J., Chen, Z., Peng, J., Liu, H., Chen, X., Wang,
835 W., and Zou, L. (2016). MEK5 overexpression is associated with the
836 occurrence and development of colorectal cancer. *BMC Cancer*, 16:302.

837 Du, N.-H., Arpat, A. B., De Matos, M., and Gatfield, D. (2014). MicroRNAs

shape circadian hepatic gene expression on a transcriptome-wide scale. *eLife*, 3:e02510.

Fijalkowska, D., Verbruggen, S., Ndah, E., Jonckheere, V., Menschaert, G., and Van Damme, P. (2017). eIF1 modulates the recognition of suboptimal translation initiation sites and steers gene expression via uORFs. *Nucleic Acids Research*, 45(13):7997–8013.

Guenthner, C. J., Luitje, M. E., Pyle, L. A., Molyneux, P. C., Yu, J. K., Li, A. S., Leise, T. L., and Harrington, M. E. (2014). Circadian rhythms of Per2::Luc in individual primary mouse hepatocytes and cultures. *PloS One*, 9(2):e87573.

Huang, N., Chelliah, Y., Shan, Y., Taylor, C. A., Yoo, S.-H., Partch, C., Green, C. B., Zhang, H., and Takahashi, J. S. (2012). Crystal Structure of the Heterodimeric CLOCK:BMAL1 Transcriptional Activator Complex. *Science*, 337(6091):189–194.

Ivanov, I. P., Loughran, G., Sachs, M. S., and Atkins, J. F. (2010). Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proceedings of the National Academy of Sciences of the United States of America*, 107(42):18056–18060.

Jackson, R. J., Hellen, C. U. T., and Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2):113–127.

Janich, P., Arpat, A. B., Castelo-Szekely, V., Lopes, M., and Gatfield, D. (2015). Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. *Genome Research*, 25(12):1848–1859.

Knoepfler, P. S., Bergstrom, D. A., Uetsuki, T., Dac-Korytko, I., Sun, Y. H., Wright, W. E., Tapscott, S. J., and Kamps, M. P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. *Nucleic Acids Research*, 27(18):3752–3761.

Kozak, M. (1987). Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Molecular and Cellular Biology*, 7(10):3438–3445.

Kozak, M. (2001). Constraints on reinitiation of translation in mammals. *Nucleic Acids Research*, 29(24):5226–5232.

Landgraf, D., Wang, L. L., Diemer, T., and Welsh, D. K. (2016). NPAS2

Compensates for Loss of CLOCK in Peripheral Circadian Oscillators. *PLOS Genetics*, 12(2):e1005882.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359.

Lavery, D. J. and Schibler, U. (1993). Circadian transcription of the cholesterol 7 alpha hydroxylase gene may involve the liver-enriched bZIP protein DBP. *Genes & Development*, 7(10):1871–1884.

Li, W., Wang, W., Uren, P. J., Penalva, L. O. F., and Smith, A. D. (2017). Riborex: fast and flexible identification of differential translation from Ribo-seq data. *Bioinformatics*, 33(11):1735–1737.

Liang, X.-h., Shen, W., Sun, H., Migawa, M. T., Vickers, T. A., and Crooke, S. T. (2016). Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nature Biotechnology*, 34(8):875–880.

Luukkonen, B. G., Tan, W., and Schwartz, S. (1995). Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *Journal of Virology*, 69(7):4086–4094.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12.

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G. D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS One*, 5(11):e13984.

Michel, A. M., Fox, G., M. Kiran, A., De Bo, C., O'Connor, P. B. F., Heaphy, S. M., Mullan, J. P. A., Donohue, C. A., Higgins, D. G., and Baranov, P. V. (2014). GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Research*, 42(D1):D859–D864.

Moffat, J., Grueneberg, D. A., Yang, X., Kim, S. Y., Kloepfer, A. M., Hinkle, G., Piqani, B., Eisenhaure, T. M., Luo, B., Grenier, J. K., Carpenter, A. E., Foo, S. Y., Stewart, S. A., Stockwell, B. R., Hacohen, N., Hahn, W. C., Lander, E. S., Sabatini, D. M., and Root, D. E. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, 124(6):1283–1298.

Mukasa, A., Wykosky, J., Ligon, K. L., Chin, L., Cavenee, W. K., and Furnari, F. (2010). Mutant EGFR is required for maintenance of glioma

growth in vivo, and its ablation leads to escape from receptor dependence. *Proceedings of the National Academy of Sciences*, 107(6):2616–2621.

Noguchi, T., Wang, L. L., and Welsh, D. K. (2013). Fibroblast PER2 circadian rhythmicity depends on cell density. *Journal of Biological Rhythms*, 28(3):183–192.

Olshen, A. B., Hsieh, A. C., Stumpf, C. R., Olshen, R. A., Ruggero, D., and Taylor, B. S. (2013). Assessing gene-level translational control from ribosome profiling. *Bioinformatics*, 29(23):2995–3002.

Poyry, T. A. A., Kaminski, A., and Jackson, R. J. (2004). What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes & Development*, 18(1):62–75.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research*, 44(W1):W83–89.

Reinert, L. S., Shi, B., Nandi, S., Mazan-Mamczarz, K., Vitolo, M., Bachman, K. E., He, H., and Gartenhaus, R. B. (2006). MCT-1 Protein Interacts with the Cap Complex and Modulates Messenger RNA Translational Profiles. *Cancer Research*, 66(18):8994–9001.

Robinson, M. D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3):R25.

Salmon, P. and Trono, D. (2007). Production and titration of lentiviral vectors. *Current Protocols in Human Genetics*, Chapter 12:Unit 12.10.

Sato, K. and Akimoto, K. (2017). Expression Levels of KMT2c and SLC20a1 Identified by Information-theoretical Analysis Are Powerful Prognostic Biomarkers in Estrogen Receptor-positive Breast Cancer. *Clinical Breast Cancer*, 17(3):e135–e142.

Schleich, S., Acevedo, J. M., Hohenberg, K. C. v., and Teleman, A. A. (2017). Identification of transcripts with short stuORFs as targets for DENR-MCTS1-dependent translation in human cells. *Scientific Reports*, 7(1):3722.

Schleich, S., Strassburger, K., Janiesch, P. C., Koledachkina, T., Miller, K. K., Haneke, K., Cheng, Y.-S., Kuchler, K., Stoecklin, G., Duncan, K. E., and Teleman, A. A. (2014). DENR-MCT-1 promotes translation re-initiation downstream of uORFs to control tissue growth. *Nature*, 512(7513):208–212.

Schulz, J., Mah, N., Neuenschwander, M., Kischka, T., Ratei, R., Schlag, P. M., Castanos-Velez, E., Fichtner, I., Tunn, P.-U., Denkert, C., Klaas, O., Berdel, W. E., Kries, J. P., Makalowski, W., Andrade-Navarro, M. A., Leutz, A., and Wethmar, K. (2018). Loss-of-function uORF mutations in human malignancies. *Scientific Reports*, 8(1):2395.

Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., Levorse, J., Dill, B. D., Schramek, D., Molina, H., Weissman, J. S., and Fuchs, E. (2017). Translation from unconventional 5' start sites drives tumour initiation. *Nature*, 541(7638):494.

Skabkin, M. A., Skabkina, O. V., Dhote, V., Komar, A. A., Hellen, C. U. T., and Pestova, T. V. (2010). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes & Development*, 24(16):1787–1801.

Starck, S. R., Tsai, J. C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., Martins-Green, M., Shastri, N., and Walter, P. (2016). Translation from the 5' untranslated region shapes the integrated stress response. *Science*, 351(6272):3867.

Suzuki, Y., Holmes, J. B., Cerritelli, S. M., Sakhuja, K., Minczuk, M., Holt,

963 I. J., and Crouch, R. J. (2010). An Upstream Open Reading Frame and the
964 Context of the Two AUG Codons Affect the Abundance of Mitochondrial
965 and Nuclear RNase H1. *Molecular and Cellular Biology*, 30(21):5123–5134.

966 Thelen, A., Scholz, A., Benckert, C., von Marschall, Z., Schroder, M.,
967 Wiedenmann, B., Neuhaus, P., Rosewicz, S., and Jonas, S. (2008). VEGF-
968 D promotes tumor growth and lymphatic spread in a mouse model of hepa-
969 tocellular carcinoma. *International Journal of Cancer*, 122(11):2471–2481.

970 Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering
971 splice junctions with RNA-Seq. *Bioinformatics*, 25(9):1105–1111.

972 Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van
973 Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Tran-
974 script assembly and quantification by RNA-Seq reveals unannotated tran-
975 scripts and isoform switching during cell differentiation. *Nature Biotech-*
976 *nology*, 28(5):511–515.

977 Watatani, Y., Ichikawa, K., Nakanishi, N., Fujimoto, M., Takeda, H.,
978 Kimura, N., Hirose, H., Takahashi, S., and Takahashi, Y. (2008). Stress-
979 induced Translation of ATF5 mRNA Is Regulated by the 5'-Untranslated
980 Region. *Journal of Biological Chemistry*, 283(5):2543–2553.

Wethmar, K. (2014). The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdisciplinary Reviews: RNA*, 5(6):765–768.

Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nature Communications*, 7:11194.

Zhong, Y., Karaletsos, T., Drewe, P., Sreedharan, V. T., Kuo, D., Singh, K., Wendel, H.-G., and Ratsch, G. (2017). RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics*, 33(1):139–141.

Figure legends

Figure 1. *Denr* knockdown engenders ribosome redistribution from CDS to 5' UTR. a) Experimental approach of the study. Ribosome profiling (RPF-seq) and parallel RNA-seq were performed on NIH3T3 mouse fibroblasts depleted for *Denr* via lentivirally expressed shRNAs, with scramble (Scr) and *Gpf* shRNAs as control conditions. Profiling was carried out in triplicate for the two *Denr* shRNAs (shRNA1, shRNAs), and for Scr

shRNA, and in duplicate for *Gfp* shRNA. We hypothesized that if DENR is required for efficient reinitiation after uORF translation, knocking down *Denr* would lead to reduced translation efficiency on the main coding sequence of DENR targets. **b)** Distribution of normalized RNA-seq read counts along the *Denr* transcript show high *Denr* knockdown efficiency. Different orange color shadings correspond to the different replicates. **c)** RPKM values for CDS-mapping RNA-seq read counts for *Denr* (left), *Mcts1* (middle), and *Eif2d* (right) indicate specific and high knockdown efficiency. Each dot corresponds to a sample and replicate and error bars indicate mean±sd. **d)** Histogram of 5' UTR-to-CDS translation (from RPF-seq) between *Denr* Kd and control cells. The analysis was limited to genes with a minimum of 20 RPF-seq reads in the 5' UTR in 3/4 of samples (n = 5148), to exclude noisy gene expression. Relative 5' UTR translation was significantly higher in *Denr* Kd cells (median = 0.32, $p < 2.2e\text{-}16$, Wilcoxon signed rank test), suggesting a ribosome footprint redistribution to 5' UTRs upon *Denr* knockdown. **e)** Histogram of 5' UTR-to-CDS mRNA abundance ratios (from RNA-seq) between *Denr* Kd and control cells, using same genes as in D. In contrast to translation, relative 5' UTR transcript levels did not change between conditions (median = 0.03, p = 0.09). **f)** Metagene analysis of normalized RPF-seq read counts across

5' UTRs. 5' UTRs were scaled to forty windows of equal size and RPF signal was averaged within each window and across genes. *Denr*-deficient cells showed an accumulation of ribosome footprints on the 5' UTR (in particular within the downstream part of the UTR), possibly due to the presence of translated uORFs.

Figure 2. Characteristics of DENR targets: decreased CDS TE, enrichment for uORFs, and functional categories. **a)** Scatterplot of *Denr* Kd-to-control ratio of mRNA abundance vs. CDS translation efficiency (TE) for all expressed genes (n = 10516). Highlighted in orange are genes with differential TE (detected by Xtail (Xiao et al., 2016), FDR < 0.1, n=240), the majority of which showed decreased TE in *Denr* Kd cells, in line with a role of DENR in efficient translation reinitiation. **b)** uORF enrichment in DENR targets. Proportion of genes with translated uORFs transcriptome-wide compared to that among the differential TE genes detected in (a). DENR targets showed enrichment for translated uORF-containing genes (p = 5.73e-4, Fisher test). **c)** Enrichment map for GO-Biological process, GO-Metabolic function and Reactome functional categories of genes with differential TE. Only categories with an FDR < 0.05 in the GO analysis were

1035 considered. Node size is proportional to the number of genes associated with
1036 the GO category, shading corresponds to the statistical significance, and
1037 edge width is proportional to the number of genes shared between functional
1038 categories.

1039 **Figure 3. Mutation of uORF start codons relieves DENR depen-**
1040 **dence. a)** Representative examples of genes with differential TE. Panels
1041 show average CDS translation efficiency in control (grey) and *Denr* Kd (red)
1042 cells. **b)** *Denr* Kd-to-Control translation on the 5' UTR (brown) and on
1043 the CDS (grey) for the same examples as in (a). DENR targets showed
1044 decreased CDS TE (a) and increased 5' UTR translation concomitant with
1045 reduced CDS translation in *Denr*-deficient cells. **c)** *(Top)* Schematic repre-
1046 sentation of the dual luciferase reporter construct, where the 5' UTR of the
1047 gene of interest – with the Wt or mutated uORF – and first few codons of
1048 the CDS were cloned upstream of firefly luciferase (FL). *(Bottom)* Results of
1049 dual luciferase assays, where the influence of the 5' UTR (Wt, and in some
1050 cases uORF mutant) on the expression of FL was measured in control and
1051 *Denr* knockdown cells, and internally normalised to RL, expressed from the
1052 same bidirectional promoter. Note that for all targets tested, depletion of

Denr reduced the F/R luciferase signal, indicating DENR dependence of efficient reporter protein expression. Mutation of uORF start codons relieved the inhibition and partially abolished DENR-dependence. For each 5' UTR assayed, F/R signals were normalised to the Scr-transduced non-mutated construct. The area of grey shading marks the non-specific effect of *Denr* loss-of-function in this assay, as judged from the reduction in signal seen for negative control reporters that contained no cloned 5' UTR (-), or irrelevant 5' UTRs selected from non-targets.

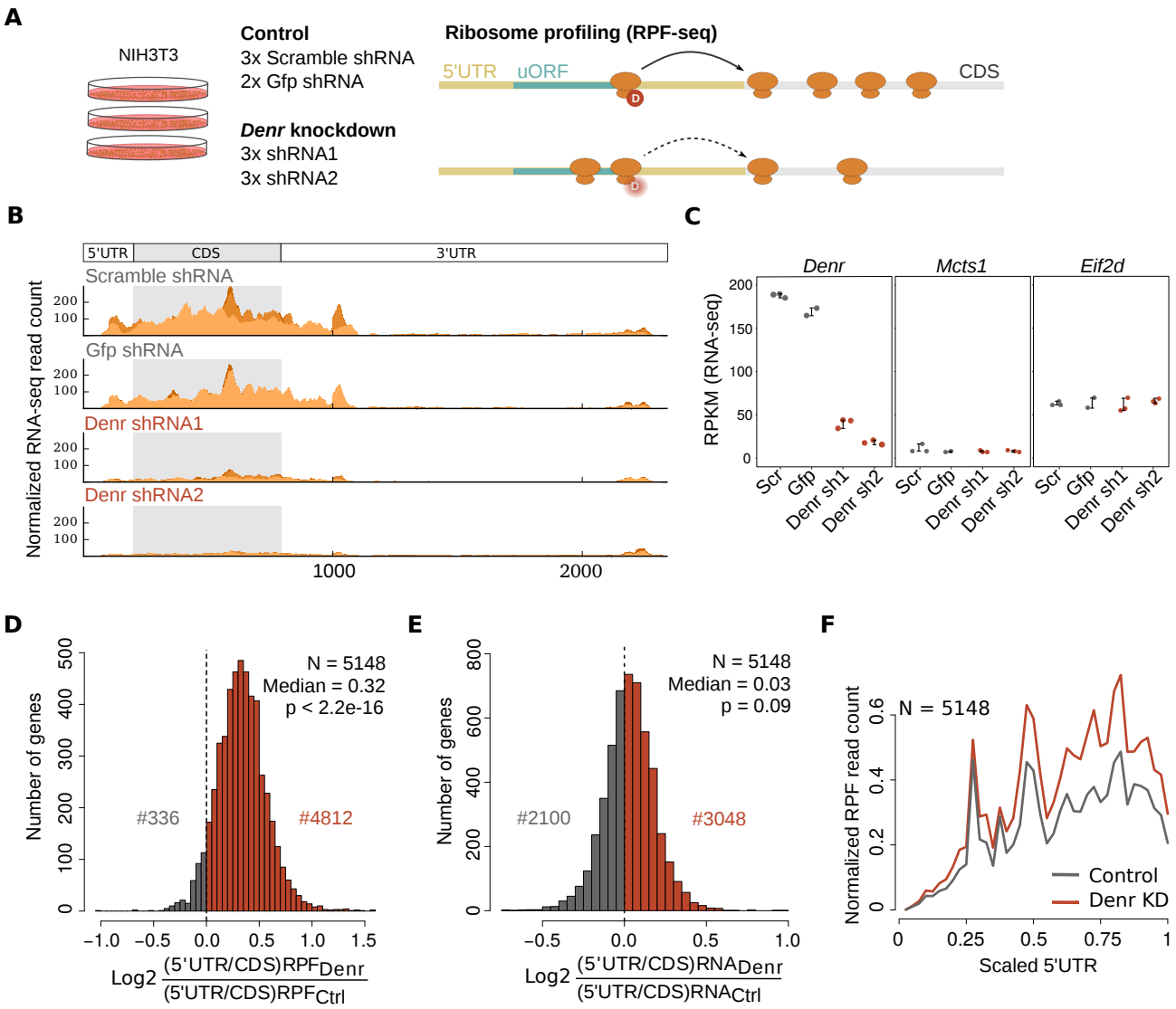
Figure 4. uORF features predictive of translation efficiency regulation by DENR. a) Schematic of the uORF features used as predictors for the modeling of CDS TE change. The number of uORFs, the start codon identity and context, the presence of overlapping uORFs, the uORF length and GC content, and the relative location within the 5' UTR were considered. Distance to the 5' cap was computed for the first uORF, and distance to the CDS start was quantified for the last uORF. **b-j)** Partial regression plots for the modeling of translation efficiency fold change (*Denr* Kd vs. Ctrl) on each uORF feature. Indicated in blue are the coefficients of the slopes for the partial regression line sand their significance levels (t-test). Note that

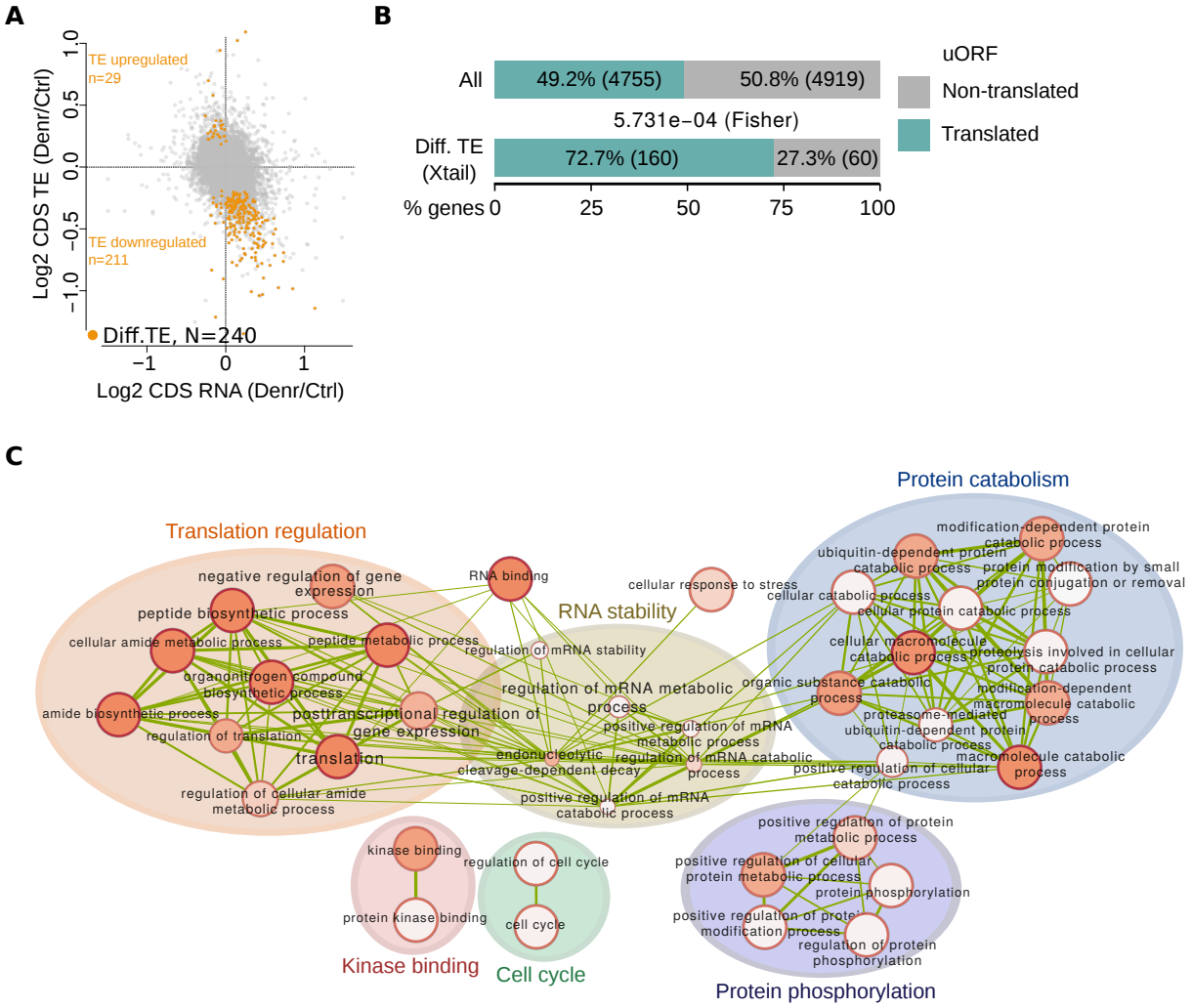
both the response and, in several cases, the predictor, are in log2 scale; thus the slopes of the partial regression models correspond to the log2 change in TE FC when doubling the corresponding explanatory variable. The number of uORFs, the proportion of AUG uORFs and the distance of the last uORF to the CDS start were the strongest predictors.

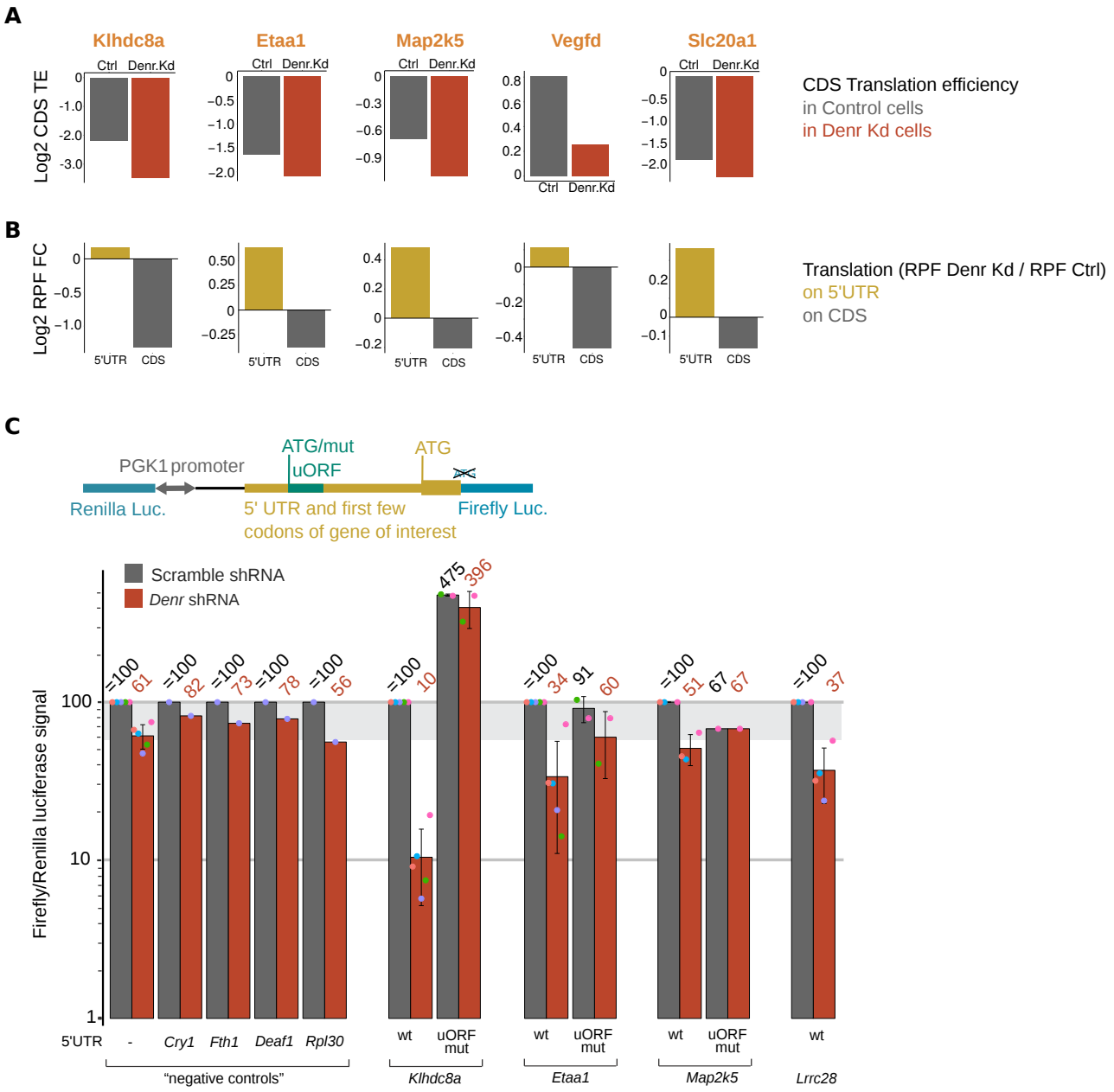
Figure 5. DENR-mediated reinitiation regulates CLOCK protein abundance and N-terminus. A) Bar graph of the average RPKM ratio (*Denr* Kd-to-control) for the main circadian core clock genes, at the level of mRNA abundance (dark shadings) and footprint abundance (light shadings). **b)** CDS translation efficiency ratio between *Denr* Kd and control cells showed that *Clock* had a reduced TE upon *Denr* depletion. **c)** Distribution of normalised RPF-seq (blue) and RNA-seq (orange) reads along the 5' UTR and first 100 nt of *Clock* CDS, in control (top) and *Denr*-Kd (bottom) cells. Green shaded areas indicate the two AUG-initiated uORFs; grey area corresponds to the annotated CDS; and brown area to the alternative CDS, in frame with the annotated one. **d)** Quantification of relative translation ratio (*Denr* Kd-to-control) on both uORFs, and on the CDS, showing increased uORF translation and reduced CDS translation in *Denr*-deficient

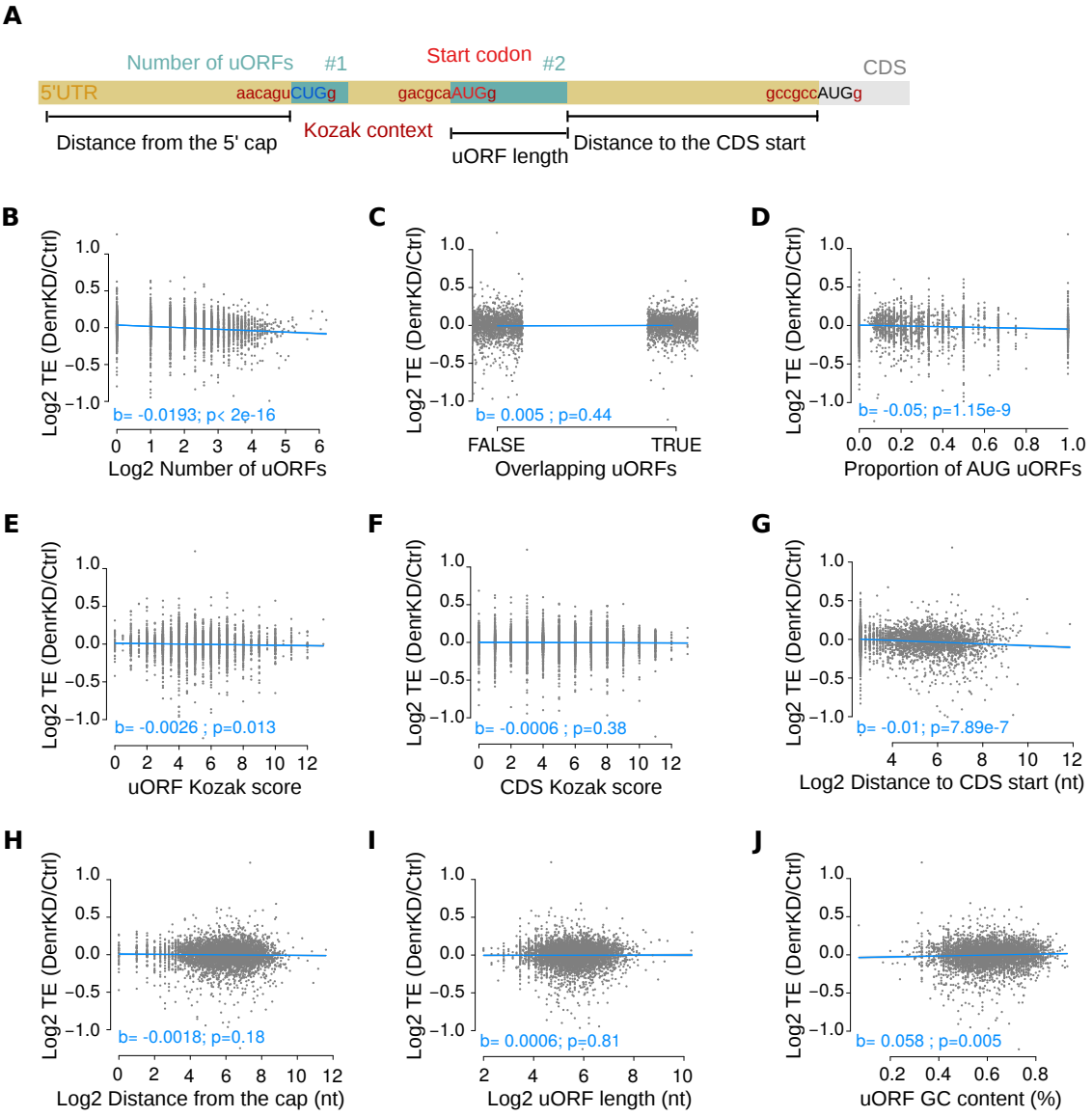
cells. **e)** Quantification of relative translation efficiency (*Denr* Kd-to-control) on both uORFs, and on the CDS showed higher TE on the overlapping uORF (uORF2). **f)** (Top) Schematic representation of the dual luciferase reporter construct, where *Clock* 5' UTR and first 14 codons of the CDS (containing both the annotated and alternative start codons) were cloned upstream of FL. (Bottom) Dual luciferase assay results for *Clock* Wt and mutants. Empty vector contained only the vector-encoded 5' UTR. FL signal was internally normalised to RL signal; all constructs and conditions were first normalised to the vaule of condition "empty vector, scr shRNA"; all *Clock* mutant constructs were then additionally normalised to the value from condition "*Clock* Wt, scr shRNA". Signal from all *Denr* shRNA-transfected *Clock* constructs were also normalised to the signal obtained for "empty vector, *Denr*-shRNA", to remove DENR effects that were not specific to the 5' UTR. **G)** Translation reinitiation on the circadian core clock component *Clock*. *Clock* transcript contains two translated AUG uORFs, the second of which shows high ribosome occupancy and overlaps with the annotated CDS. An additional in-frame AUG is located 9 amino-acids downstream of the annotated, that can give rise to a so far unannotated CLOCK isoform. Our data suggest that this second AUG is the true start codon for CLOCK

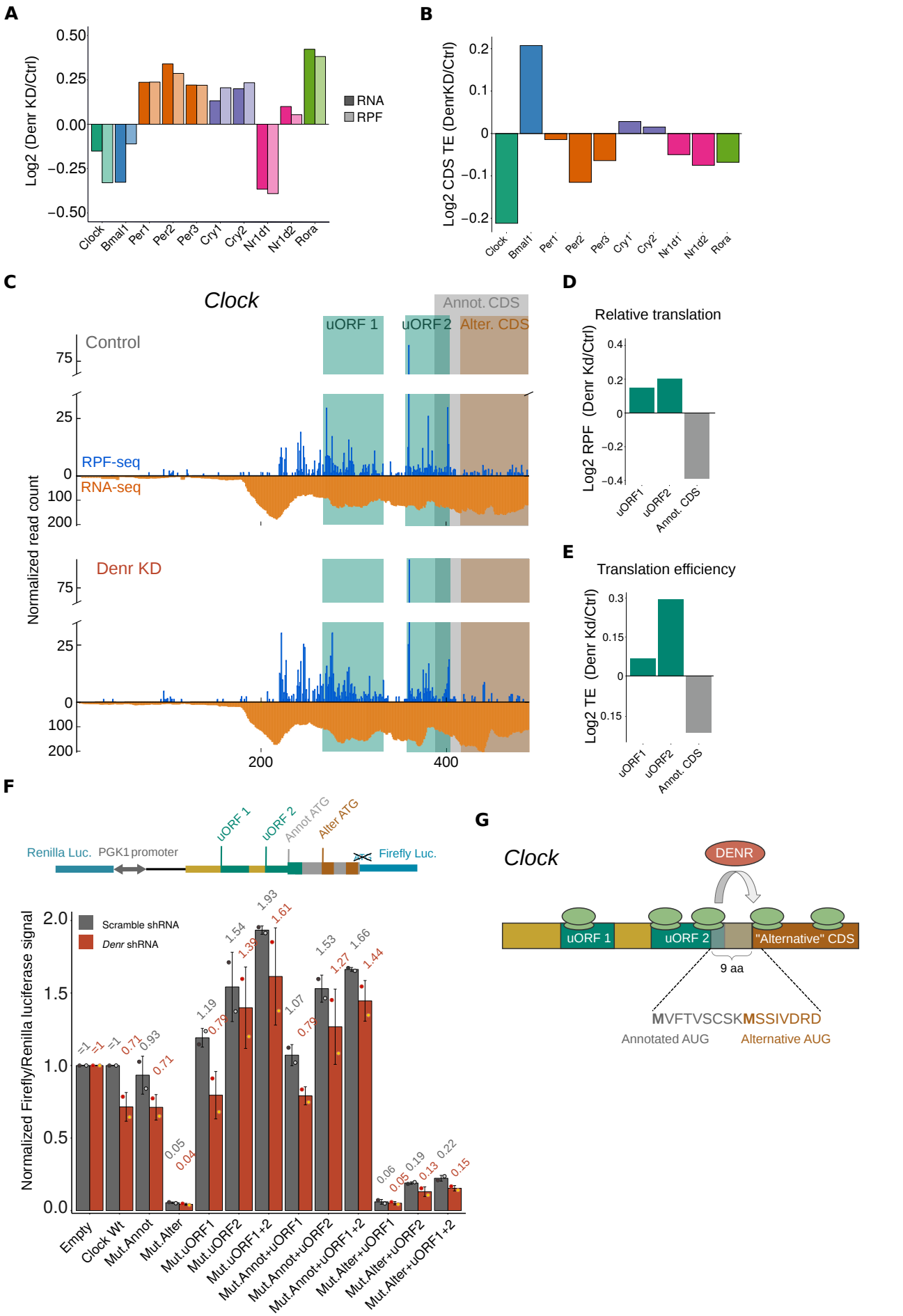
biosynthesis in NIH3T3 cells, and that DENR is necessary for translation reinitiation on the alternative CDS after translation of the second uORF.


A
NIH3T3
Control
3x Scramble shRNA
2x Gfp shRNA
Denr knockdown
3x shRNA1
3x shRNA2
Ribosome profiling (RPF-seq)
5'UTR
uORF
CDS
B
5'UTR
CDS
3'UTR
Scramble shRNA
Gfp shRNA
Denr shRNA1
Denr shRNA2
Normalized RNA-seq read count
1000
2000
C
Denr
Mcts1
Eif2d
RPKM (RNA-seq)
Scr
Gfp
Denr sh1
Denr sh2
D
N = 5148
Median = 0.32
p < 2.2e-16
#336
#4812
Number of genes
Log2 (5'UTR/CDS)RPFDenr / (5'UTR/CDS)RPFCtrl
E
N = 5148
Median = 0.03
p = 0.09
#2100
#3048
Number of genes
Log2 (5'UTR/CDS)RNADenr / (5'UTR/CDS)RNACtrl
F
N = 5148
Normalized RPF read count
Control
Denr KD
Scaled 5'UTR


A
TE upregulated
n=29
TE downregulated
n=211
Diff.TE, N=240
Log2 CDS TE (Denr/Ctrl)
Log2 CDS RNA (Denr/Ctrl)
B
All
49.2% (4755)
50.8% (4919)
5.731e−04 (Fisher)
Diff. TE (Xtail)
72.7% (160)
27.3% (60)
% genes
uORF
Non-translated
Translated
C
Translation regulation
negative regulation of gene expression
peptide biosynthetic process
cellular amide metabolic process
peptide metabolic process
organonitrogen compound biosynthetic process
amide biosynthetic process
regulation of translation
translation
regulation of cellular amide metabolic process
RNA binding
RNA stability
regulation of mRNA stability
regulation of mRNA metabolic process
cellular response to stress
positive regulation of mRNA metabolic process
endonucleolytic cleavage-dependent decay
regulation of mRNA catabolic process
positive regulation of mRNA catabolic process
Protein catabolism
modification-dependent protein catabolic process
ubiquitin-dependent protein catabolic process
protein modification by small protein conjugation or removal
cellular catabolic process
cellular protein catabolic process
cellular macromolecule catabolic process
proteolysis involved in cellular protein catabolic process
organic substance catabolic process
modification-dependent macromolecule catabolic process
proteasome-mediated ubiquitin-dependent protein catabolic process
macromolecule catabolic process
positive regulation of cellular catabolic process
Kinase binding
kinase binding
protein kinase binding
Cell cycle
regulation of cell cycle
cell cycle
Protein phosphorylation
positive regulation of protein metabolic process
positive regulation of cellular protein metabolic process
protein phosphorylation
positive regulation of protein modification process
regulation of protein phosphorylation


A
Klhdc8a
Etaa1
Map2k5
Vegfd
Slc20a1
Ctrl
Denr.Kd
Log2 CDS TE
CDS Translation efficiency
in Control cells
in Denr Kd cells
B
Log2 RPF FC
5'UTR
CDS
Translation (RPF Denr Kd / RPF Ctrl)
on 5'UTR
on CDS
C
PGK1 promoter
ATG/mut
uORF
ATG
Renilla Luc.
5' UTR and first few
codons of gene of interest
Firefly Luc.
Scramble shRNA
Denr shRNA
Firefly/Renilla luciferase signal
=100
61
82
73
78
56
10
475
396
34
91
60
51
67
67
37
5'UTR
-
Cry1
Fth1
Deaf1
Rpl30
wt
uORF mut
"negative controls"
Klhdc8a
Etaa1
Map2k5
Lrrc28


A
Number of uORFs
#1
Start codon
#2
CDS
5'UTR
aacaguCUGg
gacgcaAUGg
gccgccAUGg
Distance from the 5' cap
Kozak context
uORF length
Distance to the CDS start
B
Log2 TE (DenrKD/Ctrl)
b= -0.0193; p< 2e-16
Log2 Number of uORFs
C
b= 0.005 ; p=0.44
FALSE
TRUE
Overlapping uORFs
D
b= -0.05; p=1.15e-9
Proportion of AUG uORFs
E
b= -0.0026 ; p=0.013
uORF Kozak score
F
b= -0.0006 ; p=0.38
CDS Kozak score
G
b= -0.01; p=7.89e-7
Log2 Distance to CDS start (nt)
H
b= -0.0018; p=0.18
Log2 Distance from the cap (nt)
I
b= 0.0006; p=0.81
Log2 uORF length (nt)
J
b= 0.058 ; p=0.005
uORF GC content (%)


A
Log2 (Denr KD/Ctrl)
RNA
RPF
Clock
Bmal1
Per1
Per2
Per3
Cry1
Cry2
Nr1d1
Nr1d2
Rora
B
Log2 CDS TE (DenrKD/Ctrl)
C
Clock
Annot. CDS
uORF 1
uORF 2
Alter. CDS
Control
RPF-seq
RNA-seq
Denr KD
Normalized read count
D
Relative translation
Log2 RPF (Denr Kd/Ctrl)
uORF1
uORF2
Annot. CDS
E
Translation efficiency
Log2 TE (Denr Kd/Ctrl)
F
Renilla Luc.
PGK1 promoter
Firefly Luc.
Annot ATG
Alter ATG
Scramble shRNA
Denr shRNA
Normalized Firefly/Renilla luciferase signal
Empty
Clock Wt
Mut.Annot
Mut.Alter
Mut.uORF1
Mut.uORF2
Mut.uORF1+2
Mut.Annot+uORF1
Mut.Annot+uORF2
Mut.Annot+uORF1+2
Mut.Alter+uORF1
Mut.Alter+uORF2
Mut.Alter+uORF1+2
G
Clock
DENR
uORF 1
uORF 2
"Alternative" CDS
9 aa
MVFTVSCSKMSSIVDRD
Annotated AUG
Alternative AUG