Translational regulation of rhythmic and constitutive gene expression

Castelo-Szekely Violeta

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Center for Integrative Genomics

TRANSLATIONAL REGULATION OF RHYTHMIC AND CONSTITUTIVE GENE EXPRESSION

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Violeta CASTELO-SZEKELY

MSc. Neurosciences at Charité Universitätsmedizin, Berlin, Germany
and Vrije Universiteit, Amsterdam, The Netherlands.

Jury

Prof. David Vernez, Président
Prof. David Gatfield, Directeur de thèse
Prof. Christian Fankhauser, Co-directeur
Prof. Achim Kramer, expert
Prof. Felix Naef, expert

Lausanne 2018

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Center for Integrative Genomics

TRANSLATIONAL REGULATION OF RHYTHMIC AND CONSTITUTIVE GENE EXPRESSION

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Violeta CASTELO-SZEKELY

MSc. Neurosciences at Charité Universitätsmedizin, Berlin, Germany
and Vrije Universiteit, Amsterdam, The Netherlands.

Jury

Prof. David Vernez, Président
Prof. David Gatfield, Directeur de thèse
Prof. Christian Fankhauser, Co-directeur
Prof. Achim Kramer, expert
Prof. Felix Naef, expert

Lausanne 2018

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	David	**Vernez**
Directeur·trice de thèse	Monsieur	Prof.	David	**Gatfield**
Co-directeur·trice	Monsieur	Prof.	Christian	**Fankhauser**
Expert·e·s	Monsieur	Prof.	Achim	**Kramer**
	Monsieur	Prof.	Felix	**Naef**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Violeta Castelo Szekely

Master in Neuroscience Vrije Universiteit Amsterdam, Pays-Bas
& Charité-Universitätsmedizin, Berlin, Allemagne

intitulée

Translational regulation of rhythmic and constitutive gene expression

Lausanne, le 21 septembre 2018

pour le Doyen
de la Faculté de biologie et de médecine

Prof. David Vernez

Contents

Acknowledgements

Many are the people who have contributed, helped and made possible to carry out the work presented here, and that I would like to acknowledge:

David Gatfield, for the opportunity of a PhD that combines experimental and computational approaches, for sharing his knowledge on the RNA world and beyond, and for doing his best to make an exciting challenge out of my PhD.

Christian Fankhauser, Achim Kramer, Felix Naef and David Vernez, as members of the thesis committee, for their time, advice and knowledge to bring this work to the best end.

Peggy Janich, for teaching me first-hand the ribosome profiling technique, helping me to make quick progress.

Bulak Arpat, for introducing me to the universe of computational biology and bioinformatics with immense patience and knowledge, and for being an excellent game opponent.

Mara De Matos, for the endless clonings.

Colleagues and friends from the lab and the department, with whom I shared many a good moment, football victory and scientific discussion.

Jernej Ule and everyone in his lab, for letting me be part of such a vibrant group for one year of my PhD, and for the good memories I collected in London.

Roman Klemz, Sabrina Wendt and Achim Kramer for introducing me to the world of circadian biology back in 2012, giving me the motivation to go for a PhD in this field.

Dries Kalsbeek, for letting me continue learning about circadian physiology in his lab in 2013 and for the helpful advice he gave me before and during my PhD.

The NCCR RNA & Biology and the University of Lausanne, for funding.

Finally, I would like to thank my parents and my friends, for loving, helping, teaching, healing, amusing and bearing me. For the one-in-a-million stories. And for dancing with me. For their love knows no regulation in time or space.

Akshat, for being The Bestest.

Eli, for the Bristol Tales, for the parallel lives, and for walking along, miles away.

Abstract

Circadian rhythms govern many aspects of metabolism, physiology and behaviour and are associated with extensive oscillating gene expression. Although oscillating transcription is at the core of circadian gene expression, rhythmic post-transcriptional mechanisms can also explain numerous observations at the level of the final rhythmic protein output. In the first part of this thesis, the contribution of translation to the generation and modulation of diurnal gene expression in mouse liver and kidney was quantified using ribosome profiling. I observed that translationally engendered rhythmicity is largely organ-specific, and that tissue differences in translation efficiency served to modulate the phase and levels of protein biosynthesis from rhythmic mRNAs. Finally, upstream open reading frame (uORF) translation was identified as a novel mechanism of circadian clock regulation. Several core clock genes contained translated uORFs, and loss-of-function of DENR, a protein involved in translation reinitiation after uORF usage, led to circadian period shortening in cells.

In the second part of the thesis, I used ribosome profiling in NIH3T3 cells to annotate uORFs acting through DENR-mediated translation reinitiation transcriptome-wide, and characterised the uORF features that confer DENR dependance using a comprehensive regression model. I identified 240 transcripts that required DENR for efficient protein production, including genes associated with increased proliferation and cancer. DENR targets were enriched for the presence of translated uORFs, and were involved in functional pathways related to translation, transcript stability and the cell cycle. The presence of multiple uORFs, the uORF start codon identity and sequence context, and the distance to the main ORF, correlated with the magnitude of translational regulation, indicating that DENR promotes leaky scanning or reinitiation after strong uORF translation. Finally, within core clock genes, *Clock* was found as a DENR target. *Clock* contains two AUG-initiated uORFs, the second of which overlaps the main CDS and is followed by an additional AUG codon in frame with the annotated CDS, that could give rise to a 5' alternative CLOCK protein. Systematic mutation of the uORF, annotated or alternative CDS start codons suggested that DENR is required for reinitiation and biosynthesis of the alternative CLOCK isoform after translation of the overlapping uORF. In summary, this study provided important insights into uORF-mediated translational mechanisms that can regulate circadian clock function and gene expression at large.

Résumé

Le rythme circadien gouverne plusieurs aspects métaboliques, physiologiques et du comportement, et est associé à une forte expression oscillante des génes. L'oscillation de la transcription est à la base de la expression circadienne. Cependant, les mécanismes post-transcriptionnelles peuvent aussi expliquer certaines observations quant au niveau de la rythmicité des protéines. Dans la première partie de ma thèse, la méthode de profil ribosomique a permis de quantifier la contribution de la traduction à la génération d'oscillation de l'expression des gènes dans le foie et les reins de souris. J'ai donc pu observer que la rythmicité engendrée par la traduction est principalement organe-spécifique. Ces différences entre organes dans l'efficacité de la traduction induisent la modulation de phase et de niveau d'expression des protéines à partir des ARNm rythmiques. Finalement, nous identifions la traduction des "uORF" ("upstream open reading frame") comme un nouveau méchanisme de régulation de la fonction circadienne. Plusieurs gènes essentiels à l'horloge moléculaire contenaient des uORFs traduit. La perte de fonction de DENR, protéine impliquée dans la réinitialisation de la traduction après un uORF, conduit à un raccourcissement de la période circadienne.

Dans la deuxième partie de ma thèse, j'ai identifié les cibles de DENR avec la méthode de profil ribosomique dans la ligne cellulaire NIH3T3. Ensuite, j'ai caracterisé les propriétés des uORFs qui sont modulés par DENR avec un modèle de régression linéaire. 240 transcrits qui nécessitent DENR pour une production protéique efficace ont été identifiés, dont certains sont associés au cancer. La présence de plusieurs uORFs, l'identité et le contexte du codon d'initiation du uORF ainsi que la distance avec l'ORF principal étaient corrélés avec la magnitude de la traduction. Finalement, parmis les gènes essentiels au rythme circadien, *Clock* a été défini comme cible de DENR. *Clock* contient deux uORFs commençant par AUG dont le deuxième chevauche la séquence codante principal ("CDS") et est suivi par un codon AUG additionel. Cela pourrait entrainer la synthèse d'une protéine CLOCK altérnative. Des mutations dans les codons initiaux des uORFs ou de la CDS ont suggéré que DENR est requis pour la synthèse de l'isoforme altérnatif de CLOCK après la traduction de l'uORF chevauché. En résumé, cette étude a apporté des nouveaux méchanismes de régulation de la fonction circadienne et de la expression des génes en général basés sur la traduction des uORFs.

[Traduit de l'anglais par Mariano Schiffrin, PhD.]

Introduction

During my thesis, I was interested in the study of translational regulation of gene expression in mammals, and in particular, in the context of circadian rhythms.

In the first chapter, post-transcriptional mechanisms regulating circadian gene expression are introduced, with special interest in translational control. I then present two studies – Janich et al. [2015, 2016] and Castelo-Szekely et al. [2017] – aimed at elucidating the contribution of translational efficiency to tissue-specific rhythmic gene expression.

The second chapter takes on our finding of upstream open reading frames (uORFs) as novel regulators of circadian function, and introduces the regulatory mechanisms exerted by uORFs and by the translation reinitiation factor DENR. The chapter then presents an ongoing work studying the sequence specificities of DENR-mediated reinitiation transcriptome-wide, and the functional relevance of uORFs within circadian core clock transcripts.

1 Post-transcriptional and tissue-specific regulation of circadian gene expression

Environmental fluctuations due to the Earth's rotation are predictable; therefore being able to anticipate and adapt to them offers an advantage to any living system. To do so, most organisms have developed circadian (from the Latin *'circa'*: approximately, and *'diem'*: a day) clocks: endogenous, self-sustained, entrainable, and temperature-compensated oscillators ticking with a period of about 24 hours.

Molecular mechanism of the mammalian circadian clock: the transcriptional model

In mammals, the circadian system is organised hierarchically, with a central pacemaker located in the brain suprachiasmatic nuclei (SCN) synchronising peripheral clocks present in virtually every cell of an organism. In turn, the SCN receives photic inputs from the environment via the retino-hypothalamic tract connecting to the photosensitive retinal ganglion cells, thus entraining internal timing to geophysical time (reviewed in Mohawk et al. [2012]).

Molecularly, the clock mechanism is similar across cyanobacteria, fungi, plants, flies, and mammals [Doherty and Kay, 2010], consisting in a network of transcriptional activators and repressors interacting in feedback loops and driving the expression of core clock components [Ko and Takahashi, 2006].

In the core mammalian feedback loop, CLOCK and BMAL1 heterodimerise and bind E-box cis-regulatory enhacers to activate the transcription of target genes, including *Period* (Per1, Per2, Per3), and *Cryptochrome* (Cry1, Cry2). PER:CRY heterodimers then translocate back to the nucleus and repress CLOCK:BMAL1 activity, thus inactivating their own expression and closing the loop (Figure 1.1) [King et al., 1997; Gekakis et al., 1998; Kume et al., 1999; Bunger et al., 2000].

In a secondary feedback loop, CLOCK:BMAL1 also drive the transcription of *Rev-erb*α and *Ror*α, and these, in turn, repress and activate *Bmal1* expression, respectively, by binding to retinoic acid-related orphan receptor response elements (ROREs) present in the promoter (Figure 1.1) [Preitner et al., 2002; Sato et al., 2004; Guillaumond et al., 2005].

In order for the cycle to take $\sim$24 hours to complete, a delay must be introduced between the activation and repression parts of the loop. Post-translational modifications of core clock proteins have been shown to play a critical role by regulating stability and nuclear translocation. Phosphorylation of CLOCK leads to rhythmic DNA binding [Yoshitane et al., 2009], and phosphorylation of PER proteins by Casein Kinase 1 (CK1) δ and ϵ induces PER degradation, thus delaying their accumulation and entry into the nucleus [Eide et al., 2002; Akashi et al., 2002]. Circadian sumoylation of BMAL1, induced by CLOCK, parallels its activation and is required for normal rhythmicity [Cardone et al., 2005], whereas BMAL1 acetylation, also by CLOCK, facilitates CRY1 recruitment to the heterodimer [Hirayama et al., 2007].

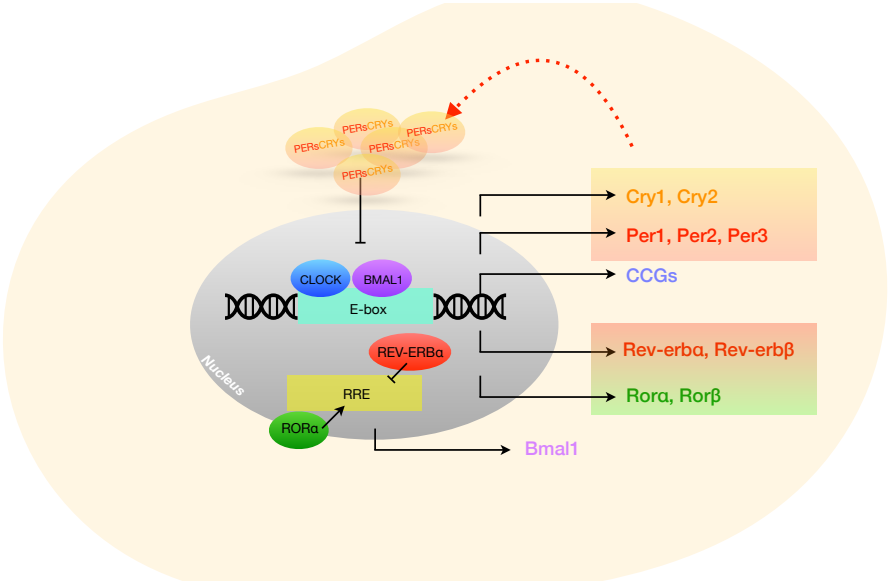


Figure 1.1: Mammalian molecular clockwork. Timekeeping relies on negative transcriptional feedback loops that give rise to oscillatory gene expression. In the core loop, the transcriptional activator CLOCK:BMAL1 drives the expression of its own repressors, PERs:CRYs; whereas in the interconnecting limb, CLOCK:BMAL1 activate the transcription of *Rev-erbα* and *Rorα*, which, in turn, repress and activate *Bmal1* expression, respectively. [CCGs: clock-controlled genes; RRE: Ror-responsive element].

The importance of core clock components for the generation and maintenance of cell-autonomous circadian rhythms is manifested by the profound impact of their disruption on the molecular oscillations and normal behaviour of the animal. For example, *Per1*$^{-/-}$, *Per2*$^{-/-}$ or *Cry1*$^{-/-}$ mice display shorter free-running periods; *Cry2*$^{-/-}$ mice exhibit longer behavioural periods; and disruption of both *Per1* and *Per2*, or *Bmal1*, results in behavioural arrhythmicity in mice [Ko and Takahashi, 2006]. Moreover, disruption of CKϵ leads to PER2 stabilisation, and a subsequent shortening of the circadian period that is implicated in familial advanced sleep phase syndrome (FASPS) [Lowrey et al., 2000; Toh et al., 2001; Xu et al., 2005], highlighting the importance of post-translational modifications.

Challenges to the transcriptional model: evidence for post-transcriptional control

CLOCK:BMAL1-activated transcription can also give rise to the rhythmic expression of clock-controlled genes (CCGs) (Figure 1.1), partly responsible for oscillating circadian outputs and functions in each tissue. However, although transcription is the driver of mRNA synthesis, protein levels also reflect regulatory processes at the post-transcriptional level, including 3' end cleavage and polyadenylation, splicing, nuclear export, RNA stability, translation and protein degradation.

Genome-wide studies comparing rhythmically transcribed mRNAs to rhythmically accumulated transcripts in mouse liver revealed a large discrepancy. Parallel quantification of nascent RNA expression or RNA Pol II loading as a measure of transcription, and steady-state mRNA accumulation using microarrays, showed that ∼25 - 70% of rhythmic mRNAs did not originate from oscillatory transcription [Menet et al., 2012; Le Martelot et al., 2012]. Similarly, quantifying steady-state mRNA levels with RNA-seq and using intronic and exonic reads as proxy for transcription and transcript abundance, respectively, estimated that only 22% of cycling transcripts were also rhythmically transcribed [Koike et al., 2012]. On the other hand, comparative analyses across organs have shown that circadian transcriptomes are largely divergent, despite a common rhythmic core clock [Panda et al., 2002; Storch et al., 2002; Zhang et al., 2014]. Together, these studies indicate that post-transcriptional mechanisms account for a significant part of the observed circadian oscillations, and that systemic cues, local oscillators and tissue-specific factors, all contribute to the generation and modulation of circadian outputs.

Evidence for post-transcriptional regulation that is time-of-day dependent has been shown at

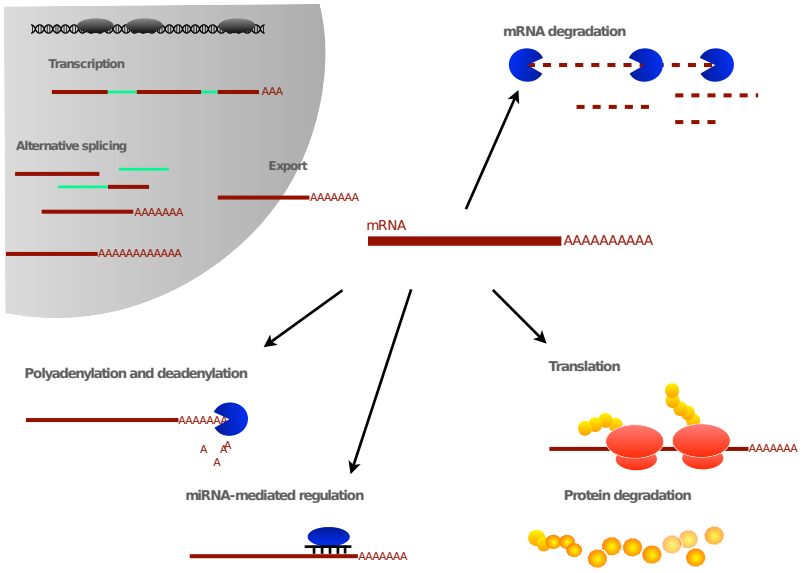


Figure 1.2: Post-transcriptional regulation of rhythmic gene expression. Circadian gene expression can be regulated at virtually every level from transcription to protein degradation. This figure illustrates mRNA processes known to be under the control of the circadian clock, including polyadenylation, translation and degradation.

virtually every stage of the mRNA life, from splicing to translation, stability and degradation (Figure 1.2). Understanding the function, extent and tissue-specificity of these processes in the generation of rhythmic gene expression is relevant for circadian physiology and behaviour.

Alternative splicing.

Alternative splicing is the process by which several transcript isoforms are produced from a single gene, through the differential inclusion/exclusion of exons, giving rise to functionally and structurally different proteins. Given the pervasiveness of the process, altering the relative expression of isoforms in a daily fashion could broadly influence the stability and translatability of mRNAs in a tissue-specific fashion.

In the only study addressing this question, the use of exon arrays in mouse liver identified surprisingly few (47, 0.4% of detectable genes) rhythmic genes regulated by circadian splicing in a tissue-specific manner, with alternative splicing phases occurring throughout the circadian

cycle. [McGlincy et al., 2012].

Cytoplasmic polyadenylation.

Circadian adenylation and deadenylation activities can lead to oscillating poly(A)-tail lengths and, in turn, to rhythmic association to polyribosomes and mRNA degradation, respectively.

Nocturnin is a circadian deadenylase highly expressed in many tissues, including kidney and liver [Wang et al., 2001]. *Noc*$^{-/-}$ mice displayed normal circadian rhythms and clock gene expression but were resistant to obesity upon high fat diet [Green et al., 2007]. Furthermore *Noc* expression paralleled that of multiple clock-controlled genes, suggesting its involvement in gating circadian output pathways.

The extent of daily variations in poly(A)-tail length in mouse liver was addressed transcriptome-wide by Kojima et al. [2012]. They observed widespread rhythmicity in poly(A) tail lengths that could mostly be explained by co-transcriptional nuclear adenylation. However, they identified ∼42 transcripts (18% of poly(A) rhythmic mRNAs) whose poly(A) tail length variation could not be explain transcriptionally or by steady-state mRNA level oscillations, and were therefore under the control of cytoplasmic circadian polyadenylation. Importantly, the peak in tail lengths of these transcripts correlated with maximal expression of polyadenylation-related proteins, such as cytoplasmic polyadenylation element binding proteins (CPEBs), and with maximal accumulation of the encoded proteins.

MicroRNAs.

MicroRNAs (miRNAs) are small (∼22 nt) non-coding RNAs that mediate mRNA silencing through translation inhibition and mRNA destabilisation, by base-pairing to complementary sequences typically located on the 3' UTR. Given that up to 60% of mammalian protein-coding

transcripts have been estimated to be regulated by miRNAs [Lewis et al., 2005; Friedman et al., 2009], it is conceivable that rhythmically expressed miRNAs contribute to the generation of oscillatory gene expression post-transcriptionally, and that miRNA-mediated regulation further modulates cyclic transcription. By genetically inactivating miRNA biogenesis in mouse liver, it was shown that the core clock was largely unaffected by miRNA loss, and that $<$2% of the mRNA oscillations were dependent on miRNAs [Du et al., 2014]. However, miRNA served to modulate the phase and amplitude of $\sim$30% of the rhythmic transcriptome [Du et al., 2014].

mRNA and protein degradation.

Daytime-dependent mRNA and protein stability can readily explain rhythms in steady-state mRNA levels from non-oscillatory transcription, and cyclic protein accumulation from constitutively expressed transcripts, respectively. Long transcript and protein half lives can also account for cases of rhythmic transcriptional and translational activities that are followed by constant steady-state levels [Le Martelot et al., 2012; Menet et al., 2012; Reddy et al., 2006]. Moreover, rhythmic half-lives can modulate the phase, amplitude and magnitude of transcript and protein rhythms, and explain the delay observed between maximal transcription and transcript abundance, and between mRNA and protein level peaks [Luck et al., 2014].

Translation efficiency.

Translational regulation by RNA-binding proteins (RBPs) has been shown to play an important role in the regulation of core clock protein expression, and for the temperature entrainment of the clock. In cultured fibroblasts, physiological temperature cycles drove the rhythmic expression of cold-inducible RNA-binding protein (CIRP), and cross-linking and immunoprecipitation (CLIP) experiments identified *Clock* among its targets [Morf et al., 2012]. CLOCK protein expression

was strongly reduced in *Cirp*-depleted cells, but total *Clock* mRNA showed only a modest decrease, consistent with a role of CIRP in controlling *Clock* translation [Morf et al., 2012]. Furthermore CIRP binding sites were significantly enriched near polyadenylation sites (PAS) and low temperatures (*Cirp* upregulation) lengthened 3'UTRs, whereas CIRP depletion shortened them, indicating that CIRP dictated the usage of PAS, and can therefore regulate circadian gene expression by controlling alternative polyadenylation [Liu et al., 2013].

RBPs and, in particular, heterogeneous nuclear ribonucleoproteins (hnRNPs), have also been implicated in degradation and cap-independent translational regulation of core clock genes. HnRNP I, also known as polypyrimidine-tract-binding protein (PTB), binds the 3'UTR of *Per2*, promoting its degradation. Importantly, PTB protein levels are in antiphase to *Per2*, supporting its stability regulatory role [Woo et al., 2009]. Similarly, hnRNP D (or ARE-binding factor 1, AUF1) mediated the rhythmic decay of *Cry1* mRNA, by binding to the 3'UTR, and rhythmic accumulation of hnRNP D was antiphasic to *Cry1* mRNA expression [Woo et al., 2010]. Finally, IRES-dependent translation has been reported for *Rev-erb*α by interaction with PTB [Kim et al., 2010], and for *Per1*, through the interaction with hnRNP Q [Lee et al., 2012a].

The above constituted individual examples of translational modulation of clock protein expression. With the advent of sequencing technologies and the improvement in mass spectrometry (MS) quantitative accuracy, these findings could be extended to a transcriptome-wide scale.

A systematic analysis of the circadian proteome by two-dimensional difference gel electrophoresis (2D-DIGE) and MS in mouse liver revealed that up to 20% of the assayed proteins fluctuated in a daytime-dependent manner, but that only half of them derived from a rhythmic transcript [Reddy et al., 2006], providing one of the earliest evidence for post-transcriptional mechanisms regulating circadian gene expression.

More recently, Robles et al. [2014] and Mauvoisin et al. [2014] in mouse liver, and Chiang et al. [2014] in the SCN, have revisited this question using high resolution MS-based proteomics together with quantification via the *in vivo* stable isotope labeling by amino acids (SILAC) technology. Although the total protein detection varies across studies, all three reported that 4-20% of the detected proteome oscillates in a circadian manner. The proportion of the circadian proteome that cannot be accounted for by an underlying transcript rhythmicity also differed among the three studies, likely due to different experimental and analytical methods. However, non-transcriptional contribution was estimated at 20 [Robles et al., 2014] to 50% [Mauvoisin et al., 2014]. Interestingly, all studies observed a phase delay between transcript and protein accumulation of 5 to $>$8 hours that is daytime dependent and is consistent with a diurnal control of protein half-life too. Moreover, the time-of-day proteome was enriched for key biological processes such as mitochondrial oxidative phosphorylation in the SCN [Chiang et al., 2014], and xenobiotic detoxification [Robles et al., 2014] and protein secretion [Mauvoisin et al., 2014] in liver, highlighting the tissue-specificity of translational and post-translational circadian regulatory mechanisms.

Jouffe et al. [2013] used microarray analysis of total and polysomal RNAs to account for diurnal variations of the mouse hepatic translatome. Polysome profiling is based on the fractionation of transcripts in a sucrose gradient according to the number of ribosomes associated to them. It therefore provides an estimate of the relative levels of translation: monosome fractions reflect the translationally arrested pool, while heavier fractions are associated with increasing translational levels. 249 probes, corresponding to $\sim$2% of the expressed genes, showed a rhythmic polysome profile and a constant mRNA abundance, with the peak in polysome association clustering at the day-to-night transition. The vast majority of these translationally regulated transcripts were

related to ribosome biogenesis and belong to the 5'-terminal oligopyrimidine tract (5'TOP) gene family.

Briefly, TOP mRNAs are characterised by the presence of a C residue at the 5'cap (only about 17% of mammalian transcripts start at C), followed by a stretch of 4-14 pyrimidines and a GC-rich sequence, and they usually have short 5'UTRs devoid of uAUGs. The location and the presence of a C at the first position are crucial for the functional importance of this cis-regulatory element [Meyuhas, 2000]. TOP mRNAs appear to be in an 'all-or-none' association state to polysomes, in which cellular growth arrest leads to the fast dissociation from polysomes and translation repression, and nutrient and energy availability triggers polysome loading and translation activation [Meyuhas, 2000]. Thus the presence of this cis-element allows a fast and coordinated response of the whole translational machinery to external growth stimuli. Although the precise regulatory mechanism has not been fully elucidated, the translation efficiency of TOP mRNAs responds to signalling via the mammalian target of rapamycin (mTOR) pathway, which is thought to activate S6 kinase and RPS6 phosphorylation [Meyuhas and Dreazen, 2009], as well as the inhibitory eIF4E-binding proteins (4E-BPs). Since 4E-BP competes with eIF4E for binding to eIF4G, it was hypothesised that loss of eIF4E-eIF4G interaction could impair TOP mRNAs translation more strongly than other mRNAs [Thoreen et al., 2012]. Additionally, it could involve one or several *trans*-acting elements recognizing the motif, as well as other elements of the initiation machinery to which the S6K and/or mTOR signals are conveyed [Hamilton et al., 2006].

Accordingly, mTORC1 expression, but also the phosphorylation state of 4E-BP1, were all rhythmic in mouse liver [Jouffe et al., 2013].

Although a large number of fractions can be obtained from a polysome gradient, these are

usually pooled to reduce the costs associated with cDNA microarrays. Pooling fractions into those containing e.g. $<$2 ribosomes, 2-n and $>$n ribosomes can limit the quantitative analysis of translational regulation to "on-off" states or large relative changes, compromising accurate calculations of translational efficiency. Moreover polysome profiling does not provide positional information on the ribosomes within ORFs or regulatory regions that can gain insights into the regulatory mechanisms underlying translational fluctuations.

By contrast, ribosome profiling is a method to determine the distribution of actively translating ribosomes on an mRNA, providing a snapshot of a cell's translatome at a given time [Ingolia et al., 2009, 2012; Ingolia, 2014; MGlincy and Ingolia, 2017]. The basis of ribosome profiling is that a ribosome covers a sequence of $\sim$30 nucleotides and protects it from nuclease digestion [Wolin and Walter, 1988]. By sequencing these ribosome footprints or ribosome-protected fragments (RPF-seq), the precise location of ribosomes and the codon being translated can be determined transcriptome-wide (Figure 1.3). Relative translation efficiency can then be quantified by the ratio of ribosome footprints to the total mRNA present (from total RNA-seq done in parallel).

Ribosome profiling has been widely used in the last decade to quantify gene-level translational control in *Drosophila* [Dunn et al., 2013], yeast [McManus et al., 2014] and mouse embryonic stem cells [Ingolia et al., 2011]; during the cell cycle [Stumpf et al., 2013] and mitosis [Tanenbaum et al., 2015]; in disease phenotypes [Schafer et al., 2015]; in the mitochondria [Rooijers et al., 2013]; by miRNAs [Bazzini et al., 2012]; to define uORFs [Bazzini et al., 2014]; to assess the coding potential of long non-coding RNAs [Guttman et al., 2013]; or to study the different stages of translation elongation itself [Lareau et al., 2014], among many others.

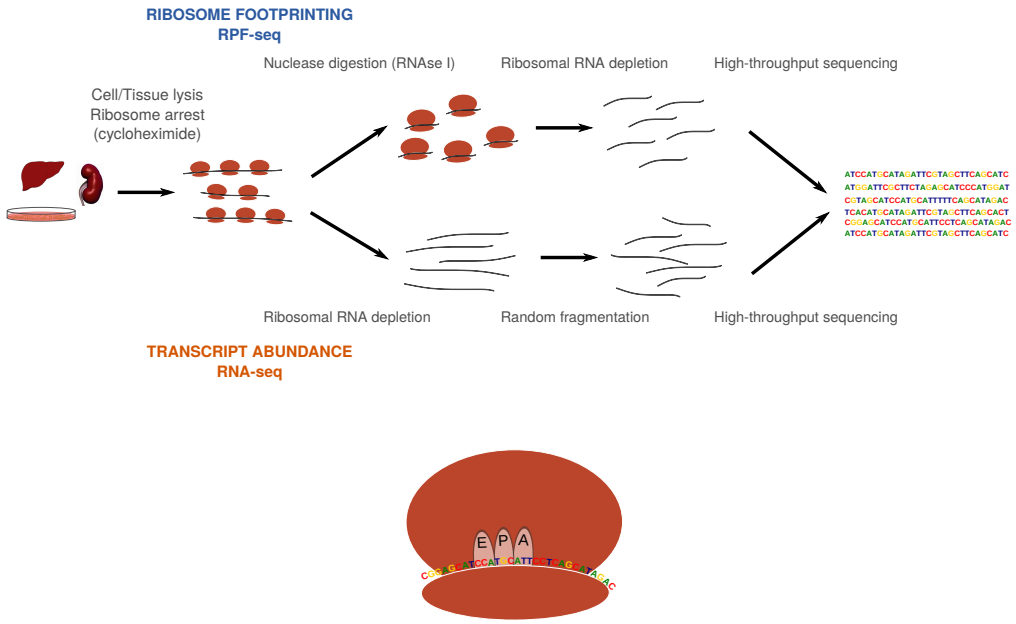


Figure 1.3: Ribosome profiling technique. *Top*: For the preparation and sequencing of ribosome-protected fragments (RPF-seq), cells/tissues are lysed in the presence of cycloheximide, to arrest elongating ribosomes. Following RNA digestion, ∼30 nt footprints are generated, from which libraries can be prepared. In parallel, total mRNA libraries (RNA-seq) are made from ribosome-depleted and randomly fragmented RNAs from the same lysate. *Bottom:* Since the ribosome protects a 30-nt fragment, the codon being translated (A-site) can be inferred as the position +15 (from the 5' of the footprint).

Investigating translational regulation around-the-clock in mouse liver and kidney

Ribosome profiling is suitable to study the dynamics of translational efficiency across the circadian cycle as well. The extent of uncoupling between transcript abundance rhythms, as measured by RNA-seq, and ribosome occupancy rhythms, from RPF-seq, can be thus quantified, identifying cases of oscillations at either level, and the phase relationship of transcripts cycling at both. Moreover, the relative stoichiometric levels of core clock proteins can be inferred, since the ribosome occupancy on a transcript is a direct measure of the absolute amount of protein synthesized, assuming that a ribosome engaged on an mRNA will produce a protein [Li et al., 2014]. Finally, using the ribosomal positional information, ribosome profiling can

gain insights into the causal sequences modulating translation efficiencies, such as uORFs or ribosomal pausing.

Using ribosome profiling around-the-clock in mouse liver, we aimed to study a new layer of post-transcriptional regulation of circadian gene expression, namely, translation efficiency, [Janich et al., 2015, 2016], and do so in a tissue-specific and comparative fashion [Castelo-Szekely et al., 2017].

Janich et al. 2015 and 2016: Results summary and contribution to the study

In Janich et al. [2015] (Research Articles, page 75) we used ribosome profiling in mouse liver to study the contribution of daytime-dependent translation to protein level oscillations that cannot be explained by rhythmic transcription.

Ribosome profiling data was collected with a 2-hour resolution in light-dark entrained mice and constituted one of the first studies applying this technique in a solid mammalian organ. We therefore verified that data was of high quality and nucleotide resolution (e.g. by looking at footprint size, and their distribution and periodicity along the transcript), and could recapitulate its temporal nature (article's figure 1).

Globally and independently of the time component of the data, translation efficiency showed an asymmetric distribution, similar to that observed in cells [Ingolia et al., 2011], with a longer negative tail suggestive of inhibitory mechanisms. In this line, 5'UTR length and the frequent presence of uORFs within them, were shown to correlate inversely with translation efficiency (article's figure 2), consistent with the predominantly negative effect of uORFs on main CDS

translation [Wethmar, 2014].

Rhythmicity detection and parameter estimation were performed on RPF and RNA-seq datasets separately based on AIC model selection, and included a 1.5 fold cutoff on amplitude. Within the core clock, rhythmic transcript abundance was a good predictor of rhythmic ribosome occupancy, and both oscillations were tightly coupled without further modulation by time-of-day dependent translation efficiency (article's figure 3). This phase synchrony was also observed globally for oscillating transcripts – the majority of rhythmic events detected in this organ – whereas transcripts cycling at either the mRNA or protein synthesis levels showed distinct peak distributions, during the night phase and at the light-to-dark transition, respectively (article's figure 4).

Robust daytime-dependent translation of non-oscillating transcripts was detected in a set of 150 genes, showing the peak of ribosomal occupancy at the onset of the dark phase. This genes were enriched for components of the translation machinery including ribosomal proteins, whose regulation has been shown to involve both the circadian and the feeding cycles [Jouffe et al., 2013; Atger et al., 2015]. Moreover, we identified transcription factors and transcripts containing iron-responsive elements, suggesting that rhythmic protein synthesis could feedback to the circadian transcriptome, and regulate iron metabolism (article's figure 5).

Finally, since the presence of uORFs was widespread in liver transcripts and modulates translation efficiency, we explored the impact of uORF translation on the circadian clock (article's figure 6). Among core clock genes, *Bmal1, Clock, Cry1, Rev-erbα and Rev-erbβ*, contained translated uORFs whose ribosome occupancy rhythmicity followed that of the main CDS. When the start codons of the uORFs detected in *Rev-erbα* were mutated in cells containing a circadian reporter, we observed increased levels of the reporter activity, indicating that uORFs

are involved in regulating *Rev-erb*α levels and possibly the magnitude of its oscillation *in vivo*. In order to study the overall effect of uORF translation on the clock function, we knocked-down *Denr* (Density Regulated Protein), a protein implicated in ribosome recycling and reinitiation after uORF translation ([Schleich et al., 2014] and chapter 2). Cells deficient for *Denr* displayed a $\sim$1.5-hour shorter period (article's figure 6), uncovering uORF translation as a novel mechanism regulating circadian function.

Janich et al. [2016] (Research articles, page 121) is a methods article that provides additional information on the experimental and analytical details to generate the ribosome profiling data used in Janich et al. [2015]. It presents an outline of the experimental workflow to prepare RPF and RNA-seq libraries (article's figure 1), and insights into some of its critical steps, such as nuclease digestion, monosome purification and rRNA depletion. For example, in order to ensure reproducible $\sim$30 nt footprint generation, nuclease digestion conditions were optimised in preliminary tests (article's figure 2), and verified for the time course experiment (article's figure 3). Moreover, the article summarizes the main steps for sequencing quality assessment, data preprocessing, and generation of per-gene count tables ready for subsequent gene expression and translation efficiency calculations.

Contribution to the studies

This study was designed by Prof. David Gatfield; whole-organ ribosome profiling was set up and performed by Dr. Peggy Janich, and data was analyzed by Dr. A. Bulak Arpat. Dr. Peggy Janich, Maykel Lopes and myself carried out the validation and follow up experiments.

I was involved in the investigations of uORF translation role on circadian clock function and, particularly, in studying the effect of impairing translation reinitiation on the circadian period in

cells (Janich et al. [2015], figure 6). This included the production of lentivirus carrying different shRNAs against *Denr*, the transduction of NIH3T3 fibroblasts carrying a circadian reporter (*Dbp-Luciferase*), the validation of knock-down efficiency by qPCR and Western blotting, and the *in vivo* measurement of bioluminescence rhythms on synchronized cells.

The identification of several core clock genes containing uORFs and their overall impact on circadian phenotype prompted us to investigate circadianly relevant uORFs and reinitiation events transcriptome-wide in a dedicated follow up study (Chapter 2).

Castelo-Szekely et al. 2017: Results summary and contribution to the study

In Castelo-Szekely et al. [2017] (Research Articles, page 127) we studied translational regulation across time, i.e. around the day, and space, i.e. across tissues.

As in Janich et al. [2015], we used ribosome profiling to assess the contribution of translational efficiency to daily gene expression rhythms in mouse kidney, an organ with important circadian functions [Bonny et al., 2013] that shares a large proportion of the expressed transcriptome with liver [Brawand et al., 2011; Zhang et al., 2014]. Furthermore, by contrasting translational rhythms in kidney to those in liver, we observed that the large tissue specificity reported for the rhythmic transcriptome [Storch et al., 2002; Zhang et al., 2014] extended to the translatome, both in transcript identities and in phase clustering.

We first verified that kidney ribosome profiling data was of comparably high quality and resolution than the liver data of the same mice from Janich et al. [2015]. Temporal disposition of the data was similarly reproduced for kidney; albeit in a reduced manner, suggestive of fewer

and/or smaller amplitude rhythms in this organ (article's figure 1).

This study revealed 178 genes with robust rhythms in transcript abundance and ribosome occupancy in both organs (article's figure 3). In contrast to liver, where both oscillations were tightly synchronised, protein production peaks preceded maximal abundance in kidney, suggesting that translation efficiency was markedly regulated in this organ and decreases over the lifetime of the transcript.

Rhythmic translation of constantly abundant transcripts was confined to less than hundred cases in kidney and, interestingly, did not include ribosomal proteins and other translational machinery components, whose translation efficiency remained constantly high along the day (article's figure 4). Since mTOR signalling triggered by feeding-fasting cycles is a known regulator of the expression of these transcripts [Jouffe et al., 2013; Atger et al., 2015], these results suggested a reduced sensitivity of kidney to systemic feeding cues. Moreover, translational oscillations in kidney showed a broader peak distribution and were of overall smaller amplitude than liver rhythms.

Finally, tissue-specific translation efficiency also modulated protein biosynthesis of core clock transcripts, leading to an overall higher convergence between tissues at the level of translation than mRNA abundance (article's figure 5).

On the other hand, the data generated in this comparative study provided insights into constitutive gene expression regulation across organs. We observed that translation efficiencies spanned a narrower range across tissues than mRNA levels, indicating a major role of transcription and mRNA decay in setting gene expression differences (article's figure 2). Nevertheless, tissue-specific translation efficiency was significant for $\sim$9% of the transcriptome and globally counteracted transcript abundance divergence. Furthermore, our anlyses suggested that tissue

specificity in translation efficiency was partly achieved by the differential expression of transcript isoforms that differed in their 5'UTRs (article's figure 2).

Contribution to the study

This study was designed by Prof. David Gatfield. Liver ribosome profiling data came from Janich et al. [2015], generated by Dr. Peggy Janich and analyzed by Dr. A. Bulak Arpat. I carried out the ribosome profiling experiment around the clock in mouse kidney, performed nearly all analyses presented in the study with the invaluable help of Dr. A. Bulak Arpat, and contributed to the writing of the manuscript.

2 Translational regulation by upstream Open Reading Frames

Gene expression regulation can occur at several stages, including translation - the process of protein assembly according to the nucleotide sequence of an mRNA. Translational regulation allows for more rapid and reversible changes in protein concentrations than transcriptional control to restore homeostasis in response to internal or environmental stimuli.

Canonical translation initiation

Translation itself can be global and transcript-specifically regulated at either of the four stages - initiation, elongation, termination, and ribosome recycling - although initiation is considered the rate-limiting step [Jackson et al., 2010]. In eukaryotes, canonical initiation involves a scanning mechanism that requires multiple initiation factors (eIFs) [Sonenberg and Hinnebusch, 2009]. It begins with the formation of the ternary complex (TC) comprised of GTP-bound eIF2 and methionyl-initiator tRNA (Met-tRNAi), and its assembly onto the 40S ribosomal subunit, facilitated by eIF1, eIF1A, eIF3 and eIF5, forming the 43S preinitiation complex (PIC).

The PIC is then recruited to the mRNA at its m7G-capped 5' end, an attachment mediated by the eIF4F complex, composed of the RNA helicase eIF4A, cap-binding protein eIF4E, and eIF4G, a scaffold protein that, in turn, binds to poly(A)-binding protein (PABC) allowing circularization of the transcript. The PIC begins scanning downstream the 5' untranslated region (5' UTR) until it encounters an initiator codon with base complementarity to Met-tRNAi. Upon recognition, eIF5 facilitates irreversible GTP hydrolysis, which triggers the formation of the scanning-arrested 48S PIC, release of eIF2-GDP, and subsequent joining of the 60S subunit, catalyzed by eIF5B, to produce 80S ribosomes ready for translation elongation.

Exceptions to the scanning mechanism involve the direct recruitment of the ribosome to the mRNA at specific internal ribosome entry sites (IRES). IRES-mediated translation initiation is frequent on viral mRNAs and can occur on some cellular mRNAs too, allowing translation on certain transcripts when translation has been globally downregulated.

The initiation rate can be controlled, leading to a global reduction of cap-dependent translation. The two best known mechanisms involve interfering with the TC recycling or with the cap recognition process [Jackson et al., 2010].

In order to allow new rounds of translation initiation, the ternary complex needs to be recycled, a process in which eIF2 is recharged with GTP by eIF2B, a guanine exchange factor (GEF). However, upon a variety of cellular stresses, including accumulation of unfolded proteins in the ER, nutrient starvation, iron deficiency or virus infection, eIF2 can be phosphorylated on Ser51 of its α subunit, rendering it a competitive inhibitor of eIF2B and therefore preventing recycling and TC reconstitution. Four protein kinases are known to phosphorylate eIF2α depending on the type of cellular stress: PKR kinase, activated by double stranded RNA and important in the antiviral response; heme-regulated kinase, most relevant in erythroid cells;

PKR-like endoplasmic reticulum kinase (PERK), activated upon ER stress; and general control non-derepressible 2 kinase (GCN2), activated by nutriend starvation [Jackson et al., 2010].

Phosphorylation also regulates the levels of eIF4F and therefore, the cap recognition competence of the complex and the recruitment of the PIC. The assembly of the complex is hindered by eIF4E-binding proteins (4E-BPs) that compete with eIF4G for binding to eIF4E. Hypophosphorylated 4E-BPs strongly bind eIF4E, preventing cap complex formation, whereas phosphorylation weakens their interaction with eIF4E. The main kinase of 4E-BPs is mammalian target of rapamycin (mTOR), an effector of the PI3K/Akt pathway that integrates signals from cellular nutrient, oxygen and energy availability [Jackson et al., 2010].

Upstream open reading frames and alternative translation initiation

Due to the scanning nature of canonical translation initiation, the first AUG encountered by the ribosome does not often correspond to that of the main protein coding sequence. Translation can thus initiate within the so-called 5' untranslated region (5' UTR) – more recently known as 5' leader – giving rise to upstream open reading frames (uORFs) that can terminate before or overlap with the coding sequence (CDS). uORFs are typically shorter than the CDS, with a mean length of 48 nt [Calvo et al., 2009], and can initiate by near-cognate start codons – with NUG generally working best – embedded in strong contexts similar to the kozak consensus sequence "gccA/GccNUGGG" [Hinnebusch, 2011]. The functional importance of AUG uORFs has been supported by the fact that their occurrence is less frequent than expected by chance [Iacono et al., 2005], and that they show a high degree of evolutionary conservation [Churbanov

et al., 2005; Matsui et al., 2007; Calvo et al., 2009; Arribere and Gilbert, 2013; Johnstone et al., 2016].

Translation of these uORFs requires translation to be reinitiated, a process in which the ribosome retains or reacquires the necessary eIFs and resumes scanning of the 5'UTR before the next start codon. Alternatively, the scanning ribosome might bypass certain uORFs, located in unfavourable contexts, in favour of a stronger uORF further downstream, or of the CDS, in a process known as leaky scanning (Figure 2.1). Translation reinitiation and leaky scanning are inefficient processes, and thus uORF-harbouring transcripts generally show reduced protein synthesis levels from the main CDS [Matsui et al., 2007; Calvo et al., 2009], and also [Janich et al., 2016; Castelo-Szekely et al., 2017]. Moreover, the peptide encoded by the uORF might interfere with the ribosome, blocking further translation or scanning, or triggering mRNA degradation by nonsense-mediated decay (NMD) (Figure 2.1). The presence of uORFs thus serves to modulate translation in a transcript-specific manner. In addition, alternative splicing and alternative promoter usage can give rise to uORF-containing and uORF-lacking isoforms in a tissue and condition-specific manner, further adding to the myriad of mechanisms through which uORFs can act (reviewed in Wethmar [2014]).

Several factors are known to affect the efficiency of reinitiation; for example, short uORFs and uORFs without strong secondary structures – usually requiring shorter translational times – are more likely to be reinitiation-competent, as the ribosome might retain the initiation machinery [Kozak, 2001]. Similarly, the location of the uORF stop codon with respect to the start of the CDS is critical for reinitiation: larger distances would allow the ribosome to regain the required initiation factors, reducing the inhibitory potential of the uORF, whereas a shorter spacing could be insufficient for reinitiation [Kozak, 1987]. In addition to the uORF length and

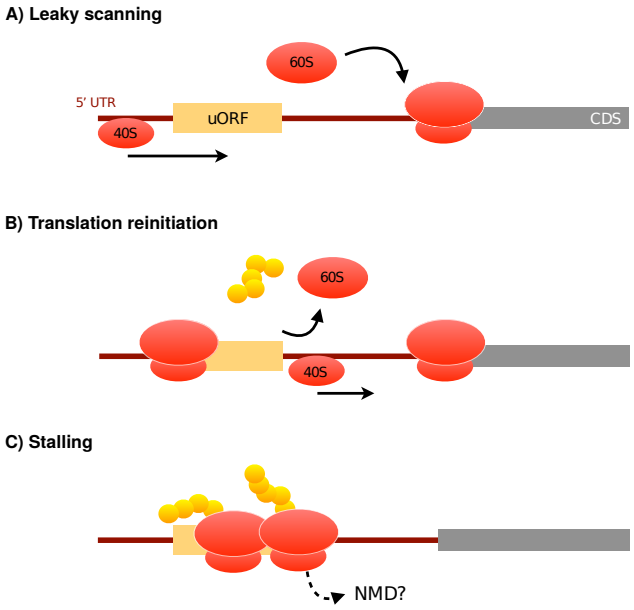


Figure 2.1: Mechanisms of uORF-mediated translational control. A) In leaky scanning, the ribosome can sometimes bypass a uORF and initiate translation at a downstream ORF. The frequency of leaky scanning depends on the efficiency of start codon recognition. B) Translation reinitiation occurs when the small ribosomal subunit remains associated to the mRNA and resumes scanning after uORF translation is terminated. C) Both the nascent uORF peptide and ribosome stalling can create a roadblock for further elongation. Alternatively, ribosome stalling and/or recognition of the uORF stop codon as premature can trigger nonsense-mediated mRNA decay (NMD).

intercistronic distance, sequences upstream, within and downstream of the uORF can all affect reinitiation efficiency and protein production [Mohammad et al., 2017] illustrating the complex regulatory mode exerted by uORFs.

As mentioned above, uORF translation can also trigger NMD, a surveillance mechanism employed by cells to selectively degrade mRNAs with premature termination codons [Behm-Ansmant and Izaurralde, 2006; Singh et al., 2008]. Briefly, during normal translation termination, eukaryotic release factor 3 (eRF3) joins the ribosome upon stop codon recognition, and, in turn, interacts with PABC, triggering ribosome subunit dissociation and peptide release. On the other hand, premature termination codons (PTCs) can be selectively recognized by the NMD pathway, which involves the interaction of eRF3 with UPF1, the NMD effector present in the exon-junction complex (EJC) deposited on the mRNA during splicing. uORF stop codons

can thus be recognised as premature and, due to the large linear distance to the poly-A, where PABC sits, and the likely presence of EJC downstream of the uORF, elicit NMD. Nonetheless, the interaction of PABC with the eIF4F complex during translation initiation, might also favour efficient translation termination, especially on shorter uORFs that retain eIF4F and eIF3, making uORFs NMD-resistant [Barbosa et al., 2013]. Although uORF-containing transcripts have been shown to be more susceptible to NMD [Mendell et al., 2004], the extent of transcript expression regulation by NMD-sensitive uORFs has not been fully elucidated.

Up to 50% of mammalian transcripts contain uORFs [Calvo et al., 2009; Lee et al., 2012b], although their functional relevance has only been experimentally validated in a relatively small number of transcripts [Wethmar et al., 2014]. The best known example of uORF-mediated translational regulation in mammals involves activating transcription factor 4 (*Atf4*), with a regulatory mode similar to that of GCN4 in yeast [Hinnebusch, 2005]. *Atf4* contains two uORFs, the second of which overlaps the CDS. In normal conditions, efficient translation reinitiation occurs at the second uORF, avoiding protein synthesis from the CDS. During nutrient deprivation stress, however, when translation is globally switched off and ternary complex levels are low, the ribosome cannot reacquire the TC after translation of the first uORF on time to restart at the second uORF. Instead, translation reinitiation occurs at the main ORF, thus allowing ATF4 protein expression, which helps restoring cellular homeostasis [Vattem and Wek, 2004].

Since the presence of uORFs generally leads to a reduction in protein synthesis, mutations or polymorphisms that create, delete or alter uORFs could influence disease-relevant phenotypes. Several hundred single nucleotide polymorphisms (SNPs) have been identified within human 5' UTRs that create or delete a uORF, leading to a 30-60% increase in protein production

in the uORF-less variant [Calvo et al., 2009]. As an example, a C-to-T polymorphism in human clotting factor XII (FXII) introduces a short uORF in the 5' UTR that leads to a ∼50% reduction plasma FXII levels, increasing the risk of thrombosis [Kanaji et al., 1998]. Similarly, rare mutations that create uORFs in several genes have also been associated to reduced protein levels and associated to a disease phenotype, including familial hypercholesterolemia (low-density lipoprotein receptor gene; LDLR), cystic fibrosis (cystic fibrosis transmembrane conductance regulator; CFTR), melanoma (cyclin-dependent kinase inhibitor 2A; CDKN2A), or congenital hyperinsulinemia(potassium inwardly-rectifying channel, subfamily J, member 11; KCNJ11) [Barbosa et al., 2013; Calvo et al., 2009].

DENR-MCTS1-mediated translation reinitiation

Despite the prevalence and functional relevance of uORFs, little is known about the factors mediating translation reinitiation.

As mentioned above, the time required for uORF translation can be critical to maintain initiation factors on the ribosome that, in turn, would allow reinitiation. Ribosomes that retain eIF3 have been shown to be competent for reinitiation on yeast *Gcn4*, through its association with reinitiation-promoting elements (RPEs) present upstream of the uORF [Mohammad et al., 2017]. In this way, eIF3 would prevent post-termination ribosome recycling – consisting in the dissociation of tRNA and mRNA from 40S subunits mediated by eRF1 and ATP-binding cassette protein ABCE1 [Dever and Green, 2012] – and allow scanning resumption. The presence of eIF1, eIF1A and eIF4F on terminating ribosomes has also been associated to efficient reinitiation [Jackson et al., 2010; Skabkin et al., 2013].

Non-canonical factors can be required for reinitiation on certain transcripts, too. Density

regulated protein (DENR) and Malignant T-cell amplified sequence 1 (MCTS1) form an heterodimer that can bind the 40S ribosomal subunit. *In vitro* biochemical studies showed that DENR-MCTS1 can recruit the initiator tRNA (Met-tRNAi) to ribosomes that place the P-site directly on an mRNA start codon, in a scanning-independent manner similar to the recruitment of ribosomes to IRES by certain virus, including hepatitis-C virus (HCV) [Skabkin et al., 2010]. Moreover, they can promote the release of deacylated tRNA and mRNA from dissociated ribosomes. DENR expression increases in cultured cells at high density but not at growth arrest [Deyo et al., 1998], and it is overexpressed in several cancers. Similarly, MCTS1 is an oncogene that promotes cell cycle progression [Prosniak et al., 1998; Hsu et al., 2005] and can bind the 5' cap complex to upregulate the translation of a subset of proliferation-related transcripts [Reinert et al., 2006]. Overexpression of MCTS1 increases tumorigenicity of breast cancer cells by promoting angiogenesis and suppressing apoptosis [Levenson et al., 2005].

The DENR-MCTS1 dimer is analogous to eIF2D, where DENR corresponds to the C' terminal part and MCT to the N' terminus of EIF2D (Figure 2.2), suggesting a functional analogy between both factors. In this line, eIF2D can also deliver Met-tRNAi to the ribosomal P-site in a GTP-independent fashion [Dmitriev et al., 2010], and induce efficient ribosome recycling [Skabkin et al., 2010]. Both DENR and MCTS1 contain RNA-binding domains: MCTS1 contains a PUA domain, involved in dsRNA binding in rRNA and tRNA, whereas DENR has a SUI domain, like that found in eIF1 (Figure 2.2), suggesting a role in start codon recognition. Moreover, biochemical studies showed that the presence of eIF1 inhibited the activity of eIF2D and that eIF1 could substitute DENR in binding to MCTS1, although less efficiently [Skabkin et al., 2010]. The structural similarities between eIF2D and DENR-MCTS1 were recently confirmed by X-ray crystallography and cryoelectron microscopy [Lomakin et al., 2017; Weisser

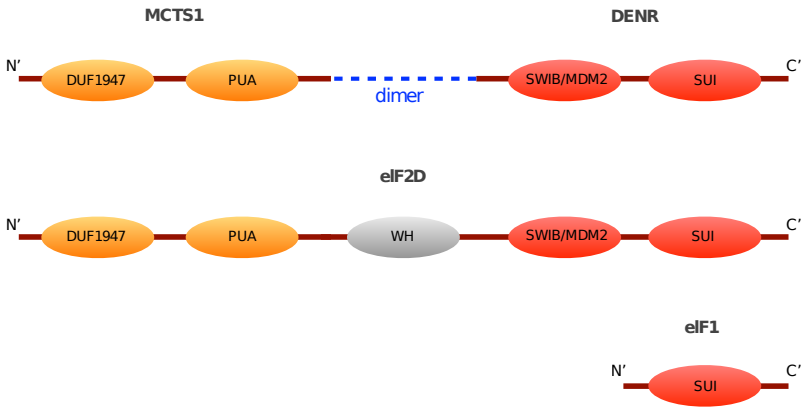


Figure 2.2: Domain organization of DENR-MCTS1 heterodimer, eIF2D, and eIF1. The protein domains of DENR and MCTS1 correspond to those in the C' terminal and N' terminal parts of eIF2D, respectively, whereas the SUI domain is shared with eIF1, and is involved in start codon recognition.

et al., 2017]. Both studies showed that the DENR-like end of eIF2D binds the 40S subunit in close proximity to the mRNA entry channel and P-site of the ribosome, and could interact with the 18S rRNA, and the acceptor arm and the anticodon stem loop of tRNA. Moreover, they revealed that the presence of eIF2D in the ribosome would be incompatible with that of eIF3, suggesting mutually exclusive mechanisms of translation reinitiation.

Involvement of DENR-MCTS1 in translation reinitiation was recently shown in Drosophila [Schleich et al., 2014]. Schleich et al. [2014] reported that *Denr*-homologue knockout flies died as pharate adults, had a larval-like epidermis and showed deficient proliferation of histoblasts. Polysome profiling of *Denr*-knockdown S2 cells showed a decreased protein synthesis for ∼100 transcripts, specifically in proliferating cells. Follow up on one of such targets, *myoblast city* (*mbc*), indicated that the 5'UTR was sufficient to confer DENR dependance to a luciferase reporter, reducing its activity without changes in mRNA levels. A series of systematic deletions on the 5'UTR led to the identification of three short uORFs with strong Kozak contexts required for DENR dependance, which was abolished upon mutation of the start codons, indicating that DENR was required for efficiently translation reinitiation. Moreover the identified targets were

enriched for the presence of uORFs containing strong Kozaks, and were related to functional categories involving in transcriptional regulation and kinases, together suggesting that DENR is necessary for efficient translation of a subset of strong uORF-containing genes involved in proliferation.

More recently, the observations made in *Drosophila* were tested in human *HeLa* cells [Schleich et al., 2017], by constructing 5' UTR reporters carrying synthetic (non-endogenous) strong uORFs. Knockdown of *Denr* or *Mcts1*, but not of *Eif2d*, resulted in reduced luciferase activity, suggesting that, like in Drosophila S2 cells, strong uORF-containing mRNAs require DENR for translation reinitiation. However, in contrast to the observations in Drosophila, DENR requirement was strongly dependent on the length of the synthetic uORF, with only uORFs coding for 1 aminoacid leading to a decrease in reporter activity upon *Denr* knockdown. Taking into account these uORF specificities, Schleich et al. [2017] defined $\sim$100 human transcripts that would be predicted to be DENR-dependant in human cells, most of which were neuronal genes and G-protein-coupled receptors genes, and further validated several candidates. By contrast, an analogous search for strong uORFs in the mouse transcriptome could not be validated in reporter assays, suggesting that the uORF features that confer DENR dependance might vary across cells and/or organisms [Schleich et al., 2017].

Investigating the sequence features of DENR-mediated translational control and the circadian clock regulation by uORFs

Together, the above-mentioned studies indicated that DENR-MCTS1, and likely also eIF2D, can regulate translation reinitiation, and defined ∼100 transcripts predicted to be under DENR control. However, a direct identification of the endogenous DENR-MCTS1 targets in mammalian cells is still lacking, as well as an *in vivo* quantification of the degree of protein synthesis regulation conferred by DENR-MCTS1.

On the other hand, the presence of translated uORFs in several core clock genes prompted us to investigate whether DENR-MCTS1 could regulate translation reinitiation within the clock circuitry, and therefore, affect circadian function. Depletion of *Denr* in mouse NIH3T3 fibroblast carrying a circadian reporter led to a circadian period shortening of up to 1.5 hours, suggesting a role of uORF translation and DENR-mediated reinitiation in efficient clock function (Janich et al. [2015], Research Articles page 75, and Chapter 1 of this thesis).

In Castelo-Szekely et al. [in preparation](Research articles, page 167), we therefore aimed to identify genes that require DENR for efficient translation transcriptome-wide in mouse NIH3T3 cells, to systematically characterise uORF/transcript features that confer DENR dependance, and to detect the DENR targets within the core clock that could explain the period phenotype.

Castelo-Szekely et al. (in preparation): Results summary and contribution to the study

In this study (Research articles, page 167), we aimed to identify the endogenous DENR targets in mammalian cells by performing ribosome profiling in NIH3T3 cells upon shRNA-mediated *Denr* depletion. We hypothesized that, if DENR is required for efficient protein synthesis of a subset of uORF-containing genes, DENR targets could be identified from changes in translation efficiency (TE) from the main CDS. Ribosome profiling is thus suitable for these analysis, as it allows both the annotation of translated uORFs and the quantification of TE differences between knockdown and control cells. Moreover, through the use of multiple regression analysis, we aimed to characterise uORF features associated with DENR dependance, expanding the observations from Schleich et al. [2014, 2017].

To explore how the translational landscape changed upon *Denr* knockdown, we quantified relative 5' UTR translation as the ratio of 5' UTR to CDS-mapping reads. We observed an increased relative translation in 5' UTRs of *Denr*-deficient cells and a higher 5' UTR footprint density, suggesting a ribosome redistribution upon *Denr* knockdown, likely due to the presence of uORFs (article's figure 1). Since 5' UTR translation served as a proxy for uORF translation, the translational shift could reflect an increased uORF usage that can be due to impaired leaky scanning or reinitiation.

In order to identify DENR targets based on a change in CDS TE, we used several available algorithms for the quantification of differential TE, and identified ∼200 transcripts with a significant decrease in TE in *Denr*-depleted cells. These targets were enriched for translated-uORF containing transcripts, were associated to functional pathways, such as translation, kinase

activity, and mRNA stability, and included many cancer-related genes (article's figure 2).

DENR dependance was validated through the use of dual reporters in which the 5' UTR of several identified targets was cloned upstream of *Firefly* luciferase. Luciferase activity was thus reduced in the absence of *Denr* and could be rescued by mutating the uORF start codon. Although not initially identified in our analyses, reporter assays also revealed DENR dependance for the mouse homologue of *mbc* identified in Schleich et al. [2014] as a DENR target (article's figure 3).

We next investigated which transcript specificities can confer DENR dependance by modelling the change in CDS TE observed upon *Denr* knockdown on several uORF features that can regulate uORF TE and efficient translation reinitiation. The analysis revealed that the number of uORFs, the presence of AUG uORFs in strong Kozak contexts, and the distance of the last uORF to the start of the main ORF were associated to decreased CDS TE in *Denr*-deficient cells (article's figure 4).

Finally, this study also revealed *Clock* as a target of DENR. *Clock* contains two translated AUG uORFs, the second of which overlaps the CDS. Following the second uORF's stop codon, *Clock* CDS contains a second start site, in frame with the annotated one, thus potentially giving rise to an alternative N' truncated protein, that would be 9 aminoacids shorter, when the overlapping uORF is translated. Through a series of luciferase reporter assays, in which the start codon of the uORFs, main ORF or alternative ORF were systematically mutated, we observed that uORF 2 (the overlapping one), strongly reduces CLOCK protein expression, and was dependent on DENR for initiation on the alternative CDS (article's figure 5). Moreover, mutation of the annotated CDS start codon did not result in a change in protein expression compared to *Clock Wt*, suggesting that this ATG is not being used in NIH3T3 cells, or even,

that the start codon of CLOCK is wrongly annotated. By contrast, mutating the alternative ATG led to a 95% reduction in luciferase signal, consistent with the predominant expression of the shorter isoform.

Contribution to the study

I participated in the study design together with Prof. David Gatfield; performed all experiments, with helpful assistance from Mara de Matos and Angelica Liechti, who carried out all clonings for reporter construction. I analysed the data and wrote the manuscript with input and supervision from Prof. David Gatfield. The manuscript is currently in preparation. The recent finding of *Clock* as a DENR target opened new exciting questions that we are currently addressing experimentally and that we count on adding to the final manuscript (See Discussion).

Discussion

Tissue-specific translational regulation of circadian gene expression

In the first part of this thesis, two fundamental questions of circadian genomics were addressed, namely, its post-transcriptional and tissue-specific regulation (Research articles, pages 75, 121 and 127). Given that proteins are closer to the final functional rhythmic output, quantifying protein synthesis rates around-the-day is of high relevance to understand circadian gene expression, physiology and behaviour.

The analysis of daytime-dependent translation efficiency revealed that 8-9% of transcripts with protein synthesis rhythms showed constant steady-state mRNA levels in liver and kidney, globally corresponding to 1-2% of the transcriptome. Moreover, translational rhythms were largely exclusive to each organ, despite the large overlap in expressed transcriptomes. The latter suggested that, similar to the observed differences in oscillatory transcriptomes [Storch et al., 2002; Panda et al., 2002; Zhang et al., 2014; Mure et al., 2018], tissue-specific factors and different integration of systemic cues (e.g. feeding) largely determine rhythmic protein output.

In liver, an organ with key metabolic functions, biosynthesis of ribosomal proteins (RPs) is rhythmically regulated, gating the high energy consuming process of translation to the time of

the day with highest nutrient availability. Accumulation of RPs during the dark/active phase could result in a global increase in protein synthesis in this phase, which would be missed in our analysis (see below). This hypothesis was recently confirmed in night-time-fed animals, showing that not only total protein content oscillated, but also whole liver size and mass changed accordingly [Sinturel et al., 2017]. This further indicated that the number of ribosomes might be rate-limiting for protein synthesis, through the precise production of RPs and the synthesis of rRNAs in excess [Sinturel et al., 2017]. The opposite, however - an accumulation of RPs during the day in day-time-fed animals - was not observed [Sinturel et al., 2017], suggesting that there is more to it than a response to nutrient availability and that both circadian and metabolic signals could control RP biogenesis. Given that TOR expression itself is clock-regulated in liver [Jouffe et al., 2013], the local clock could, in this scenario, gate nutrient-derived mTOR signalling by controlling its expression. Given its smaller role in energy homeostasis, it is likely that kidney responds less pronouncedly to feeding/fasting signals, therefore showing a dampened or rather constant expression of translational machinery components. In this line, TOR expression and activity also showed reduced-amplitude rhythms in kidney [Jouffe et al., 2013]. Moreover, kidney contains a more heterogeneous cell population, making it plausible that different cell types with specialised functions show cell type-specific translational rhythms, as shown for rhythmic mRNA accumulation as well [Zuber et al., 2009]. Elucidating RP and other cases of translational control observed in liver – including the rhythmic regulation of iron homeostasis through *Fth1*/*Ftl1* – could be the object of future research.

It is noteworthy, that ribosome profiling as it was implemented in these studies would miss global translational changes occurring between time-points, e.g. a global translation upregulation during night-time due to increased RP production. These would be lost upon the

global normalization procedure that assumes that most transcripts do not change their ribosome occupancy, and that is applied between samples (i.e. timepoints) to correct for sequencing depth differences. In future studies, the use of spike-ins (e.g. ribosome footprints derived from mRNA of a different species) and a normalization to this external reference will be useful in cases where a lack of global changes cannot be assumed.

Beyond the contribution of protein synthesis rates to rhythm generation, tissue-specific translation efficiency served to modulate the phase differences between the rhythmic transcriptome and translatome. In kidney but not in liver, maximal protein biosynthesis preceded mRNA level peaks, suggesting that differences in mRNA polyadenylation kinetics and stability across organs can regulate translation efficiency rhythms, as has been shown before in mouse liver alone [Kojima et al., 2012].

Quantification of ribosome footprints on the CDS serves as an estimation of absolute protein production [Li et al., 2014]. Our work therefore allowed the comparison of core clock stoichiometries across organs, and showed that TEs counteracted mRNA levels differences in all but *Clock*, *Bmal1*, *Rev-erb*β, and *Ror*α transcripts. Presumably, rhythmic translation efficiency that is tissue-specifically regulated, serve to modulate core clock components production, which in turn, may adjust organ-specific rhythmicity parameters. For example, since increased PER2 levels give rise to long free-running period [D'Alessandro et al., 2015], the higher PER proteins production in kidney could explain the longer period observed in kidney vs. hepatic explants [Yoo et al., 2004]. In the interconnecting limb too, the higher synthesis of activators (RORs) and lower of repressors (REV-ERBs) observed in liver – the latter potentially via differential uORF translation – were consistent with known differences in the active state of their targets, mainly involved in hepatic metabolism [Fang and Lazar, 2015]. Using protein synthesis rather

than transcript levels as a read-out will be of use to understand the functional properties of the clock circuitry in other organs as well.

Translational contribution to constitutive gene expression divergence between tissues

The transcriptome and translatome data generated in Janich et al. [2015] and Castelo-Szekely et al. [2017] provided insights into constitutive gene expression control as well. Understanding the mechanisms behind tissue-specific regulation of gene expression, and the relative contributions of transcription and translation to define protein levels, is a subject of current research in the gene expression field at large, and an ongoing matter of debate [Vogel and Marcotte, 2012; Silva and Vogel, 2016]. Recent studies have shown that mRNA levels correlated only poorly with protein abundance, indicating post-transcriptional regulation, but while some studies suggested that mRNA-to-protein ratio differs between genes but is constant across tissues [Edfors et al., 2016], others observed a significant inter-organ variation [Franks et al., 2017] suggestive of tissue-specificity as well.

The analysis of differential translation efficiency across organs showed that this was of smaller scale than mRNA abundance differences, similar to the narrower dynamic range of TE observed within tissues. These findings indicated that gene expression divergence is largely defined at the transcriptome level and that protein biosynthesis had a modulatory role. Nevertheless, organ-specific TE, affecting $\sim$9% of the shared transcriptome, served to counteract mRNA level differences leading to a larger similarity in protein production across organs, a phenomenon similar to that reported between yeast [McManus et al., 2014] and rat [Schafer et al., 2015] strains, and in line with a higher evolutionary conservation of the proteome [Schrimpf et al.,

2009; Laurent et al., 2010; Khan et al., 2013]. It it therefore likely that a stronger selection pressure acts to maintain precise protein biosynthesis rates than RNA levels.

Several mechanisms may be responsible for tissue-specific gene regulation post-transcriptionally. In Castelo-Szekely et al. [2017], I observed that transcript diversity that affects the 5' UTR, as well as GC content or 5'UTR length, were associated to TE differences. These results highlight the relevance of translation regulatory mechanisms involving 5' UTRs, and that could include tissue-specific uORF, RBP or translation machinery components expression and usage. In particular, differential uORF translation between organs, as well as differential expression of transcript isoforms containing or lacking uORFs, can thus serve to quickly, reversibly, and cell-specifically regulate protein biosynthesis.

The genome-wide data generated in our studies can serve as a resource for future studies in the field; for example, of the correlation of protein levels to both translational rates and mRNA abundance in different organs, or for a more comprehensive study of uORF usage across organs.

DENR-mediated translation regulation

Translation efficiency can be regulated through both *cis*- and *trans*- acting factors, such as uORFs and RBPs, respectively. uORF usage is pervasive in the mammalian transcriptome, and generally leads to a reduced translation efficiency on the main CDS [Wethmar, 2014]. In the analyses of the liver translatome, we found that also within core clock transcripts, uORF translation was abundant, and that circadian period length in NIH3T3 cells could be regulated by DENR-mediated translation reinitiation [Janich et al., 2015], revealing a novel mechanism of circadian clock regulation. In the second part of this thesis (Research articles, page 167), DENR targets in NIH3T3 cells were identified, including one among core clock genes, gaining

insights into the sequence specificities that confer DENR dependance.

DENR was recently identified as a eukaryotic reinitiation factor [Schleich et al., 2014], and its targets in mammalian cells have been predicted through an *in-silico* approach that used the DENR dependance measured for several synthetic uORF-containing reporters – of varying length or Kozak strength – to extrapolate translational regulation genome-wide [Schleich et al., 2017]. Using ribosome profiling in *Denr*-deficient cells, I quantified CDS TE changes *in vivo*, identifying 240 transcripts that needed DENR for efficient protein production, and that were enriched for uORF-containing genes.

DENR and its heterodimerization partner MCTS1 are oncogenes [Prosniak et al., 1998; Reinert et al., 2006], upregulated, among other, in prostate cancer according to public cancer genomic databases [Cerami et al., 2012]; and many of the identified targets are thus proliferation-related genes, upregulated in several cancer types, or associated to increased risk and poor prognosis. Kelch Domain Containing 8A (KLHDC8A) protein is overexpressed in human gliomas that become resistant to epidermal growth factor receptor (EGFR) silencing, the commonly used therapeutic approach, allowing aggressiveness maintenance in these tumors through an unknown EGFR-independent pathway [Mukasa et al., 2010]. It is thus plausible that in the absence of EGFR dependance, a DENR-mediated translational control comes in place to increase KLHDC8A protein levels. Similarly, Vascular endothelial growth factor D (VEGF-D), or Mitogen-Activated Protein Kinase Kinase 5 (MAP2K5) are upregulated in different cancers [Thelen et al., 2008; Diao et al., 2016]. DENR could therefore be an attractive drug target candidate: decreasing DENR levels could, in turn, reduce the oncogenic levels of the proteins it regulates in proliferating, but not in quiescent, cells.

A comparison of the targets identified in our study and those found previously [Schleich et al.,

2014, 2017] returned little overlap. All but one (*Tmem60*) of the eight targets computationally predicted and validated in HeLa cells by Schleich et al. [2017] were not expressed in NIH3T3 cells. *Tmem60* was reported to result in a 25% downregulation upon *Denr* knockdown, whereas in our system a ∼8% decrease in CDS TE was quantified. Similarly, from the ensemble of 104 targets predicted by Schleich et al. [2017], >90% were not expressed or lacked a mouse orthologue in NIH3T3 cells, and others (e.g. *Lca5*, *Mllt11* or *Dsel*) were regulated to a lesser extent. It is therefore likely that DENR targets and their degree of regulation is cell-type and organism-specific. Moreover, fundamental differences in experimental and analytical approaches could underlie the differences in the translation regulation magnitude observed between studies. Nevertheless, *Dock1*, the mouse orthologue of Drosophila's *mbc* shown in Schleich et al. [2014] to be a strong DENR target in flies, also showed DENR dependance in our reporter assays, although it was not detected in our differential TE analysis, indicating that our list might not be fully exhaustive.

Several transcript/uORF properties can control the efficiency of uORF by-pass and translation reinitiation. The start codon and context can influence the strength of its recognition; sequences upstream, within or downstream a uORF can be recognised by *trans*-acting factors affecting reinitiation [Mohammad et al., 2017]; and uORF length, structural complexity and time required for decoding, as well as the intercistronic distance, can affect reinitiation efficiency [Poyry et al., 2004; Kozak, 1987, 2001]. What rules dictate DENR dependance on certain uORFs but not on others are thus intriguing. Using all annotated uORF-containing transcripts, I identified four prominent features that added predictive value to a comprehensive model of CDS TE regulation exerted by DENR. The number of uORFs, the start codon identity and sequence context, and the distance to the main ORF, correlated with the magnitude of translational regulation. The

presence of several uORFs was already known to lead to a CDS TE downregulation, by requiring successive reinitiation events [Calvo et al., 2009], and DENR regulated transcripts would not be an exception. Similarly, DENR would be required for efficient reinitiation on uORFs located further from the CDS, consistent with the idea that longer distances allow reacquisition of the initiation machinery on time for CDS translation [Kozak, 1987]. Finally, AUG-starting uORFs and those within strong Kozak contexts showed the strongest DENR dependance, as was shown in fly [Schleich et al., 2014] and human cells [Schleich et al., 2017], suggesting a conserved mechanism of action. On the other hand, the length of the uORFs was not correlated with the efficiency of DENR-mediated regulation. Notably, already between Drosophila and human cells, DENR dependance on uORF length was quantitatively different: whereas in S2 cells, up to 4 amino acid-coding uORFs were found to be regulated by DENR [Schleich et al., 2014], uORFs coding for only 1 amino acid were strongly DENR-dependent in HeLa cells [Schleich et al., 2017]. Therefore, the length specificities for DENR regulation could differ between species. Moreover, deriving genome-wide DENR dependance from synthetic uORF-containing reporters [Schleich et al., 2017] could be biased to the results obtained from a selected number of constructs. It is therefore not unlikely that the *in vivo* targets contain uORFs of varying length, since it is the time required for translation rather than the linear length *per se* what determines scanning resumption [Barbosa et al., 2013; Calvo et al., 2009; Poyry et al., 2004].

Translational regulation in uORF-containing genes can occur through several different mechanisms: leaky scanning, in which uORFs are bypassed; reinitiation, in which the ribosome resumes scanning after uORF translation; and other mechanisms involving ribosome stalling and NMD (Wethmar [2014] and Figure 2.1). DENR-MCTS1 can bind the small ribosomal subunit, close to the P-site and mRNA entry channel [Lomakin et al., 2017; Weisser et al.,

2017]. However, the iClip experiments that I carried out in collaboration with Prof. Jernej Ule did not reveal DENR crosslinking to mRNAs (data not shown), suggesting that DENR travels on the ribosome or transiently interacts with it, although recognition and binding to specific motifs around uORFs could be possible through a so far unknown factor. The presence of DENR-MCTS1 potentially clashes sterically with that of other canonical translation initiation factors, e.g. certain eIF3 subunits [Lomakin et al., 2017; Weisser et al., 2017]. Given that eIF3 is implicated in translation termination and ribosome recycling [Beznoskova et al., 2013], it is plausible that DENR-MCTS1 recruitment to the ribosome displaces eIF3 to then promote reinitiation on a downstream ORF through DENR's SUI motif, involved in start codon recognition [Yoon and Donahue, 1992]. Alternatively, DENR-MCTS1 could compete with, and replace, eIF1 for cap-dependent scanning, given that simultaneous DENR-MCTS1 and eIF2-tRNAi presence on the ribosome is not incompatible [Lomakin et al., 2017]. Overexpression of DENR in proliferating cells would shift the balance to more DENR-MCTS1-containing ribosomes, leading to increased CDS translation through skipping of strong uORF that would normally be translated in a eIF1-containing ribosome. Discerning between leaky scanning and reinitiation would gain insights into the precise mechanism of action of DENR-MCTS1. Translation complex profile sequencing (TCP-seq), a newly developed technique derived from ribosome profiling, could be used to monitor the translation dynamics of scanning and terminating ribosomes [Archer et al., 2016]. TCP-seq allows the recovery of small ribosomal subunit footprints – which differ in length from those captured by 80S monosomes – around both the uORF start and stop codons, and could thus be used to elucidate the mode of action of DENR-MCTS1 and other translation (re)initiation factors.

The structural and domain organisation similarity between the DENR-MCTS1 heterodimer

and eIF2D (Skabkin et al. [2010]; Lomakin et al. [2017]; Weisser et al. [2017] and Figure 2.2) suggests similar modes of regulation. However it is still unknown whether both factors regulate the same or different sets of transcripts and whether they act in different cell types/cellular contexts, what are the sequence specificities of the transcripts and uORFs regulated by each factor, and whether their action confers a similar degree of protein synthesis regulation to their targets. It would thus be of interest to carry out a comparative study between both factors, following an approach analogous to this study on eIF2D. Interestingly, Fijalkowska et al. [2017] performed ribosome profiling on *Eif1*-depleted human cells and characterised uORF features associated to eIF1-mediated translational control. Like DENR, eIF1 contains a SUI domain involved in start codon recognition [Kasperaitis et al., 1995], and previous *in vitro* studies showed that both factors can promote reinitiation [Skabkin et al., 2010, 2013]. However, in contrast to DENR, which is necessary for efficient translation of strong AUG uORF-containing mRNAs, eIF1 regulated transcripts with near-AUG uORFs, indicating that, despite the protein domain similarity, these two factors are involved in reinitiation on distinct classes of uORFs.

CLOCK protein isoforms generated through reinitiation

Among core clock transcripts, *Clock* was identified as a DENR target, showing a ∼20% reduction in protein synthesis rate in *Denr*-depleted cells. We annotated two highly translated AUG-initiated uORFs on *Clock*'s 5' UTR, the second of which overlaps with the main CDS. The presence of an additional AUG, downstream of the second uORF's stop codon and in frame with the annotated CDS start site, could give rise to a 5' truncated CLOCK protein isoform, that would be 9 amino acids shorter (Figure 3.1). A series of systematic mutations suggested that DENR was necessary for reinitiation on the alternative CDS after translation of the overlapping

uORF, revealing that alternative translation initiation could give rise to two CLOCK protein isoforms. Nevertheless, our results suggested that the currently annotated, longer CLOCK isoform, is not highly expressed in NIH3T3 cells, and could even be missannotated, taking into account our and published ribosome profiling data (see article's Supplementary Figure 8). In the future, it will be exciting to investigate whether both isoforms can be detected *in vivo*, by for example, N-terminal proteomic experiments in mouse liver.

The N-terminal portion of CLOCK contains the bHLH domain, which together with its counterpart in BMAL1, is involved in DNA binding. Mutations affecting the heterodimer interface have been shown to impair CLOCK:BMAL1 stability and binding activity and lead to altered circadian period phenotypes [Huang et al., 2012]. Since the alternative CLOCK protein identified here would only lack the first nine amino acids of the full CLOCK isoform, it is likely that the bHLH domain itself (starting at aa-26) would remain intact (Figure 3.1). However, N-terminal motifs upstream of the DNA-binding domains in other bHLH transcription factors have also been shown to be required for cooperative DNA binding [Knoepfler et al., 1999]. It is therefore tempting to speculate that long and short forms of CLOCK protein could have distinct DNA binding properties, affecting transcriptional activation, and thus giving functional differences in the clock.

We had previously observed that *Denr*-deficient cells displayed a shorter circadian period (Janich et al. [2015] and Chapter 1 of this thesis). Although we cannot conclude that reduced CLOCK protein synthesis underlies this phenotype, it is reassuring that Clock$^{-}/^{-}$ fibroblasts also showed a similar period shortening [Landgraf et al., 2016]. *In vivo* mutation of the overlapping uORF through the CRISPR-Cas system could be useful to further assess whether uORF translation is relevant for circadian function.

MVFTVSCSKM ***SSIVDRDDSS*** ***IFDGLVEEDD*** ***KDK***AKRVSRN KSEKKRDQF NVLIKELGSM 60
LPGNARKMDK STVLQKSIDF LRKHKETTAQ SDASEIRQDW KPTFLSNEEF TQLMLEALDG 120
FFLAIMTDGS IIYVSESVTS LLEHLPSDLV DQSIFNFIPE GEHSEVYKIL STHLLESDSL 180
TPEYLKSKNQ LEFCCHMLRG TIDPKEPSTY EYVRFIGNFK SLTSVSTSTH NGFEGTIQRT 240
HRPSYEDRVC FVATVRLATP QFIKEMCTVE EPNEEFTSRH SLEWKFLFLD HRAPPIIGYL 300
PFEVLGTSGY DYYHVDDLEN LAKCHEHLMQ YGKGKSCYYR FLTKGQQWIW LQTHYYITYH 360
QWNSRPEFIV CTHTVVSYAE VRAERRRELG IEESLPETAA DKSQDSGSDN RINTVSLKEA 420
LERFDHSPTP SASSRSSRKS SHTAVSDPSS TPTKIPTDTS TPPRQHLPAH EKMTQRRSSF 480
SSQSINSQSV GPSLTQPAMS QAANLPIPQG MSQFQFSAQL GAMQHLKDQL EQRTRMIEAN 540
IHRQQEELRK IQEQLQMVHG QGLQMFLQQS NPGLNFGSVQ LSSGNSNIQQ LTPVNMQGQV 600
VPANQVQSGH ISTGQHMIQQ QTLQSTSTQQ SQQSVMSGHS QQTSLPSQTP STLTAPLYNT 660
MVISQPAAGS MVQIPSSMPQ NSTQSATVTT FTQDRQIRFS QGQQLVTKLV TAPVACGAVM 720
VPSTMLMGQV VTAYPTFATQ QQQAQTLSVT QQQQQQQQQP PQQQQQQQQS SQEQQLPSVQ 780
QPAQAQLGQP PQQFLQTSRL LHGNPSTQLI LSAAFPLQQS TFPPSHHQQH QPQQQQQLPR 840
HRTDSLTDPS KVQPQ I

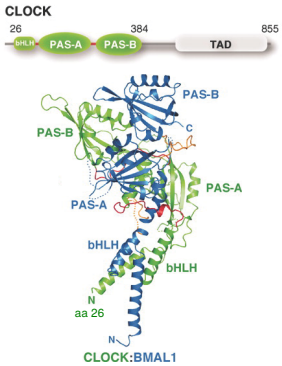


Figure 3.1: CLOCK sequence and structure. On top, CLOCK amino acid sequence, with both the annotated and alternative Methionines in green. Highlighted in bold are the peptides detected in the circadian nuclear proteome study by Wang et al. [2017], although sequence coverage was fairly limited (38.42%). On the bottom, CLOCK protein domain organisation and CLOCK:BMAL1 heterodimer structure (from Huang et al. [2012]). CLOCK crystal structure was resolved starting from amino acid 26, which lies in the bHLH domain that interacts with DNA.

The presence of several uORFs, one of them overlapping the CDS, is reminiscent of the uORF arrangement observed in *Atf4*. Protein synthesis in *Atf4* follows an on/off regulatory mode, in which translation is abolished in normal cellular homeostasis, because translation on the overlapping uORF avoids ribosomal assembly on the CDS start. During nutrient starvation, however, when levels of ternary complex are low, the ribosome fails to assemble on the overlapping uORF after translation of the first uORF, thus allowing ATF4 protein production [Vattem and Wek, 2004]. It will thus be interesting to investigate whether *Clock* uORFs are responsive to different stress signals, such as glucose, serum or oxygen deprivation, oxydative stress, etc, and whether DENR is involved in the stress response to shift the balance between the different CLOCK isoforms.

Bibliography

Akashi, M., Tsuchiya, Y., Yoshino, T., and Nishida, E. (2002). Control of intracellular dynamics of mammalian period proteins by casein kinase I epsilon (CKIepsilon) and CKIdelta in cultured cells. *Molecular and Cellular Biology*, 22(6):1693–1703.

Archer, S. K., Shirokikh, N. E., Beilharz, T. H., and Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature*, advance online publication.

Arribere, J. A. and Gilbert, W. V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Research*, 23(6):977–987.

Atger, F., Gobet, C., Marquis, J., Martin, E., Wang, J., Weger, B., Lefebvre, G., Descombes, P., Naef, F., and Gachon, F. (2015). Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. *PNAS*, 112(47):E6579–E6588.

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

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., and Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO Journal*.

Bazzini, A. A., Lee, M. T., and Giraldez, A. J. (2012). Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science*, 336(6078):233–237.

Behm-Ansmant, I. and Izaurralde, E. (2006). Quality control of gene expression: a stepwise assembly pathway for the surveillance complex that triggers nonsense-mediated mRNA decay. *Genes & Development*, 20(4):391–398.

Beznoskova, P., Cuchalova, L., Wagner, S., Shoemaker, C. J., Gunisova, S., von der Haar, T., and Valasek, L. S. (2013). Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. *PLoS genetics*, 9(11):e1003962.

Bonny, O., Vinciguerra, M., Gumz, M. L., and Mazzoccoli, G. (2013). Molecular bases of circadian rhythmicity in renal physiology and pathology. *Nephrol. Dial. Transplant.*, 28(10):2421–2431.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grutzner, F., Bergmann, S., Nielsen, R., Paabo, S., and Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369):343–348.

Bunger, M. K., Wilsbacher, L. D., Moran, S. M., Clendenin, C., Radcliffe, L. A., Hogenesch, J. B., Simon, M. C., Takahashi, J. S., and Bradfield, C. A. (2000). Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell*, 103(7):1009–1017.

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.

Cardone, L., Hirayama, J., Giordano, F., Tamaru, T., Palvimo, J. J., and Sassone-Corsi, P. (2005). Circadian clock control by SUMOylation of BMAL1. *Science (New York, N.Y.)*, 309(5739):1390–1394.

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

Chiang, C.-K., Mehta, N., Patel, A., Zhang, P., Ning, Z., Mayne, J., Sun, W. Y. L., Cheng, H.-Y. M., and Figeys, D. (2014). The Proteomic Landscape of the Suprachiasmatic Nucleus Clock Reveals Large-Scale Coordination of Key Biological Processes. *PLOS Genetics*, 10(10):e1004695.

Churbanov, A., Rogozin, I. B., Babenko, V. N., Ali, H., and Koonin, E. V. (2005). Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes. *Nucleic Acids Research*, 33(17):5512–5520.

D'Alessandro, M., Beesley, S., Kim, J. K., Chen, R., Abich, E., Cheng, W., Yi, P., Takahashi, J. S., and Lee, C. (2015). A tunable artificial circadian clock in clock-defective mice. *Nature Communications*, 6:8587.

Dever, T. E. and Green, R. (2012). The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 4(7):a013706.

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

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

Dmitriev, S. E., Terenin, I. M., Andreev, D. E., Ivanov, P. A., Dunaevsky, J. E., Merrick, W. C., and Shatsky, I. N. (2010). GTP-independent tRNA Delivery to the Ribosomal P-site by a Novel Eukaryotic Translation Factor. *Journal of Biological Chemistry*, 285(35):26779–26787.

Doherty, C. J. and Kay, S. A. (2010). Circadian Control of Global Gene Expression Patterns. *Annual review of genetics*, 44:419–444.

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.

Dunn, J. G., Foo, C. K., Belletier, N. G., Gavis, E. R., and Weissman, J. S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. *eLife*, 2.

Edfors, F., Danielsson, F., Hallstrom, B. M., Kall, L., Lundberg, E., Ponten, F., Forsstrom, B., and Uhlen, M. (2016). Geneâspecific correlation of RNA and protein levels in human cells and tissues. *Molecular Systems Biology*, 12(10):883.

Eide, E. J., Vielhaber, E. L., Hinz, W. A., and Virshup, D. M. (2002). The circadian regulatory proteins BMAL1 and cryptochromes are substrates of casein kinase Iepsilon. *The Journal of Biological Chemistry*, 277(19):17248–17254.

Fang, B. and Lazar, M. A. (2015). Dissecting the Rev-erba Cistrome and the Mechanisms Controlling Circadian Transcription in Liver. *Cold Spring Harbor Symposia on Quantitative Biology*, 80:233–238.

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.

Franks, A., Airoldi, E., and Slavov, N. (2017). Post-transcriptional regulation across human tissues. *PLOS Computational Biology*, 13(5):e1005535.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1):92–105.

Gekakis, N., Staknis, D., Nguyen, H. B., Davis, F. C., Wilsbacher, L. D., King, D. P., Takahashi, J. S., and Weitz, C. J. (1998). Role of the CLOCK protein in the mammalian circadian mechanism. *Science (New York, N.Y.)*, 280(5369):1564–1569.

Green, C. B., Douris, N., Kojima, S., Strayer, C. A., Fogerty, J., Lourim, D., Keller, S. R., and Besharse, J. C. (2007). Loss of Nocturnin, a circadian deadenylase, confers resistance to hepatic steatosis and diet-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23):9888–9893.

Guillaumond, F., Dardente, H., Giguere, V., and Cermakian, N. (2005). Differential control of Bmal1 circadian transcription by REV-ERB and ROR nuclear receptors. *Journal of Biological Rhythms*, 20(5):391–403.

Guttman, M., Russell, P., Ingolia, N., Weissman, J., and Lander, E. (2013). Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell*, 154(1):240–251.

Hamilton, T. L., Stoneley, M., Spriggs, K. A., and Bushell, M. (2006). TOPs and their regulation. *Biochemical Society Transactions*, 34(1):12–16.

Hinnebusch, A. G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology*, 59:407–450.

Hinnebusch, A. G. (2011). Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiology and Molecular Biology Reviews*, 75(3):434–467.

Hirayama, J., Sahar, S., Grimaldi, B., Tamaru, T., Takamatsu, K., Nakahata, Y., and Sassone-Corsi, P. (2007). CLOCK-mediated acetylation of BMAL1 controls circadian function. *Nature*, 450(7172):1086–1090.

Hsu, H.-L., Shi, B., and Gartenhaus, R. B. (2005). The MCT-1 oncogene product impairs cell cycle checkpoint control and transforms human mammary epithelial cells. *Oncogene*, 24(31):4956–4964.

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.

Iacono, M., Mignone, F., and Pesole, G. (2005). uAUG and uORFs in human and rodent 5 untranslated mRNAs. *Gene*, 349:97–105.

Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics*, 15(3):205–213.

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., and Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*, 7(8):1534–1550.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, 324(5924):218–223.

Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4):789–802.

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., and Gatfield, D. (2016). Analyzing the temporal regulation of translation efficiency in mouse liver. *Genomics Data*, 8:41–44.

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.

Johnstone, T. G., Bazzini, A. A., and Giraldez, A. J. (2016). Upstream ORFs are prevalent translational repressors in vertebrates. *The EMBO Journal*.

Jouffe, C., Cretenet, G., Symul, L., Martin, E., Atger, F., Naef, F., and Gachon, F. (2013). The Circadian Clock Coordinates Ribosome Biogenesis. *PLoS Biology*, 11(1).

Kanaji, T., Okamura, T., Osaki, K., Kuroiwa, M., Shimoda, K., Hamasaki, N., and Niho, Y. (1998). A Common Genetic Polymorphism (46 C to T Substitution) in the 5'-Untranslated Region of the Coagulation Factor XII Gene Is Associated With Low Translation Efficiency and Decrease in Plasma Factor XII Level. *Blood*, 91(6):2010–2014.

Kasperaitis, M. A., Voorma, H. O., and Thomas, A. A. (1995). The amino acid sequence of eukaryotic translation initiation factor 1 and its similarity to yeast initiation factor SUI1. *FEBS letters*, 365(1):47–50.

Khan, Z., Ford, M. J., Cusanovich, D. A., Mitrano, A., Pritchard, J. K., and Gilad, Y. (2013). Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures. *Science*, 342(6162):1100–1104.

Kim, D.-Y., Woo, K.-C., Lee, K.-H., Kim, T.-D., and Kim, K.-T. (2010). hnRNP Q and PTB modulate the circadian oscillation of mouse Rev-erb a via IRES-mediated translation. *Nucleic Acids Research*, 38(20):7068–7078.

King, D. P., Zhao, Y., Sangoram, A. M., Wilsbacher, L. D., Tanaka, M., Antoch, M. P., Steeves, T. D., Vitaterna, M. H., Kornhauser, J. M., Lowrey, P. L., Turek, F. W., and Takahashi, J. S. (1997). Positional cloning of the mouse circadian clock gene. *Cell*, 89(4):641–653.

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.

Ko, C. H. and Takahashi, J. S. (2006). Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, 15(suppl 2):R271–R277.

Koike, N., Yoo, S.-H., Huang, H.-C., Kumar, V., Lee, C., Kim, T.-K., and Takahashi, J. S. (2012). Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science (New York, N.Y.)*, 338(6105):349–354.

Kojima, S., Sher-Chen, E. L., and Green, C. B. (2012). Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes & Development*, 26(24):2724–2736.

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.

Kume, K., Zylka, M. J., Sriram, S., Shearman, L. P., Weaver, D. R., Jin, X., Maywood, E. S., Hastings, M. H., and Reppert, S. M. (1999). mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell*, 98(2):193–205.

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.

Lareau, L. F., Hite, D. H., Hogan, G. J., and Brown, P. O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *eLife*, 3:e01257.

Laurent, J. M., Vogel, C., Kwon, T., Craig, S. A., Boutz, D. R., Huse, H. K., Nozue, K., Walia, H., Whiteley, M., Ronald, P. C., and Marcotte, E. M. (2010). Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics*, 10(23):4209–4212.

Le Martelot, G., Canella, D., Symul, L., Migliavacca, E., Gilardi, F., Liechti, R., Martin, O., Harshman, K., Delorenzi, M., Desvergne, B., Herr, W., Deplancke, B., Schibler, U., Rougemont, J., Guex, N., Hernandez, N., and Naef, F. (2012). Genome-Wide RNA Polymerase II Profiles and RNA Accumulation Reveal Kinetics of Transcription and Associated Epigenetic Changes During Diurnal Cycles. *PLoS Biology*, 10(11).

Lee, K.-H., Woo, K.-C., Kim, D.-Y., Kim, T.-D., Shin, J., Park, S. M., Jang, S. K., and Kim, K.-T. (2012a). Rhythmic Interaction between Period1 mRNA and hnRNP Q Leads to Circadian Time-Dependent Translation. *Molecular and Cellular Biology*, 32(3):717–728.

Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., and Qian, S.-B. (2012b). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proceedings of the National Academy of Sciences*, 109(37):E2424–E2432.

Levenson, A. S., Thurn, K. E., Simons, L. A., Veliceasa, D., Jarrett, J., Osipo, C., Jordan, V. C., Volpert, O. V., Satcher, R. L., and Gartenhaus, R. B. (2005). MCT-1 Oncogene Contributes to Increased In vivo Tumorigenicity of MCF7 Cells by Promotion of Angiogenesis and Inhibition of Apoptosis. *Cancer Research*, 65(23):10651–10656.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J. (2014). Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell*, 157(3):624–635.

Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., and Yan, J. (2013). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Scientific Reports*, 3.

Lomakin, I. B., Stolboushkina, E. A., Vaidya, A. T., Zhao, C., Garber, M. B., Dmitriev, S. E., and Steitz, T. A. (2017). Crystal Structure of the Human Ribosome in Complex with DENR-MCT-1. *Cell Reports*, 20(3):521–528.

Lowrey, P. L., Shimomura, K., Antoch, M. P., Yamazaki, S., Zemenides, P. D., Ralph, M. R., Menaker, M., and Takahashi, J. S. (2000). Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science (New York, N.Y.)*, 288(5465):483–492.

Luck, S., Thurley, K., Thaben, P. F., and Westermark, P. O. (2014). Rhythmic Degradation Explains and Unifies Circadian Transcriptome and Proteome Data. *Cell Reports*, 9(2):741–751.

Matsui, M., Yachie, N., Okada, Y., Saito, R., and Tomita, M. (2007). Bioinformatic analysis of post-transcriptional regulation by uORF in human and mouse. *FEBS letters*, 581(22):4184–4188.

Mauvoisin, D., Wang, J., Jouffe, C., Martin, E., Atger, F., Waridel, P., Quadroni, M., Gachon, F., and Naef, F. (2014). Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1):167–172.

McGlincy, N. J., Valomon, A., Chesham, J. E., Maywood, E. S., Hastings, M. H., and Ule, J. (2012). Regulation of alternative splicing by the circadian clock and food related cues. *Genome Biology*, 13:R54.

McManus, C. J., May, G. E., Spealman, P., and Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Research*, 24(3):422–430.

Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F., and Dietz, H. C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature Genetics*, 36(10):1073–1078.

Menet, J. S., Rodriguez, J., Abruzzi, K. C., and Rosbash, M. (2012). Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife*, 1.

Meyuhas, O. (2000). Synthesis of the translational apparatus is regulated at the translational level. *European Journal of Biochemistry*, 267(21):6321–6330.

Meyuhas, O. and Dreazen, A. (2009). Ribosomal protein S6 kinase from TOP mRNAs to cell size. *Progress in Molecular Biology and Translational Science*, 90:109–153.

MGlincy, N. J. and Ingolia, N. T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. *Methods*, 126(Supplement C):112–129.

Mohammad, M. P., MunzarovaÂ Pondelickova, V., Zeman, J., Gunisova, S., and Valasek, L. S. (2017). In vivo evidence that eIF3 stays bound to ribosomes elongating and terminating on short upstream ORFs to promote reinitiation. *Nucleic Acids Research*, 45(5):2658–2674.

Mohawk, J. A., Green, C. B., and Takahashi, J. S. (2012). CENTRAL AND PERIPHERAL CIRCADIAN CLOCKS IN MAMMALS. *Annual review of neuroscience*, 35:445–462.

Morf, J., Rey, G., Schneider, K., Stratmann, M., Fujita, J., Naef, F., and Schibler, U. (2012). Cold-Inducible RNA-Binding Protein Modulates Circadian Gene Expression Posttranscriptionally. *Science*, 338(6105):379–383.

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.

Mure, L. S., Le, H. D., Benegiamo, G., Chang, M. W., Rios, L., Jillani, N., Ngotho, M., Kariuki, T., Dkhissi-Benyahya, O., Cooper, H. M., and Panda, S. (2018). Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science*, page eaao0318.

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.

Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., Schultz, P. G., Kay, S. A., Takahashi, J. S., and Hogenesch, J. B. (2002). Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock. *Cell*, 109(3):307–320.

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.

Preitner, N., Damiola, F., Luis-Lopez-Molina, Zakany, J., Duboule, D., Albrecht, U., and Schibler, U. (2002). The Orphan Nuclear Receptor REV-ERBa Controls Circadian Transcription within the Positive Limb of the Mammalian Circadian Oscillator. *Cell*, 110(2):251–260.

Prosniak, M., Dierov, J., Okami, K., Tilton, B., Jameson, B., Sawaya, B. E., and Gartenhaus, R. B. (1998). A Novel Candidate Oncogene, MCT-1, Is Involved in Cell Cycle Progression. *Cancer Research*, 58(19):4233–4237.

Reddy, A. B., Karp, N. A., Maywood, E. S., Sage, E. A., Deery, M., O'Neill, J. S., Wong, G. K. Y., Chesham, J., Odell, M., Lilley, K. S., Kyriacou, C. P., and Hastings, M. H. (2006). Circadian Orchestration of the Hepatic Proteome. *Current Biology*, 16(11):1107–1115.

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.

Robles, M. S., Cox, J., and Mann, M. (2014). In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. *PLoS Genetics*, 10(1).

Rooijers, K., Loayza-Puch, F., Nijtmans, L. G., and Agami, R. (2013). Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nature Communications*, 4.

Sato, T. K., Panda, S., Miraglia, L. J., Reyes, T. M., Rudic, R. D., McNamara, P., Naik, K. A., FitzGerald, G. A., Kay, S. A., and Hogenesch, J. B. (2004). A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron*, 43(4):527–537.

Schafer, S., Adami, E., Heinig, M., Rodrigues, K. E. C., Kreuchwig, F., Silhavy, J., van Heesch, S., Simaite, D., Rajewsky, N., Cuppen, E., Pravenec, M., Vingron, M., Cook, S. A., and Hubner, N. (2015). Translational regulation shapes the molecular landscape of complex disease phenotypes. *Nature Communications*, 6.

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.

Schrimpf, S. P., Weiss, M., Reiter, L., Ahrens, C. H., Jovanovic, M., Malmstrom, J., Brunner, E., Mohanty, S., Lercher, M. J., Hunziker, P. E., Aebersold, R., Mering, C. v., and Hengartner, M. O. (2009). Comparative Functional Analysis of the Caenorhabditis elegans and Drosophila melanogaster Proteomes. *PLOS Biology*, 7(3):e1000048.

Silva, G. M. and Vogel, C. (2016). Quantifying gene expression: the importance of being subtle. *Molecular Systems Biology*, 12(10):885.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A Competition between Stimulators and Antagonists of Upf Complex Recruitment Governs Human Nonsense-Mediated mRNA Decay. *PLOS Biology*, 6(4):e111.

Sinturel, F., Gerber, A., Mauvoisin, D., Wang, J., Gatfield, D., Stubblefield, J. J., Green, C. B., Gachon, F., and Schibler, U. (2017). Diurnal Oscillations in Liver Mass and Cell Size Accompany Ribosome Assembly Cycles. *Cell*, 169(4):651–663.e14.

Skabkin, M., Skabkina, O., Hellen, C. T., and Pestova, T. (2013). Reinitiation and Other Unconventional Posttermination Events during Eukaryotic Translation. *Molecular Cell*, 51(2):249–264.

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.

Sonenberg, N. and Hinnebusch, A. G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*, 136(4):731–745.

Storch, K.-F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F. C., Wong, W. H., and Weitz, C. J. (2002). Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417(6884):78–83.

Stumpf, C. R., Moreno, M. V., Olshen, A. B., Taylor, B. S., and Ruggero, D. (2013). The Translational Landscape of the Mammalian Cell Cycle. *Molecular Cell*, 52(4):574–582.

Tanenbaum, M. E., Stern-Ginossar, N., Weissman, J. S., and Vale, R. D. (2015). Regulation of mRNA translation during mitosis. *eLife*, 4:e07957.

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

Thoreen, C. C., Chantranupong, L., Keys, H. R., Wang, T., Gray, N. S., and Sabatini, D. M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*, 485(7396):109–113.

Toh, K. L., Jones, C. R., He, Y., Eide, E. J., Hinz, W. A., Virshup, D. M., Ptacek, L. J., and Fu, Y. H. (2001). An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science (New York, N.Y.)*, 291(5506):1040–1043.

Vattem, K. M. and Wek, R. C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences*, 101(31):11269–11274.

Vogel, C. and Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13(4):227–232.

Wang, J., Mauvoisin, D., Martin, E., Atger, F., Galindo, A. N., Dayon, L., Sizzano, F., Palini, A., Kussmann, M., Waridel, P., Quadroni, M., Dulic, V., Naef, F., and Gachon, F. (2017). Nuclear Proteomics Uncovers Diurnal Regulatory Landscapes in Mouse Liver. *Cell Metabolism*, 25(1):102–117.

Wang, Y., Osterbur, D. L., Megaw, P. L., Tosini, G., Fukuhara, C., Green, C. B., and Besharse, J. C. (2001). Rhythmic expression of Nocturnin mRNA in multiple tissues of the mouse. *BMC Developmental Biology*, 1:9.

Weisser, M., Schafer, T., Leibundgut, M., Bohringer, D., Aylett, C. H. S., and Ban, N. (2017). Structural and Functional Insights into Human Re-initiation Complexes. *Molecular Cell*, 67(3):447–456.e7.

Wethmar, K. (2014). The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley interdisciplinary reviews. RNA*, 5(6):765–778.

Wethmar, K., Barbosa-Silva, A., Andrade-Navarro, M. A., and Leutz, A. (2014). uORFdb– a comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Research*, 42(D1):D60–D67.

Wolin, S. L. and Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *The EMBO journal*, 7(11):3559–3569.

Woo, K.-C., Ha, D.-C., Lee, K.-H., Kim, D.-Y., Kim, T.-D., and Kim, K.-T. (2010). Circadian Amplitude of Cryptochrome 1 Is Modulated by mRNA Stability Regulation via Cytoplasmic hnRNP D Oscillation. *Molecular and Cellular Biology*, 30(1):197–205.

Woo, K.-C., Kim, T.-D., Lee, K.-H., Kim, D.-Y., Kim, W., Lee, K.-Y., and Kim, K.-T. (2009). Mouse period 2 mRNA circadian oscillation is modulated by PTB mediated rhythmic mRNA degradation. *Nucleic Acids Research*, 37(1):26–37.

Xu, Y., Padiath, Q. S., Shapiro, R. E., Jones, C. R., Wu, S. C., Saigoh, N., Saigoh, K., Ptacek, L. J., and Fu, Y.-H. (2005). Functional consequences of a CKIdelta mutation causing familial advanced sleep phase syndrome. *Nature*, 434(7033):640–644.

Yoo, S.-H., Yamazaki, S., Lowrey, P. L., Shimomura, K., Ko, C. H., Buhr, E. D., Siepka, S. M., Hong, H.-K., Oh, W. J., Yoo, O. J., Menaker, M., and Takahashi, J. S. (2004). PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proceedings of the National Academy of Sciences of the United States of America*, 101(15):5339–5346.

Yoon, H. J. and Donahue, T. F. (1992). The suil suppressor locus in Saccharomyces cerevisiae encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. *Molecular and Cellular Biology*, 12(1):248–260.

Yoshitane, H., Takao, T., Satomi, Y., Du, N.-H., Okano, T., and Fukada, Y. (2009). Roles of CLOCK phosphorylation in suppression of E-box-dependent transcription. *Molecular and Cellular Biology*, 29(13):3675–3686.

Zhang, R., Lahens, N. F., Ballance, H. I., Hughes, M. E., and Hogenesch, J. B. (2014). A circadian gene expression atlas in mammals: Implications for biology and medicine. *PNAS*, 111(45):16219–16224.

Zuber, A. M., Centeno, G., Pradervand, S., Nikolaeva, S., Maquelin, L., Cardinaux, L., Bonny, O., and Firsov, D. (2009). Molecular clock is involved in predictive circadian adjustment of renal function. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38):16523–16528.

Research articles

1. Peggy Janich, Alaaddin Bulak Arpat, **Violeta Castelo-Szekely**, Maykel Lopes and David Gatfield. *Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames.* Genome Research (2015) 25:1848-59.

2. Peggy Janich, Alaaddin Bulak Arpat, **Violeta Castelo-Szekely** and David Gatfield. *Analyzing the temporal regulation of translation efficiency in mouse liver.* Genomics Data (2016) 8:41-4.

3. **Violeta Castelo-Szekely**, Alaaddin Bulak Arpat, Peggy Janich and David Gatfield. *Translational contributions to tissue specificity in rhythmic and constitutive gene expression.* Genome Biology (2017) 18:116.

4. **Violeta Castelo-Szekely**, Mara De Matos and David Gatfield. *uORF-mediated regulation of CLOCK protein abundance identified by translatome-wide mapping of targets of reinitation factor DENR.* [In preparation].

Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames

Peggy Janich, Alaaddin Bulak Arpat, Violeta Castelo-Szekely, Maykel Lopes and David Gatfield

Genome Research (2015) 25:1848-59

Research

Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames

Peggy Janich,[1,3] Alaaddin Bulak Arpat,[1,2,3] Violeta Castelo-Szekely,[1] Maykel Lopes,[1] and David Gatfield[1]

[1]*Center for Integrative Genomics, Génopode, University of Lausanne, 1015 Lausanne, Switzerland;* [2]*Vital-IT, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland*

Mammalian gene expression displays widespread circadian oscillations. Rhythmic transcription underlies the core clock mechanism, but it cannot explain numerous observations made at the level of protein rhythmicity. We have used ribosome profiling in mouse liver to measure the translation of mRNAs into protein around the clock and at high temporal and nucleotide resolution. We discovered, transcriptome-wide, extensive rhythms in ribosome occupancy and identified a core set of approximately 150 mRNAs subject to particularly robust daily changes in translation efficiency. Cycling proteins produced from nonoscillating transcripts revealed thus-far-unknown rhythmic regulation associated with specific pathways (notably in iron metabolism, through the rhythmic translation of transcripts containing iron responsive elements), and indicated feedback to the rhythmic transcriptome through novel rhythmic transcription factors. Moreover, estimates of relative levels of core clock protein biosynthesis that we deduced from the data explained known features of the circadian clock better than did mRNA expression alone. Finally, we identified uORF translation as a novel regulatory mechanism within the clock circuitry. Consistent with the occurrence of translated uORFs in several core clock transcripts, loss-of-function of *Denr*, a known regulator of reinitiation after uORF usage and of ribosome recycling, led to circadian period shortening in cells. In summary, our data offer a framework for understanding the dynamics of translational regulation, circadian gene expression, and metabolic control in a solid mammalian organ.

[Supplemental material is available for this article.]

The mammalian circadian system consists of a master pacemaker in the brain's suprachiasmatic nuclei (SCN) that synchronizes subsidiary oscillators present in most cell types. In the liver and other organs, up to 15% of gene expression shows daily oscillations that are driven directly by local clocks, or by systemic signals such as feeding and body-temperature rhythms (Vollmers et al. 2009; Mohawk et al. 2012; Zhang et al. 2014). Of the molecular mechanisms potentially accounting for rhythmic gene expression, transcription has been extensively studied, notably within the core clock circuitry consisting of transcriptional activators (mainly CLOCK; ARNTL/BMAL1; RORA, RORB, RORC) and repressors (mainly PER1, 2; CRY1, 2; NR1D1/REV-ERB alpha; NR1D2/REV-ERB beta). Their interactions in negative feedback loops generate transcriptional oscillations not only of clock genes, but genome-wide (Mohawk et al. 2012), which has led to the view that transcription represents the dominant driver of gene expression rhythms. However, post-transcriptional mechanisms likely contribute as well. In extension to earlier work showing poor overlap between liver proteome and transcriptome rhythms (Reddy et al. 2006), two recent studies have indicated that 20% (Robles et al. 2014) to 50% (Mauvoisin et al. 2014) of cyclically accumulating proteins are expressed from nonoscillating mRNAs. Conceivably, these protein rhythms are generated at the level of translation and/or protein stability. In other fields, translational regulation is emerging as key to understanding the overall moderate correlations between mRNA and protein abundances (Vogel and Marcotte 2012); a role in rhythmic gene expression is thus conceivable as well.

Time of day–dependent translation is not unprecedented in mammals. Transcripts encoding ribosomal proteins (RPs) associate with polysomes preferentially at the beginning of the night, coincident with feeding time (Jouffe et al. 2013). These mRNAs contain 5′-terminal oligopyrimidine (5′-TOP) motifs that are regulated by the nutrient-sensitive mammalian target of rapamycin complex 1 (TORC1) pathway (Meyuhas and Kahan 2015). Another documented mechanism for rhythmic translation involves daily dynamics in poly(A) tail length and the rhythmic activity of cytoplasmic polyadenylation element-binding proteins (CPEBs) (Kojima et al. 2012). Despite such individual examples, a comprehensive and quantitative analysis of rhythmic translation from a mammalian organ is still lacking.

We have used ribosome profiling (RPF-seq), a method based on the massively parallel sequencing of ribosome-protected mRNA footprints (Ingolia et al. 2009), to determine the positions of translating ribosomes transcriptome-wide and to establish a quantitative, high-resolution map of the mouse liver translatome around the clock. From RPF-seq and matching whole-transcriptome sequencing (RNA-seq) data, we determined the relationship between translation and mRNA abundance rhythms, uncovered the set of mRNAs for which these rhythms are uncoupled, and

[3]**These authors contributed equally to this work.**
Corresponding author: david.gatfield@unil.ch
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.195404.115. Freely available online through the *Genome Research* Open Access option.

25:1848–1859 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/15; www.genome.org

calculated translation efficiencies transcriptome-wide. Moreover, we inferred the relative levels of clock protein biosynthesis and identified upstream open reading frame (uORF) translation as a novel regulatory mechanism within the clock circuitry. Altogether, our study reveals key features of rhythmic protein biosynthesis and the impact of translational control on gene expression in a solid, highly differentiated mammalian organ with well-studied functions.

Results

Ribosome profiling in liver around the clock

We collected time-resolved ribosome profiling data from 48 male mice entrained to light-dark cycles and euthanized at 2-h intervals around the clock. We assembled liver extracts into two independent replicate time series (12 timepoints, *Zeitgeber* time ZT0 to 22) (Fig. 1A), prepared ribosome footprints and matching total RNA, converted them into sequenceable libraries (Fig. 1B), and sequenced them with high coverage (Supplemental Table S1). Our protocol yielded high-quality footprints that mainly mapped to protein coding sequences (CDS) and were depleted from untranslated regions (UTRs) of mRNAs (Fig. 1C,D; Supplemental Table S1). Moreover, the predominant footprint length of 29–30 nt (Fig. 1E) allowed the precise identification of translated codons. The alignment of CDS-mapping reads relative to the position of the ribosome's aminoacyl tRNA-site (A-site; inferable from footprint length and sequence) (Ingolia et al. 2011) thus revealed excellent reading frame preference (Fig. 1F) and captured the CDS triplet codon composition transcriptome-wide (Fig. 1G). These characteristics were absent in the RNA-seq data, as expected (Fig. 1D–G; Supplemental Fig. S1A). Moreover, the quantification of CDS-mapping reads showed high reproducibility across biological replicates (Supplemental Fig. S1B,C). Finally, principal component analysis (PCA) on the ensemble of data sets (Fig. 1H) separated RPF-seq and RNA-seq data on PC1 and recapitulated its cyclic nature with near-perfect temporal resolution (PC2 and PC3); i.e., the covariates in the experimental design (RNA/RPF and factor time) were retrieved by an unsupervised method for sample clustering. We concluded that the data sets were of high technical quality and would be suitable for comprehensive analyses of rhythmic and constitutive translation in the liver.

Hallmarks of translational regulation in liver

Ribosome profiling from mammalian tissues is still relatively uncommon, and we therefore started with a characterization of general properties of the translatome data, independently of its time resolution. From the ratio of CDS-mapping RPF-seq to RNA-seq reads, we first computed relative ribosome occupancies, which can be interpreted as relative translational efficiencies (TEs) because each footprint reflects the synthesis of an individual protein molecule and, importantly, integrating read numbers across the entire CDS corrects for local variation in footprint density (Ingolia et al. 2011). Briefly, while the local speed of translation elongation may vary (e.g., ribosome pausing due to RNA structure or codon usage) and represents a source of inhomogeneous footprint distribution on a given CDS, average translation speeds across genes appear to be rather constant (Ingolia et al. 2011). It should be noted, however, that a possible influence of local variation on overall translation speed of an mRNA has been suggested (Dana and Tuller 2012) and is a current topic of debate (Ingolia 2014).

In mouse embryonic stem cells (mESCs), TEs cover an approximately 10-fold range and have an asymmetric distribution indicative of an intrinsic upper limit that is transcript-specifically decreased by inhibitory mechanisms (Ingolia et al. 2011). TEs in liver showed a broader dynamic range and analogous asymmetry (Fig. 2A). While multiple mechanisms are likely involved in establishing transcript-specific ribosome occupancies, several simple transcript features have previously been observed to correlate with TEs. In yeast, ORF length and translation rate correlate inversely, presumably due to a selection for faster translation initiation on loci encoding short proteins (Arava et al. 2003; Ingolia et al. 2009). Also in the liver data set, CDS lengths explained a significant proportion of variance in ribosome occupancies ($R^2 = 0.16$; $P = 1.26 \times 10^{-160}$) (Fig. 2B). 5′ UTR ($R^2 = 0.047$; $P = 2.33 \times 10^{-46}$) and 3′ UTR lengths ($R^2 = 0.015$; $P = 4.29 \times 10^{-16}$) correlated with TEs as well (Fig. 2B), and the predictive power of the 5′ UTR length remained significant even after correction for interdependence of UTR and CDS lengths (Supplemental Fig. S2A–C). These results are consistent with a prominent role for 5′ UTRs in translational control and with the idea that translation regulatory elements—of which longer 5′ UTRs potentially contain more—are predominantly inhibitory.

Within 5′ UTRs, uORFs are emerging as important *cis*-regulatory elements that control CDS translation, usually in an inhibitory fashion (Wethmar et al. 2014). Six percent of RPF-seq reads fell into annotated 5′ UTRs (Fig. 1C), albeit with pronounced transcript-specific variability (Supplemental Fig. S3A), suggestive of abundant uORF usage in the liver. uORFs are short and often poorly conserved (Churbanov et al. 2005) and frequently initiate at near-cognate (non-AUG) start codons (Ingolia et al. 2011), complicating prediction just from sequence. To explore uORF usage in the liver, we therefore compiled a uORF-enriched transcript set based on whether the 5′ UTRs harbored sequence stretches (1) embraced by AUG initiation and stop codons and (2) covered by footprints with distinct reading frame preference. Despite these simple criteria that miss, for example, non-AUG-initiated uORFs, the detected transcripts showed significantly lower main ORF TEs (difference in location of red vs. gray densities on the ordinate of –0.525 corresponds to >30% TE reduction; $P = 2.2 \times 10^{-16}$; Wilcoxon rank-sum test) (Fig. 2C), whereas the TEs of transcripts lacking translated uORFs were slightly increased (Fig. 2C, blue). Importantly, the correlation of low TEs with translated uORFs was independent of 5′ UTR (or CDS/3′ UTR) lengths (Supplemental Fig. S2D–F). Instead, linear regression analysis uncovered that the underlying cause for the correlation of TE with 5′ UTR length (see above) (Fig. 2B) was uORF presence rather than 5′ UTR length per se (Supplemental Fig. S2, cf. G–K and A–C). Finally, uORF-containing mRNAs were globally less abundant (Fig. 2C, red density on abscissa), possibly reflecting the activity of the nonsense-mediated mRNA decay (NMD) pathway that selectively degrades mRNAs with premature termination codons and that is known to act on uORF-containing transcripts (Mendell et al. 2004). We concluded that uORF usage is frequent in the liver and likely a major determinant of transcript-specific TEs, correlating with reduced protein biosynthesis from the CDS. Of note, we also analyzed another feature that may have been expected to correlate with TE, i.e., the presence of pause sites. By using an analogous approach to Ingolia et al. (2011) to identify local variation in footprint density that exceeded the CDS median (more precisely, we used the similar, but for low translated transcripts, less stochastic "trimean") by a certain threshold, we found that the presence of such sites had no predictive power for translation rates (Supplemental Fig. S4A,B).

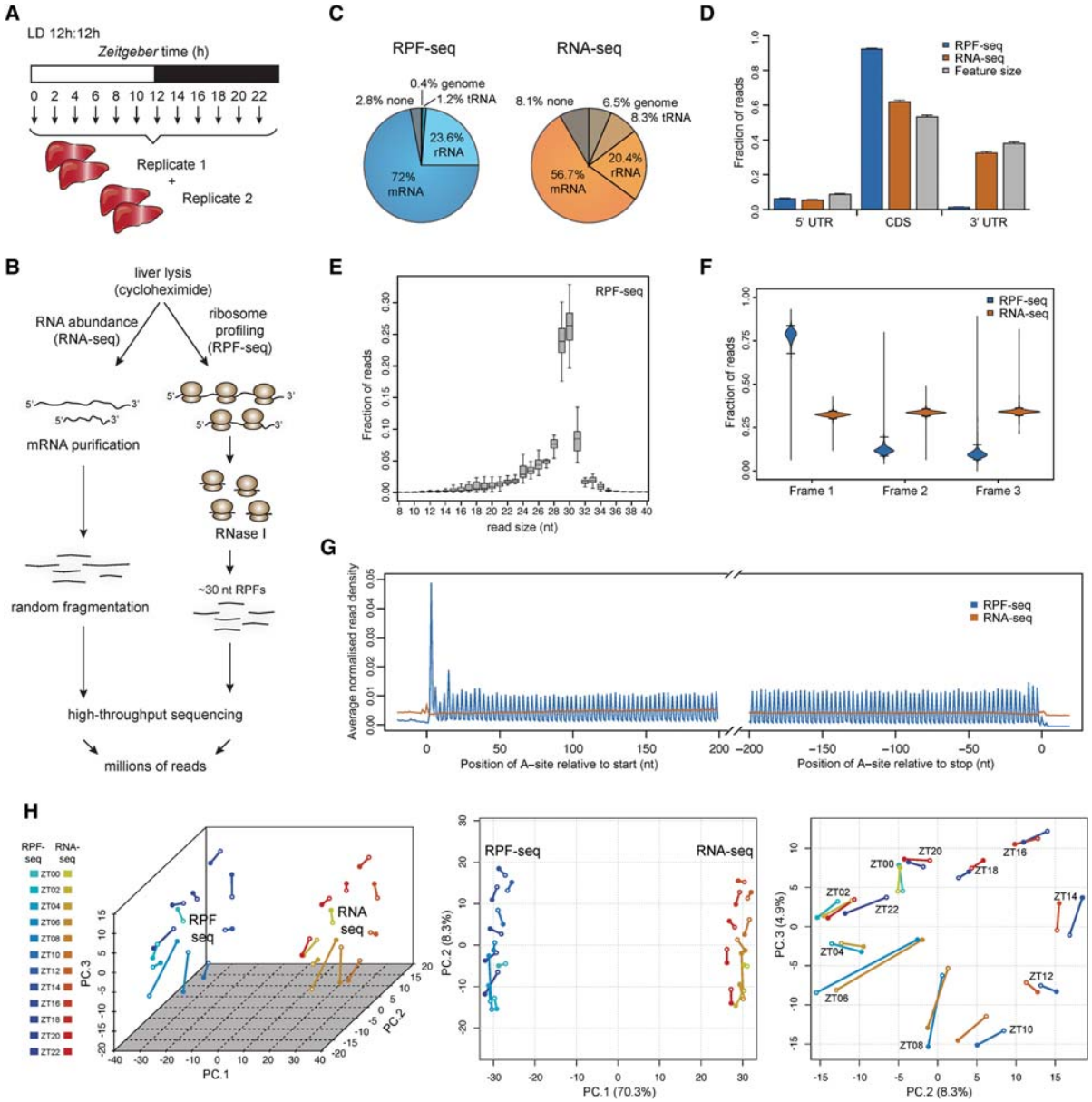


Figure 1. Time-resolved ribosome profiling data from mouse liver. (*A*) Overview of the experimental design for liver sampling over the 24-h cycle of the day. Forty-eight livers were collected and assembled into two replicate time series of 12 timepoints around the clock (each sample representing a pool of two mice). (*B*) Overview of the main steps in the protocol for the preparation of RNA abundance (RNA-seq) and ribosome profiling (RPF-seq) data; for details, see Supplemental Material. (*C*) Mapping summary of RPF-seq and RNA-seq reads across all replicates and timepoints. Note that RPF-seq reads were enriched for mRNAs, as expected. For detailed mapping outcome, see Supplemental Table S1. (*D*) Read distribution within 5′ UTRs, CDS, and 3′ UTRs for RPF-seq (blue) and RNA-seq data (orange) compared with the distribution expected by chance, which is determined by the feature sizes (gray; $N = 10829$). Note the enrichment of RPF reads within CDS, the depletion from 3′ UTRs, and considerable amounts of reads (6%) within the 5′ UTR. (*E*) Insert size distribution of RPF-seq reads across all replicates and timepoints shows that the majority of footprints are 29–30 nt in length. Box-and-whisker plots: midline, median; box, 25th and 75th percentiles. Whiskers extend to the minimum and maximum values within 1.5 times the inter-quartile range from the box. (*F*) Frame analysis for RPF-seq and RNA-seq reads within the CDS (using genes for which the expressed transcript isoforms define one main translated CDS/protein—called single protein isoform genes—with an RPF-RPKM [reads per kilobase per million mapped reads] value >5 and fulfilled a few other minor criteria described in Supplemental Material; $N = 3793$). RPF-seq reads show a clear preference for reading frame 1 (the annotated frame), whereas RNA-seq reads distribute equally across the three reading frames, as expected. Violin plots extend to the range of the data, with horizontal lines marking the 2.5% and 97.5% quantiles. (*G*) Read density distribution of RPF-seq and RNA-seq reads within 200 nt from the start or −200 nt from the stop codons reveals a 3-nt periodicity of RPF reads within coding sequences. The analysis used only transcripts from single protein isoform genes (see *F*) with RPF-RPKM > 5 and CDS > 400 nt ($N = 3237$) and quantified the number of reads per nucleotide based on the A-site prediction as described in the Supplemental Material. (*H*) Principal component (PC) analysis of RPF-seq and RNA-seq data sets, using the top-ranked 4000 genes (see Supplemental Methods). The first three PCs explain 70.3%, 8.3%, and 4.9% of total variation, respectively (3D scatter plot, *left* panel). While PC1 mainly reflected variance attributable to differences between the mRNA abundance and footprint data sets (*middle* panel), PC2 and PC3 resolved mainly variance attributable to factor time (*right* panel). Note that the timepoints assemble to a near-perfect "clock" in the PC2 versus PC3 representation. A scree plot showing contributions of further PCs can be found in Supplemental Figure S1D.

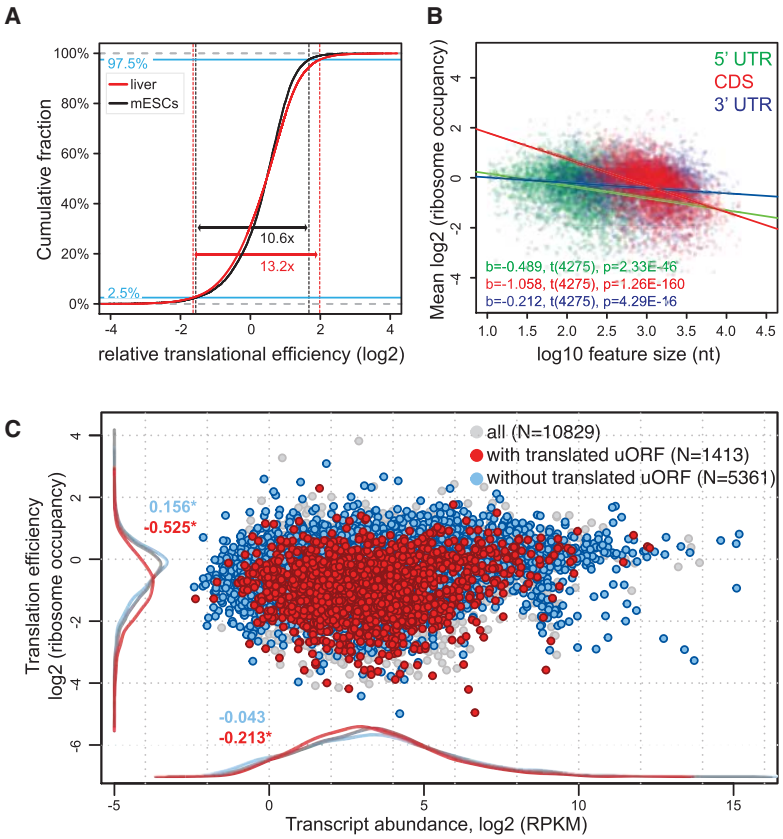


Figure 2. Analysis of translation efficiencies in mouse liver. (*A*) TE distribution in mouse liver (red curve; representing all 24 ZTs from 10829 genes, i.e., $N = 129870$ individual data points) compared with that in mESCs (black curve; data from Ingolia et al. 2011; $N = 10217$ genes). Liver data were adjusted for mean to ensure comparability with mESC data. The asymmetrical nature of the TE distribution that has previously been reported for mESCs and that is indicative of an intrinsic upper limit to translation rates (Ingolia et al. 2011) is also observed in the liver. Note that the TE range is significantly broader in the liver, where 95% of data fall into a 13.2-fold TE range, compared with a 10.6-fold range in mESCs ($P < 2 \times 10^{-5}$, permutation test). (*B*) Correlation between TEs and 5′ UTR, CDS, or 3′ UTR lengths. Analysis was performed on transcripts from genes for which the transcriptomics analysis showed that a single protein isoform was produced, and that had an RNA-seq RPKM value >5, and 5′ and 3′ UTR lengths ≥10 nt ($N = 4277$). Linear regression lines for each group are plotted *over* the data points, and related *t*-test results of the regression slopes are reported in the plot area with the same color code. Inverse and statistically significant correlation between TE and feature length was thus apparent for all three features, with predictive value CDS > 5′ UTR > 3′ UTR. (*C*) Scatter plot of TEs (ordinate) versus transcript abundances (TAs; abscissa) averaged over timepoints and replicates. Highlighted are transcripts from single protein isoform genes, which do (red) or do not (blue) contain at least one translated AUG-initiated uORF. Density curves of TEs and TAs for highlighted data points are plotted on the margins with same color code. uORF translation is thus associated with a pronounced TE decrease and a slight decrease in transcript abundances. Numbers on density curves reflect the location shift ($\log_2$ values of the median calculated from the differences across all timepoints) relative to all transcripts. Transcripts with translated uORF: TE, $P < 2.2 \times 10^{-16}$; TA, $P = 1.3 \times 10^{-5}$ (Wilcoxon rank-sum test). Transcripts without translated uORF: TE, $P < 2.2 \times 10^{-16}$; TA, $P = 0.287$ (Wilcoxon rank-sum test).

In the core clock and globally, rhythmic mRNA abundance is a good predictor of footprint rhythms

We used the clock genes for first temporal analyses of the data sets. As illustrated by the anti-phasic expression of *Arntl* and *Nr1d1*, core clock transcripts were detected with high coverage and oscillated in both the RNA-seq and RPF-seq data sets (Fig. 3A). Read count integration over the CDS indicated that for all core clock components, footprint profiles closely matched mRNA abundance rhythms (Fig. 3B). We concluded that the rhythmic biosynthesis of core clock proteins was determined by mRNA availability with no further regulation by time of day–dependent translation (Supplemental Fig. S5A). We next conducted transcriptome-wide rhythmicity analyses. Applying a more than 1.5-fold peak-to-trough amplitude cut-off, we identified oscillations in the RNA-seq and RPF-seq data sets that affected in both cases ~17% of the protein-coding transcriptome (almost 1900 mRNAs) (Fig. 4A; Supplemental Table S2). However, mRNA abundance and ribosome occupancy rhythms showed different peak phase distributions (Fig. 4B,C). In good agreement with previous reports (Le Martelot et al. 2012; Zhang et al. 2014), a majority of mRNAs thus showed maximal abundance during the night, with an enrichment around ZT15–ZT19 (Fig. 4B). In contrast, maximal translation was prevalent at the beginning of the dark phase, with a dominant peak around ZT15–ZT16 (Fig. 4C). These different distributions resulted from transcripts that were unique to either data set rather than from phase delays occurring between mRNA accumulation and translation, because the intersecting set of 1192 "mRNA and footprints rhythmic" transcripts (Fig. 4A) showed near-identical RNA-seq and RPF-seq oscillations (Fig. 4D–G; Supplemental Fig. S6A). We concluded that whenever both mRNA abundance and ribosome occupancy cycled, they globally did so in sync. Similar to the core clock components (Fig. 3), most rhythmic mRNAs were thus translated concomitant with their cellular accumulation and had constant TEs. Distinct out-of-phase translation was indeed confined to rather few exceptional cases (Supplemental Fig. S6B). Finally, it is noteworthy that the TEs of "mRNA and footprints rhythmic" transcripts were slightly increased compared with the global population of expressed transcripts (location shift of dark blue vs. gray densities of 0.106 corresponds to ~8% higher TEs; $P = 2.2 \times 10^{-05}$; Wilcoxon rank-sum test) (Fig. 4H). No enrichment or depletion for rhythmic genes was seen with regard to AUG-initiated uORFs or pause sites ($P = 0.533$ and $P = 0.315$, respectively; Fisher's exact test).

Widespread time of day–dependent translation of nonrhythmic mRNAs

More than one-third (682/1874) of genes that cycled at the footprint level did not have a rhythmic mRNA (Fig. 4A); globally, these transcripts showed decreased TEs (location shift of –0.105 corresponds to a 7% TE reduction; $P = 0.002$) (Fig. 4H, light blue). However, closer inspection of the underlying RPF-seq and

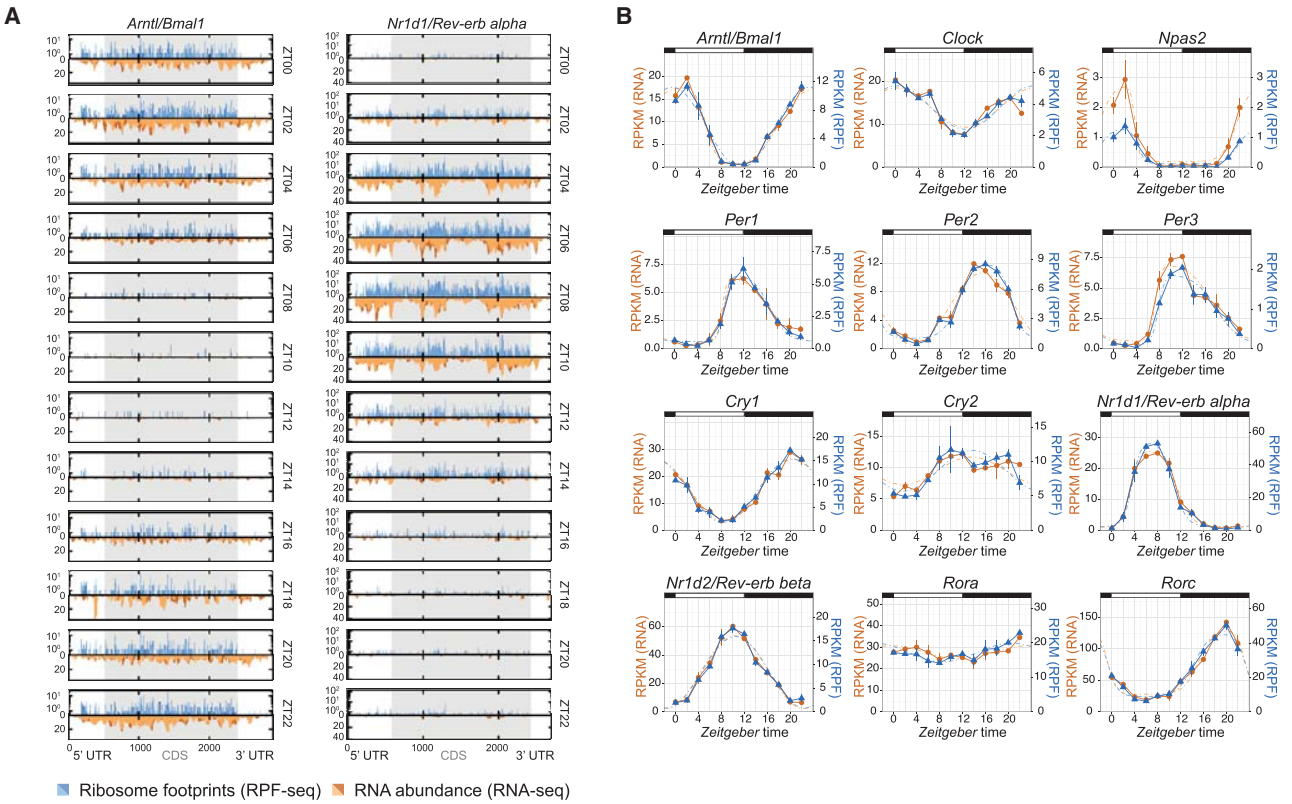


Figure 3. Core clock transcripts show mRNA abundance and ribosome occupancy in sync. (*A*) Time-resolved distribution (ZT00 to ZT22; arranged vertically) of normalized counts of ribosome profiling (RPF-seq; blue) and RNA abundance (RNA-seq; orange) reads along the *Arntl/Bmal1* and *Nr1d1/Rev-erb alpha* transcripts. Different color shadings (dark/light orange and blue) indicate the two biological replicates. Gray shading of the boxes marks the CDS; UTRs are in white. (*B*) RPKM values of CDS-mapping RPF-seq (blue) and RNA-seq (orange) data for circadian core clock genes around the 24-h daily cycle. Means per timepoint are plotted; error bars, the two biological replicates. Dashed lines represent rhythmic curve fittings to the data.

RNA-seq profiles indicated that in many cases, nonrhythmicity assignments at the mRNA level had resulted from noise or low amplitudes (close to the imposed 1.5-fold cut-off) in expression profiles that otherwise still appeared to be rhythmic (false-negatives caused by "cliff effects") (Supplemental Fig. S6C). To refine the "mRNA flat–footprints rhythmic" assignments, we used the analytical framework *Babel* (Olshen et al. 2013) to first identify all transcripts that had significant TE differences over timepoints (and/or whose TEs significantly deviated from the global transcript population), and subsequently performed the rhythmicity analyses on these. This strategy resulted in a high-confidence set of 147 rhythmically translated, but otherwise nonoscillating, mRNAs (Fig. 5A; Supplemental Table S3). Their RPF-seq profiles showed a striking phase distribution with a dominant peak around the day-to-night transition (ZT10–ZT16) (Fig. 5B). Gene ontology (GO) analyses revealed enrichment for mRNAs encoding components of the protein biosynthesis machinery, including RPs, elongation factors, and poly(A) binding proteins, whose translation underwent a characteristic upsurge starting from ZT10 (Fig. 5C,D). Of note, increased polysome association at the beginning of the night has previously been described for this class of transcripts (Jouffe et al. 2013), which all contain 5′-TOP motifs that are regulated by TORC1 (Meyuhas and Kahan 2015). Of 79 RPs, 35 were contained in the high-confidence list, and—with the exception of eight proteins whose mRNAs were undetectable or translationally invariable—visual inspection confirmed that most other RPs shared a similar RPF-seq profile as well (Supplemental Fig. S7A; Supplemental Dataset 1). In summary, our data extend previous findings (Jouffe et al. 2013) and precisely quantify the coordination of protein biosynthesis within the translational apparatus in mouse liver. Of note, our study uncovers a peculiarity of RP gene expression, i.e., particularly high mRNA abundances paired with low TEs, which undergo coordinated upsurge/translational de-repression prior to the day-to-night transition (Supplemental Fig. S7B; Supplemental Movie M1).

Other GO terms and individual "mRNA flat–footprints rhythmic" transcripts caught our attention as well. The electron transport chain components that the GO analysis identified (Fig. 5C) corresponded throughout to mitochondrially encoded transcripts characterized by a translational spike at ZT12 (Supplemental Fig. S8A; Supplemental Dataset 1). As Western blot analysis did not reveal any oscillations at the protein level (Supplemental Fig. S8B), the significance of these translational rhythms remains to be uncovered. We next verified for other examples of rhythmic translation whether protein abundances oscillated. This was indeed the case for geranylgeranyl diphosphate (GGPP) synthase 1 (*Ggps1*), encoding a key branchpoint enzyme in the mevalonate pathway (Supplemental Fig. S8C,D). GGPP is important for the C20-prenylation of proteins and for the regulation of the nuclear receptor NR1H3/LXR alpha (Forman et al. 1997). Moreover, two transcription factors, Deformed epidermal autoregulatory factor 1 (*Deaf1*)

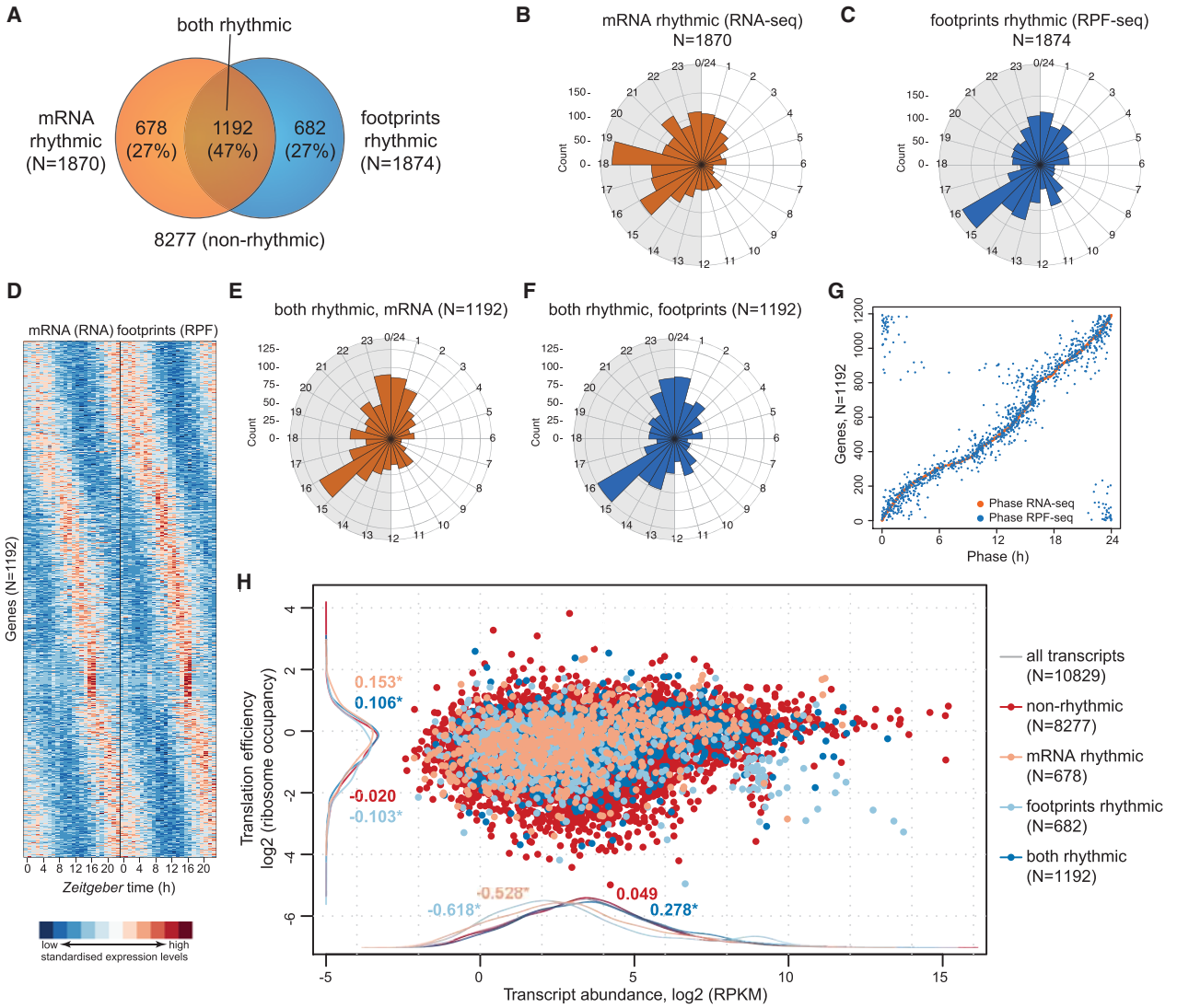


Figure 4. Transcriptome-wide analysis of transcript abundance and ribosome occupancy rhythms. (*A*) Venn diagram summarizing the result of rhythmicity detection in the RNA-seq and RPF-seq data. Of a total of 10,829 expressed protein-coding loci, 1870 showed mRNA rhythms and 1874 had footprint rhythms (in both cases 17% of all) with a >1.5× peak-to-trough amplitude (FDR < 0.05); 1192 transcripts were common to both sets. (*B*) Phase histograms for the transcripts in *A*, showing the peak phase distribution of mRNA abundance (RNA-seq) rhythms over the 24-h cycle in orange. The length of the spoke indicates how many transcripts peaked at a specific time. (*C*) As in *B* for footprint rhythm (RPF-seq) in blue. Note the different phase distribution of ribosome occupancy rhythms compared with RNA abundance rhythms depicted in *B*. (*D*) Heat map of rhythms at the level of mRNA abundance (RNA-seq; *left*) and footprints (RPF-seq; *right*) for the overlapping set from *A* (1192 genes). Transcripts are sorted by the phase of maximal ribosome occupancy. For the representation, mRNA abundances and translation levels are standardized within each gene (row) and independently for RNA-seq and RPF-seq columns (*Z*-scores). (*E*) Phase histograms showing the phase distribution of "mRNA and footprints rhythmic" transcripts from the overlap in *A* ($N = 1192$) for mRNA accumulation (RNA-seq) in orange. (*F*) As in *E*, but for footprints (RPF-seq) in blue. Note that the distribution is near-identical to that in *E*. (*G*) Phase correlation plot of the "mRNA and footprints rhythmic" genes ($N = 1192$). Each row contains two dots marking the phase of maximal mRNA abundance (orange) and the phase of maximal footprints (blue) for each gene. Genes are ordered according to the phase of the mRNA. (*H*) Scatter plot of TE versus transcript abundances for the genes classified into the different rhythmicity categories shown in *A* (values averaged over timepoints and replicates). Density curves are plotted on the margins with the same color code. Numbers on density curves reflect the location shift ($\log_2$ values) relative to all transcripts (gray). (*) Significance at the 0.05 level (Wilcoxon rank-sum test). The plot shows that rhythmic genes (i.e., those with rhythmic mRNAs [light salmon], with rhythmic footprints [light blue], or with rhythmic mRNA and footprints [dark blue]) are significantly different at the level of TE and of transcript abundances than transcripts of nonrhythmic genes (red).

and Max interactor 1 (*Mxi1*), showed robust greater than twofold rhythms in translation (Fig. 5E; Supplemental Fig. S8E). For DEAF1, steady-state protein levels oscillated as well (Fig. 5E), raising the interesting possibility that translational rhythmicity is propagated to transcriptional target genes and thus contributes to shaping the rhythmic transcriptome. In accordance, previously reported DEAF1 targets (Yip et al. 2009) were significantly enriched for genes that are rhythmically transcribed (enrichment 1.6-fold; $P = 0.009$) (Fig. 5F; identified by Du et al. 2014). Notably, the majority of DEAF1 target gene pre-mRNAs peaked between ZT6 and

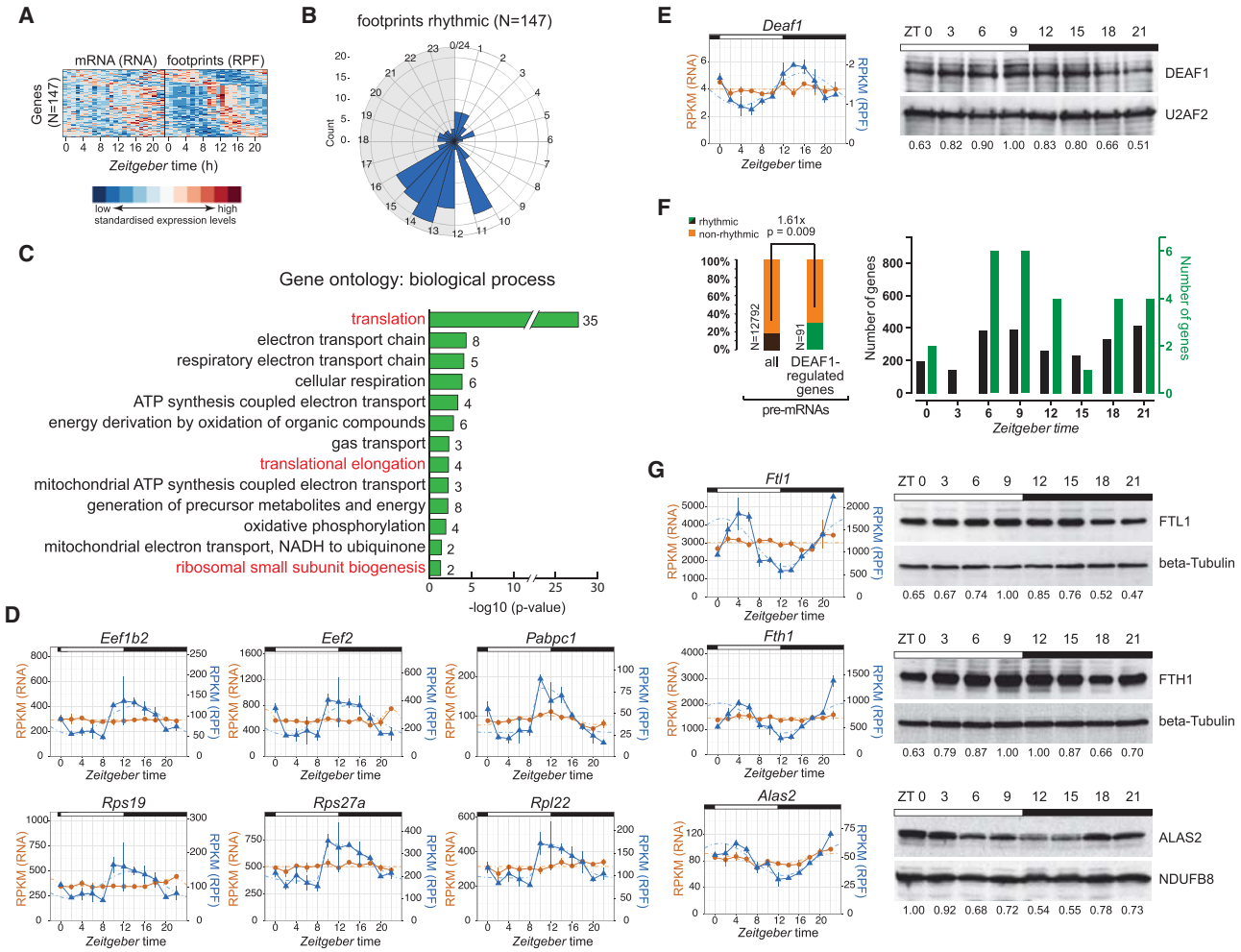


Figure 5. Daytime-dependent translational control of protein biosynthesis machinery components, of transcription factors, and in iron metabolism. (*A*) Heat map of "mRNA flat–footprints rhythmic" genes identified after *Babel* analysis (significant changes in ribosome occupancy) showing the mRNA abundance (RNA-seq; *left*) and footprint (RPF-seq; *right*) data for 147 genes. Transcripts are sorted by the phase of maximal ribosome occupancy. For the representation, mRNA abundances and translation levels are standardized within each gene (row) and independently for RNA-seq and RPF-seq columns (*Z*-scores). Note that this high-confidence set of 147 transcripts has clear rhythms at the translational level, but not so at the mRNA abundance level. (*B*) Phase histogram showing the phase distribution of footprint data across the 24-h cycle. Generally, there was a strong enrichment of ribosome occupancy maxima at the light-to-dark transition. This distribution is significantly nonrandom ($P = 1.12 \times 10^{-10}$; $R = 0.395$; Rayleigh uniformity test) and significantly different from those in Figure 4C ($P < 7.6 \times 10^{-08}$; $W = 32.80$; Watson–Wheeler test for homogeneity of angles). (*C*) Gene ontology (GO) analysis of biological process based on genes detected in *A*. Bar graph shows $\log_{10}$ of *P*-values for each category. Numbers *next* to the bars represent the number of genes within each category. Marked in red are the GO terms connected to the protein biosynthesis machinery. (*D*) RPF-seq (blue) and RNA-seq (orange) profiles for several rhythmically translated transcripts connected to the protein biosynthesis machinery—ribosomal proteins, translation factors, and poly(A) binding proteins—around the 24-h cycle. Means per timepoint are plotted; error bars, replicates. Dashed lines represent rhythmic curve fittings. (*E*, *left*) RNA-seq (orange) and RPF-seq (blue) profile as in *D* for transcription factor *Deaf1* around the clock. (*Right*) Western blot analysis for DEAF1 (and loading control U2AF2) in liver nuclear extracts. Numbers *below* the panels show relative levels of DEAF1 protein after normalization to U2AF2. Data from one representative time series are shown ($N = 2$). (*F*, *left*) Rhythmically transcribed genes (from Du et al. 2014) are 1.6-fold enriched in DEAF1-regulated genes (Fisher's exact test $P = 0.009$) (Yip et al. 2009), consistent with the idea that rhythmic translation and protein accumulation of DEAF1 leads to rhythmic transcriptional activity on its target genes. (*Right*) Peak abundances of rhythmically transcribed genes around the clock. (Black) All rhythmically transcribed genes; (green) rhythmic DEAF1-regulated genes. (*G*, *left*) RNA-seq and RPF-seq profiles for *Ftl1*, *Fth1*, and *Alas2*, similar as in *D*. Note the high-amplitude rhythms in translation from flat mRNAs. (*Right*) Confirmation of protein rhythmicity by Western blot analysis of total protein extracts (FTL1, FTH1) or mitochondrial extracts (ALAS2) prepared from mouse liver. Protein levels were normalized to beta tubulin or NDUFB8, respectively. Protein data from one representative time series are shown ($N = 2$–3).

ZT12, coinciding with maximal DEAF1 protein abundance (Fig. 5E,F).

Finally, our data revealed that the well-known case of translational control through iron-responsive elements (IREs) undergoes high-amplitude oscillations. IREs are stem–loops found in mRNAs involved in iron, oxygen, and energy metabolism (see Anderson et al. 2012 and references therein). Depending on the cytosolic iron concentration and other cues (see Discussion), IREs located in 5′ or 3′ UTRs regulate mRNA translation and degradation, respectively. Ferritin heavy and light chain 1 (*Fth1*, *Ftl1*), involved in iron storage, as well as aminolevulinic acid synthase 2 (*Alas2*), a rate-limiting enzyme of heme synthesis, showed high-amplitude

oscillations in translation that led to protein rhythmicity (Fig. 5G). All three transcripts contain IREs in their 5′ UTRs, whereas transferrin receptor (*Tfrc*) is a case where a 3′ UTR–borne IRE controls mRNA turnover (Anderson et al. 2012). *Tfrc* mRNA (and footprint) profiles were rhythmic in our data set (Supplemental Dataset 1); moreover, previous RNA-seq data quantifying pre-mRNA and mRNA levels around the clock (Du et al. 2014) revealed that the abundance of *Tfrc* mRNA, but not of its pre-mRNA (and hence, its transcription), oscillated (Supplemental Fig. S8F). These analyses suggest that regulation by IREs is overall under time of day–dependent control. Of note, besides the temporal regulation of IRE-containing transcripts, we observed high-amplitude oscillations in gene expression (RNA-seq and RPF-seq) throughout key steps in iron metabolism (Supplemental Fig. S8G). These findings indicated widespread rhythmic regulation of iron homeostasis that has been largely overlooked so far.

Core clock transcripts show a broad range of TEs and abundant uORF usage

Although core clock transcript TEs were temporally invariable (see above) (Fig 3B; Supplemental Fig. S5A), further analyses of the data revealed features of regulation that were potentially of functional importance. First, we noticed that the intrinsic TEs of clock mRNAs spanned a greater than sixfold range from the poorest (*Clock, Per3*) to the best (*Nr1d1*) translators (Supplemental Fig. S5A,B, upper panel). TEs and mRNA abundances together define the amounts of protein that are produced, and RPF-seq RPKMs are hence a direct readout of relative protein biosynthesis levels. We therefore used our footprint counts to precisely quantify the stoichiometry at which the clock proteins are produced. Importantly, these estimates explained known features of the clock better than did RNA expression data alone. Notably, *Clock* is in excess of its heterodimerization partner *Arntl* at the transcript level (integrated over the day, about 1.6-fold more *Clock* than *Arntl* mRNA), but due to TE differences, about 1.5-fold more ARNTL than CLOCK protein is produced (Supplemental Fig. S5B; for similar results obtained when peak levels rather than daily amounts were considered, see Supplemental Fig. S5C), which is consistent with the conjecture that ARNTL is in excess over CLOCK (Huang et al. 2012). For the main positive and negative limb components, our data indicated daily biosynthesis at a ratio of CLOCK(1.0):ARNTL(1.5):PER1(0.6):PER2(1.1):CRY1(2.3):CRY2(2.3), i.e., overall similar levels of produced proteins (Supplemental Fig. S5B). Finally, it is established that in the interconnecting limb NR1D1/REV-ERB alpha represents the dominant *Rev-erb* paralog in liver (Preitner et al. 2002; Bugge et al. 2012). While *Nr1d2/Rev-erb beta* is nevertheless significantly more abundant at the mRNA level, the amount of biosynthesized NR1D1 protein exceeds the NR1D2 paralog by greater than twofold due to greater than fivefold differences in TEs (Supplemental Fig. S5B, lower panel). In summary, these analyses suggest that TE is an important factor in establishing the appropriate clock protein output. Conceivably, it may represent an additional layer at which the core clock can undergo regulation.

The striking correlation of uORF usage with TEs (Fig. 2C) prompted us to explore whether translated uORFs were present in core clock transcripts. Intriguingly, *Arntl, Clock, Cry1, Nr1d1*, and *Nr1d2,* all showed considerable ribosome occupancy in their 5′ UTRs and contained one or more AUG-initiated uORFs (Fig. 6A) with footprint coverage that showed clear frame preference, indicative of their active translation (Fig. 6B). To investigate how uORFs regulated the rhythmic production of a clock protein, we chose *Nr1d1*, for which RPF-seq reads on uORFs 1 and 2 showed particularly high coverage (Fig. 6A), frame bias (Fig. 6B), and rhythmicity in sync with the main ORF (Supplemental Fig. S9A). Moreover, uORF1 was remarkably long (192 nt) and conserved in mammals, potentially coding for a 63-amino-acid polypeptide (Supplemental Fig. S9B). We constructed a lentiviral reporter gene from an *Nr1d1* genomic fragment that contained promoter sequences conferring rhythmic transcription (Stratmann et al. 2010), exon 1 (5′ UTR and codons 1–10), intron 1, and a modified exon 2, in which firefly luciferase (FL) was fused to NR1D1 after amino acid 15 (Fig. 6C). Moreover, we designed reporter variants in which the predicted uORF initiation codons were mutated or the uORFs were deleted altogether (Supplemental Fig. S9C). When expressed in a circadian model cell line, NIH3T3 fibroblasts, all constructs showed comparable high-amplitude bioluminescence rhythms in the detrended data (Fig. 6C), indicating that uORFs were dispensable for rhythmic protein expression per se. Importantly, the differences in absolute luciferase signals that we observed between reporters (Supplemental Fig. S9D,E) could have resulted from altered TEs or simply reflect unequal lentiviral titers, transduction efficiencies, cell numbers, or similar. We therefore measured the effect of the uORFs on main ORF translation in an independent assay in which the 5′ UTRs were cloned upstream of FL CDS in a lentiviral vector that also expressed an internal control gene, *Renilla* luciferase (RL), driven from the same bidirectional promoter (Fig. 6D, top; Du et al. 2014). These analyses revealed that uORF deletions, or the subtle initiation codon mutations, led to increased levels of FL reporter activity (Fig. 6D, bottom). Moreover, uORF1 and uORF2 had an additive effect, as judged by mutants in which uORF1 and uORF2 AUGs were mutated to alanine codons either singly (mutants M1A-uORF1 and M1A-uORF2) or in combination (M1A-uORF1+2). We next measured RNA expression levels of both luciferases, which allowed us to estimate the relative contributions that altered RNA stability (lighter shading in Fig. 6D, bottom) and translation regulation (darker shading in Fig. 6D, bottom) made to the observed increases in reporter protein output. These analyses suggested dual contribution by both mechanisms, in line with the initial observation (Fig. 2C) of decreased TE and mRNA abundance of uORF-containing transcripts (the latter possibly involving regulation through the NMD pathway). Extrapolated to the regulation of the endogenous *Nr1d1* transcript in vivo, the prediction from these results would be that uORF1+2 could regulate the magnitude of NR1D1 oscillations.

There was evidence for translated uORFs in several core clock transcripts (Fig. 6A,B), but the net regulatory effect of uORF translation on clock function would likely be difficult to predict from studying each case individually. In order to estimate the overall impact of uORF translation on the clock, we took advantage of the recent discovery that Density Regulated Protein (DENR) is implicated in ribosome recycling after translation termination and acts as a selective regulator of reinitiation at the main CDS after uORF usage (Skabkin et al. 2013; Schleich et al. 2014). In the absence of DENR, proteins whose expression is regulated by uORFs are thus produced at lower levels (Schleich et al. 2014). We down-regulated endogenous *Denr* in NIH3T3 cells carrying a circadian *Dbp-Luciferase* reporter gene (Stratmann et al. 2012) using three different shRNAs (Fig. 6E). *Denr*-deficient cells showed robust period shortening of free-running circadian oscillations by up to 1.5 h (Fig. 6F); moreover, a short period phenotype was also observed using a second reporter, *Arntl-Luciferase* (Supplemental

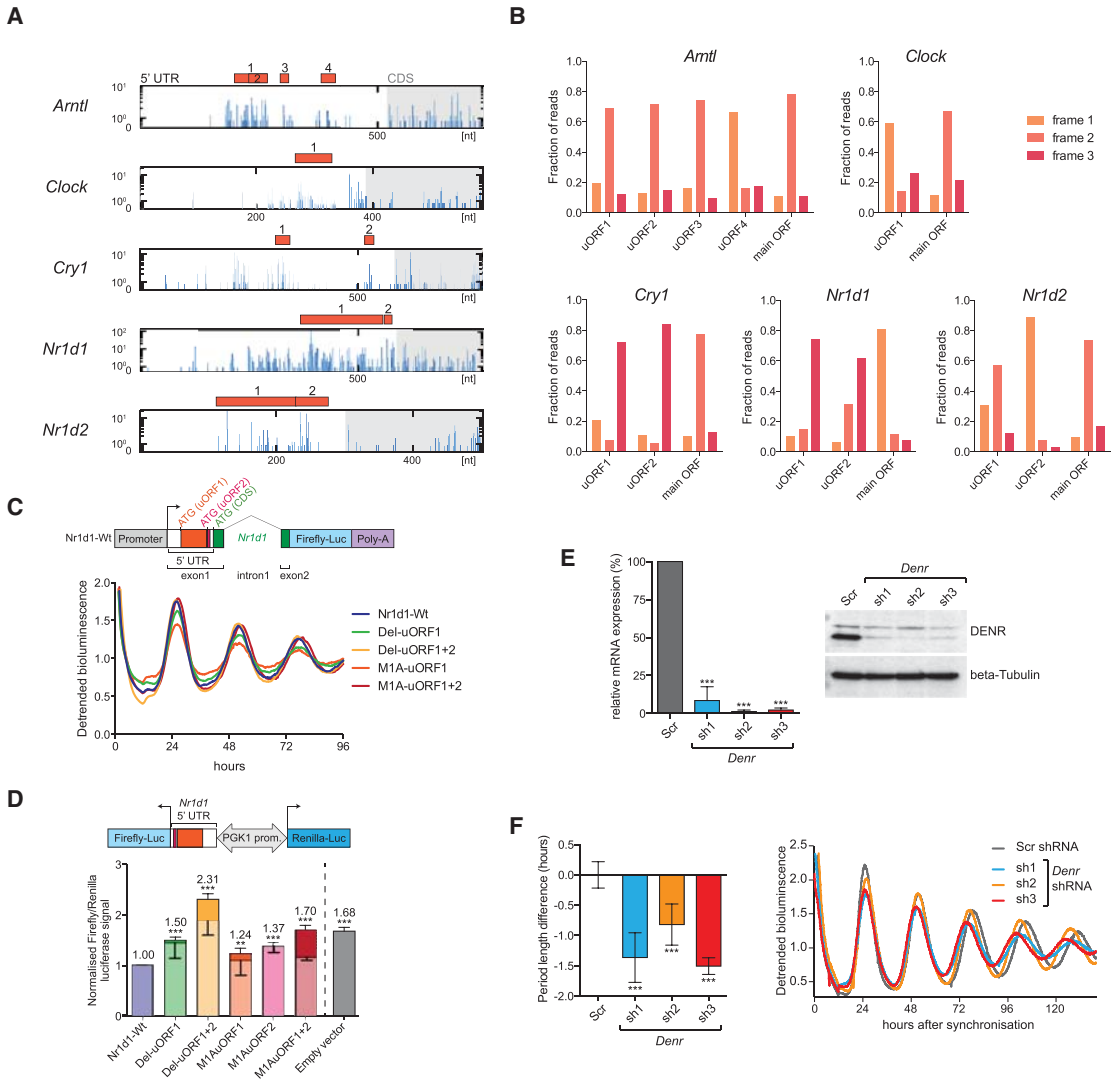


Figure 6. uORF translation is prevalent in core clock transcripts and impacts clock functions. (*A*) The 5′ UTRs of the depicted clock transcripts all contained at least one translated AUG-initiated uORF. Distribution of raw read counts of RPF reads (blue) along the 5′ UTR (white region in box) and the first 200 nt of the CDS (gray shaded region) is shown for the timepoint with maximal CDS translation. Red filled boxes indicate AUG-initiated uORFs within the 5′ UTR. Predicted uORFs for each gene are serially numbered. (*B*) Frame preference of uORF-mapping footprints. The fractions of footprints aligning to the three reading frames are shown for the uORFs shown in *A* and for the main ORF (CDS). Frame definition is relative to the annotated 5′ end of the transcript; please note that frame definition is different from that in Figure 1F. Most uORFs are thus covered by footprints that have a similar degree of frame preference as the main ORF-mapping footprints, indicating that uORF-mapping reads likely originate from processive translation. (*C*, *top*) Schematic showing the wild-type *Nr1d1*-firefly luciferase (FL) reporter gene consisting of a genomic *Nr1d1* fragment in which FL (blue) is expressed in fusion with the first 15 amino acids of NR1D1 (green). In exon 1, the location of uORF1 and -2 (red) and their predicted start codons within the 5′ UTR is shown. (*Bottom*) Real-time bioluminescence recordings of luciferase rhythms in NIH3T3 cells lentivirally transduced with the *Nr1d1*-FL reporter (Wt) and various mutants in which either uORFs are deleted (Del mutants) or uORF initiation codons are mutated to an alanine codon (M1A mutants). Cells were synchronized with dexamethasone. Raw bioluminescence was detrended using a 24-h moving average, and one representative replicate of a total $N = 3–9$ is shown. (*D*, *top*) Schematic representation of the dual luciferase reporter construct to measure how the *Nr1d1* 5′ UTR (Wt/mutants) influences the expression of the FL CDS. From the same bidirectional promoter, *Renilla* luciferase (RL) is expressed for internal control. (*Bottom*) Results of dual luciferase assay where FL signals were internally normalized to RL. Empty vector (gray) only contained the vector-encoded 5′ UTR. Experiments were performed in NIH3T3 cells ($N = 2–4$ independent experiments of triplicates). Lighter shading of the bars indicates the proportion of the increase that can be attributed to increased FL mRNA abundance in the mutants (measured by qRT-PCR), leaving the remainder of the increase (darker shading) attributable to translation. Note that whenever the translated uORF1 and -2 of the *Nr1d1* 5′ UTR are deleted (Del mutants) or just the initiation codons are mutated to alanine codons (M1A mutants), the inhibitory activity of the *Nr1d1* 5′ UTR is relieved. uORF1 and uORF2 appear to have an additive inhibitory effect on main ORF translation (cf. M1A uORF1 and -2 single mutants with the double mutant). For a schematic of the mutants, see also Supplemental Figure S9C. (*E*, *left*) Relative mRNA levels (RT-qPCR; normalized to expression of control gene *Nudt4*) of *Denr* in DBP-Luciferase reporter–expressing NIH3T3 cells transduced with scramble shRNA (Scr; serving as control) or three different shRNAs targeting *Denr* ($N = 3$). All shRNAs reduced *Denr* expression to <10%. (*Right*) Western blot analysis for DENR indicated efficient depletion at the protein level. Beta tubulin served as loading control. (*F*) shRNA-mediated *Denr* knockdown causes a short period phenotype of free-running circadian rhythms in NIH3T3 cells. (*Left*) Summary of period change engendered by *Denr* shRNAs 1–3 relative to Scr (control) shRNA in cells expressing the *DBP-Luciferase* reporter ($N = 3–8$). (*Right*) Representative bioluminescence tracks of Scr (control) and *Denr* shRNA-transduced DBP-Luciferase cells. Depending on the shRNA, the period of free-running circadian oscillations was 0.8–1.5 h shorter. (*D–F*) Bar graphs, mean ± SD; (**) $P < 0.01$, (***) $P < 0.001$ (*t*-test). *P*-value in *D* refers to differences in FL/RL activities (darker shading).

Fig. S9F). We concluded that DENR has important functions in the regulation of the mammalian circadian clock.

Discussion

How translation efficiency contributes to temporal gene expression is a largely unexplored facet of chronobiology. Translation is one step closer than the mRNA to the relevant output of most gene expression, the protein. Our ribosome profiling data should therefore be of wide interest to the research community and complement the many transcriptome data sets that are already available. By providing a resource consisting of transcriptome-wide RPF-seq/RNA-seq and TE plots (Supplemental Dataset 1), associated rhythmicity parameters (Supplemental Table S2), and high-confidence transcript lists from *Babel* analysis (Supplemental Table S3), we wish to facilitate the widespread use of our data. This resource provides a number of straightforward opportunities for exciting future endeavors, one example being the interesting cases of genes whose mRNAs, but not the footprint profiles, oscillate (Fig. 4A; Supplemental Fig. S6C), which we were unable to further investigate within the scope of this study.

Our study provides part of the explanation for the longstanding enigma that the mRNAs of many oscillating proteins show constant abundance over the day (Reddy et al. 2006). Two recent reports have estimated that 20% (Robles et al. 2014) to 50% (Mauvoisin et al. 2014) of protein rhythms are engendered by translation, protein degradation, or secretion. Our high-confidence set of approximately 150 nonoscillating mRNAs that undergo robust daily TE rhythms corresponds to ~8% of all detected rhythmically biosynthesized proteins. Considering the conservative selection criteria that were applied, the true extent of translationally driven rhythmicity may even be higher.

How does the translatome data correlate with the rhythmic proteome? The answer to this question is less straightforward than expected. A first complication lies in the poor overlap of the proteomics studies; although both report on almost 200 rhythmic proteins (Mauvoisin et al. 2014; Robles et al. 2014), <20% are shared between the studies, and the particularly interesting "mRNA flat–protein rhythmic" class has only three proteins in common (Supplemental Fig. S10A,B). As the overlap in the total detected proteome (about 3000–5000 polypeptides in both studies) is >50%, the differences seen in the rhythmic sets are not just a matter of proteome coverage. Differences in mass-spectrometric and, very likely, rhythmicity detection methodology may have caused the discrepancies; sophisticated meta-analyses on all available raw data using comparable algorithms and statistical parameters would thus be of great value. Another concern when comparing RPF-seq and proteome data sets is that MS data are inevitably biased for abundant (i.e., highly expressed and/or stable) polypeptides, whereas rhythmic TEs predominantly affected mRNAs whose abundance was below average (Fig. 4H). Many translationally regulated transcripts are hence not covered by the proteome data. It is reassuring that, despite such limitations, several translationally rhythmic transcripts are part of the rhythmic proteome (e.g., FTH1, EEF1A1, and EEF2 in study by Robles et al. 2014).

Transcripts encoding components of the protein biosynthesis machinery stand out among the rhythmically translated mRNAs. Their preferential association with polysomes at the light-to-dark transition has been reported before; it likely gates the energy-consuming ribosome biogenesis to the appropriate time when nutrients are plentiful, and involves regulatory cues from both feeding (via TORC1-regulated 5′-TOP motifs) and from the clock (Jouffe et al. 2013). As we observed increased TEs on these mRNAs already at ZT10 (Fig. 5D), i.e., ~2 h before the main surge in food intake in ad libitum fed animals (Adamovich et al. 2014), we consider it likely that the mechanism entails more than a simple, immediate reaction to nutrients. Moreover, the relatively variable up-regulation seen across biological replicates (Fig. 5D, RPF-seq error bars) is remarkable for genetically identical animals and could point to a behavioral component contributing to the regulatory mechanism. Interestingly, and reminiscent of the timing in the liver, increased ribosome association of mRNAs in *Drosophila* occurs at phases of relative behavioral quiescence, just prior to locomotor activity bouts (Huang et al. 2013). It remains to be explored whether this similarity is indicative of mechanistic parallels. Another exciting open question concerns the possibility that the rhythmic biosynthesis of components of the translational apparatus contributes to daily changes in overall translation rate that have been reported and that may involve mTOR signaling and a noncanonical cytoplasmic role for ARNTL (Lipton et al. 2015).

Among the other cases of TE rhythmicity, only a few were directly suggestive of an underlying mechanism, as was the case for mRNAs encoding iron metabolic proteins that all contain IREs. IREs are bound by iron regulatory proteins (IRPs) 1 and 2 (encoded by the genes *Aco1* and *Ireb2*, respectively), which sense intracellular iron levels by distinct mechanisms and respond to other metabolic signals as well (for review, see Anderson et al. 2012). IRP1 assembles a 4Fe-4S cluster in response to increased iron availability, which precludes IRE binding and permits translation. Other signals, such as NO, H_2O_2, and O_2, also influence the Fe-S cluster and IRP1 activity. IRP2 is regulated by protein degradation via FBXL5, an E3 ubiquitin ligase stabilized by iron and oxygen. We did not observe high-amplitude rhythms in the expression of IRPs or their known regulators (Supplemental Fig. S11; please note low amplitude rhythms for *Ireb2*), and it is conceivable that rhythmicity occurs at the level of available, bioactive iron in the hepatocyte (day/night changes in hepatic total iron have been reported) (Unger et al. 2009), of O_2 pressure/consumption (Peek et al. 2013), or of reactive oxygen species (Khapre et al. 2011). Together with the oscillations in mRNA abundance seen for multiple other iron metabolic genes, the rhythmic regulation of IRE-containing transcripts uncovers a previously unappreciated extent of temporal control in this physiologically important pathway.

It is noteworthy that clock genes showed constant TEs, indicating exclusion from time of day–dependent translational control. The considerable delays between mRNA and protein accumulation that have been reported for several core clock components (e.g., Lee et al. 2001) must therefore have other, post-translational origins. Nevertheless, our study has unveiled important insights into how translation contributes to core clock regulation. First, the CDS-mapping RPF-seq reads allow estimating relative biosynthesis rates of core clock proteins, which will likely add to a better quantitative understanding of the clock mechanism. Moreover, the footprint profiles from several clock mRNAs showed hallmarks of regulation that, however, may be operative not in a temporal fashion but under other (e.g., environmental, metabolic, cell-type–specific) conditions yet to be defined. In this context, the high number of translated uORFs within the core clock transcripts is particularly striking. uORF translation is generally viewed as inhibitory for protein production from the main ORF (Wethmar et al. 2014) and represents an attractive

mechanism for how clock protein levels (and consequently clock parameters) could be adjusted post-transcriptionally. It is tempting to speculate that one or several of the identified core clock uORFs are implicated in the short period phenotype observed in *Denr*-depleted cells. Of note, there is growing evidence for cell-type–specific uORF usage (e.g., Ingolia et al. 2011), and it is also largely unexplained how certain core clock parameters can be strikingly tissue specific (e.g., >2 h longer free-running period in kidney vs. lung) (Yoo et al. 2004). It is conceivable that cell-type–specific differences in clock protein concentration and/or stoichiometry are involved (Lee et al. 2011) and that tissue-specific uORF usage and translation rates contribute. Altogether, our results suggest that the circadian system represents a particularly suitable paradigm for future studies of uORF biology.

Methods

Animals

For time series experiments, 12-wk-old male mice (C57BL/6J; Janvier Labs) were entrained for 2 wk to LD 12:12 with free access to food and water and were euthanized at indicated *Zeitgeber* times (ZT0 corresponding to "lights on") by decapitation after anesthesia (isoflurane). Livers were removed and processed either directly or flash-frozen in liquid N_2. All experimental procedures were approved by the Veterinary Office of the Canton Vaud (authorization VD2376).

Ribosome profiling and RNA-seq

RPF-seq and RNA-seq libraries were generated using Ribo-Zero and ARTseq ribosome profiling kits (Epicentre) and sequenced on an Illumina HiSeq 2500. Detailed protocols, including for lysate preparation, are described in the Supplemental Material.

Bioinformatic analysis of ribosome profiling and RNA-seq

Adapter-trimmed, size-filtered sequencing reads (lengths 26–35 nt and 21–60 nt for RPF-seq and RNA-seq, respectively) were mapped sequentially to mouse rRNA, human rRNA, mt-tRNA, mouse tRNA, mouse cDNA (Ensembl release 75), and, finally, mouse genomic sequences (GRCm38.p2). cDNA-mapping reads were counted toward 5′ UTR, CDS, and 3′ UTR per gene basis. CDS counts were normalized by the upper quantile method and transformed into modified RPKM values. TEs were calculated as the ratio of RPF-RPKM to mRNA-RPKM. For detailed information on bioinformatic analysis, see the Supplemental Material.

Protein analyses

Total, nuclear, and mitochondrial protein extracts were prepared from two to three individual mice per timepoint and analyzed by SDS-PAGE and Western blotting according to standard protocols. Figures show one representative time series. Detailed experimental protocols and antibodies are described in the Supplemental Material.

Cloning

The generation of lentiviral luciferase reporter plasmids containing wild-type/mutant fragments of the *Nr1d1* genomic region is described in the Supplemental Material. For the generation of lentiviral shRNA expression vectors targeting *Denr*, sequences from the TRC shRNA Library at the Broad Institute were cloned into pLKO.1puro backbone vector (Addgene no. 10878) (Moffat et al. 2006); sequences/clones are listed in Supplemental Material. Scramble shRNA (Addgene no. 1864) served as control.

Cell culture

Cell culture, lentiviral production/transduction, the recording of circadian bioluminescence rhythms, and the dual luciferase assays followed standard methods. Detailed experimental protocols and additional references can be found in the Supplemental Material.

RT-qPCR

RNA was extracted with TriFast peqGOLD (PEQLAB), reverse-transcribed with SuperScript II (Invitrogen), and amplified with SYBR green rox master mix (Roche) and gene-specific primers (Supplemental Material) on a Stratagene Mx3000P apparatus (Agilent). Relative expression levels were determined using the ΔΔCt method.

GO analysis

GO analysis was carried out on the footprints rhythmic set after Babel analysis (Olshen et al. 2013) using the DAVID bioinformatics resource (Huang et al. 2009).

Data access

The sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE67305.

Acknowledgments

We thank staff at the Lausanne Genomic Technologies Facility for high-throughput sequencing support, Ioannis Xenarios and staff at Vital-IT for computational support, and Mara De Matos for technical assistance. We thank Felix Naef and Laura Symul for help with model selection algorithms. D.G. was supported by Swiss National Science Foundation (professorship grants 128399, 157528); National Center for Competence in Research (NCCR) RNA & Disease; Fondation Pierre Mercier; Fondation Leenaards; Olga Mayenfisch Stiftung; SystemsX.ch StoNets consortium; and the University of Lausanne. P.J. was supported by Human Frontiers Science Program long-term fellowship LT000158/2013-L.

References

Adamovich Y, Rousso-Noori L, Zwighaft Z, Neufeld-Cohen A, Golik M, Kraut-Cohen J, Wang M, Han X, Asher G. 2014. Circadian clocks and feeding time regulate the oscillations and levels of hepatic triglycerides. *Cell Metab* **19:** 319–330.

Anderson CP, Shen M, Eisenstein RS, Leibold EA. 2012. Mammalian iron metabolism and its control by iron regulatory proteins. *Biochim Biophys Acta* **1823:** 1468–1483.

Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. 2003. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **100:** 3889–3894.

Bugge A, Feng D, Everett LJ, Briggs ER, Mullican SE, Wang F, Jager J, Lazar MA. 2012. Rev-erbα and Rev-erbβ coordinately protect the circadian clock and normal metabolic function. *Genes Dev* **26:** 657–667.

Churbanov A, Rogozin IB, Babenko VN, Ali H, Koonin EV. 2005. Evolutionary conservation suggests a regulatory function of AUG triplets in 5′-UTRs of eukaryotic genes. *Nucleic Acids Res* **33:** 5512–5520.

Dana A, Tuller T. 2012. Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput Biol* **8:** e1002755.

Du NH, Arpat AB, De Matos M, Gatfield D. 2014. MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. *eLife* **3:** e02510.

Forman BM, Ruan B, Chen J, Schroepfer GJ Jr, Evans RM. 1997. The orphan nuclear receptor LXRα is positively and negatively regulated by distinct products of mevalonate metabolism. *Proc Natl Acad Sci* **94:** 10588–10593.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44–57.

Huang N, Chelliah Y, Shan Y, Taylor CA, Yoo SH, Partch C, Green CB, Zhang H, Takahashi JS. 2012. Crystal structure of the heterodimeric CLOCK: BMAL1 transcriptional activator complex. *Science* **337:** 189–194.

Huang Y, Ainsley JA, Reijmers LG, Jackson FR. 2013. Translational profiling of clock cells reveals circadianly synchronized protein synthesis. *PLoS Biol* **11:** e1001703.

Ingolia NT. 2014. Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* **15:** 205–213.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147:** 789–802.

Jouffe C, Cretenet G, Symul L, Martin E, Atger F, Naef F, Gachon F. 2013. The circadian clock coordinates ribosome biogenesis. *PLoS Biol* **11:** e1001455.

Khapre RV, Kondratova AA, Susova O, Kondratov RV. 2011. Circadian clock protein BMAL1 regulates cellular senescence in vivo. *Cell Cycle* **10:** 4162–4169.

Kojima S, Sher-Chen EL, Green CB. 2012. Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes Dev* **26:** 2724–2736.

Le Martelot G, Canella D, Symul L, Migliavacca E, Gilardi F, Liechti R, Martin O, Harshman K, Delorenzi M, Desvergne B, et al. 2012. Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles. *PLoS Biol* **10:** e1001442.

Lee C, Etchegaray JP, Cagampang FR, Loudon AS, Reppert SM. 2001. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* **107:** 855–867.

Lee Y, Chen R, Lee HM, Lee C. 2011. Stoichiometric relationship among clock proteins determines robustness of circadian rhythms. *J Biol Chem* **286:** 7033–7042.

Lipton JO, Yuan ED, Boyle LM, Ebrahimi-Fakhari D, Kwiatkowski E, Nathan A, Guttler T, Davis F, Asara JM, Sahin M. 2015. The circadian protein BMAL1 regulates translation in response to S6K1-mediated phosphorylation. *Cell* **161:** 1138–1151.

Mauvoisin D, Wang J, Jouffe C, Martin E, Atger F, Waridel P, Quadroni M, Gachon F, Naef F. 2014. Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. *Proc Natl Acad Sci* **111:** 167–172.

Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* **36:** 1073–1078.

Meyuhas O, Kahan T. 2015. The race to decipher the top secrets of TOP mRNAs. *Biochim Biophys Acta* **1849:** 801–811.

Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK, et al. 2006. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124:** 1283–1298.

Mohawk JA, Green CB, Takahashi JS. 2012. Central and peripheral circadian clocks in mammals. *Annu Rev Neurosci* **35:** 445–462.

Olshen AB, Hsieh AC, Stumpf CR, Olshen RA, Ruggero D, Taylor BS. 2013. Assessing gene-level translational control from ribosome profiling. *Bioinformatics* **29:** 2995–3002.

Peek CB, Affinati AH, Ramsey KM, Kuo HY, Yu W, Sena LA, Ilkayeva O, Marcheva B, Kobayashi Y, Omura C, et al. 2013. Circadian clock NAD$^+$ cycle drives mitochondrial oxidative metabolism in mice. *Science* **342:** 1243417.

Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U. 2002. The orphan nuclear receptor REV-ERBα controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110:** 251–260.

Reddy AB, Karp NA, Maywood ES, Sage EA, Deery M, O'Neill JS, Wong GK, Chesham J, Odell M, Lilley KS, et al. 2006. Circadian orchestration of the hepatic proteome. *Curr Biol* **16:** 1107–1115.

Robles MS, Cox J, Mann M. 2014. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet* **10:** e1004047.

Schleich S, Strassburger K, Janiesch PC, Koledachkina T, Miller KK, Haneke K, Cheng YS, Kuchler K, Stoecklin G, Duncan KE, et al. 2014. DENR–MCT-1 promotes translation re-initiation downstream of uORFs to control tissue growth. *Nature* **512:** 208–212.

Skabkin MA, Skabkina OV, Hellen CU, Pestova TV. 2013. Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol Cell* **51:** 249–264.

Stratmann M, Stadler F, Tamanini F, van der Horst GT, Ripperger JA. 2010. Flexible phase adjustment of circadian *albumin D site-binding protein* (*Dbp*) gene expression by CRYPTOCHROME1. *Genes Dev* **24:** 1317–1328.

Stratmann M, Suter DM, Molina N, Naef F, Schibler U. 2012. Circadian *Dbp* transcription relies on highly dynamic BMAL1-CLOCK interaction with E boxes and requires the proteasome. *Mol Cell* **48:** 277–287.

Unger EL, Earley CJ, Beard JL. 2009. Diurnal cycle influences peripheral and brain iron levels in mice. *J Appl Physiol* **106:** 187–193.

Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13:** 227–232.

Vollmers C, Gill S, DiTacchio L, Pulivarthy SR, Le HD, Panda S. 2009. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proc Natl Acad Sci* **106:** 21453–21458.

Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A. 2014. uORFdb: a comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Res* **42(**Database issue**):** D60–D67.

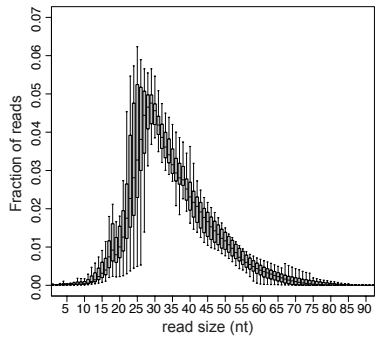
Yip L, Su L, Sheng D, Chang P, Atkinson M, Czesak M, Albert PR, Collier AR, Turley SJ, Fathman CG, et al. 2009. Deaf1 isoforms control the expression of genes encoding peripheral tissue antigens in the pancreatic lymph nodes during type 1 diabetes. *Nat Immunol* **10:** 1026–1033.

Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, Siepka SM, Hong HK, Oh WJ, Yoo OJ, et al. 2004. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc Natl Acad Sci* **101:** 5339–5346.

Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. 2014. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc Natl Acad Sci* **111:** 16219–16224.

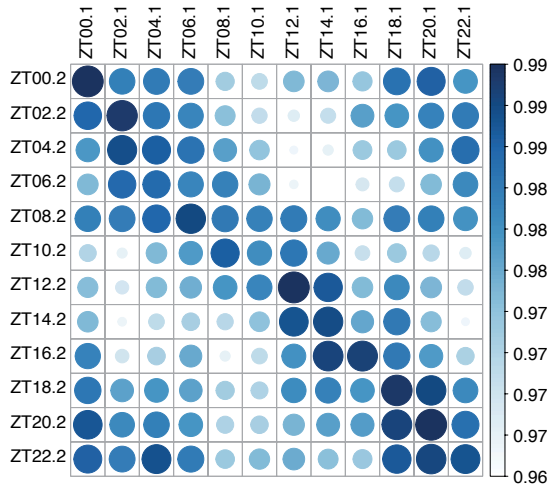
Received June 4, 2015; accepted in revised form October 14, 2015.

A



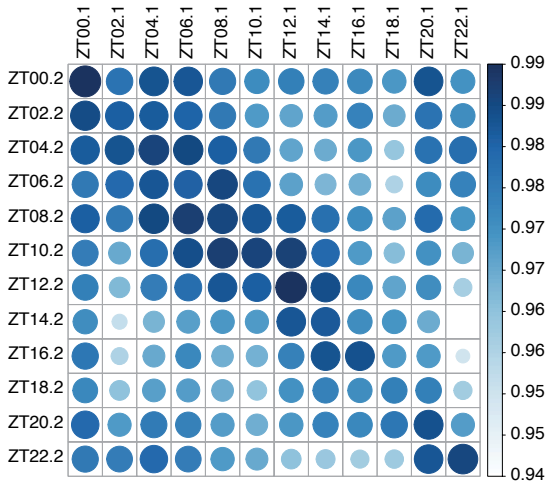
B

RPF-seq

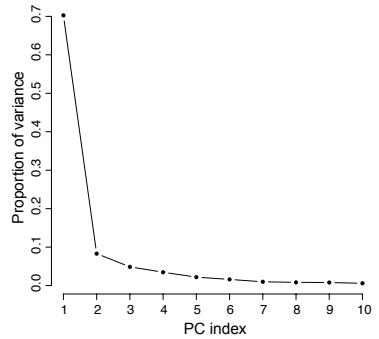


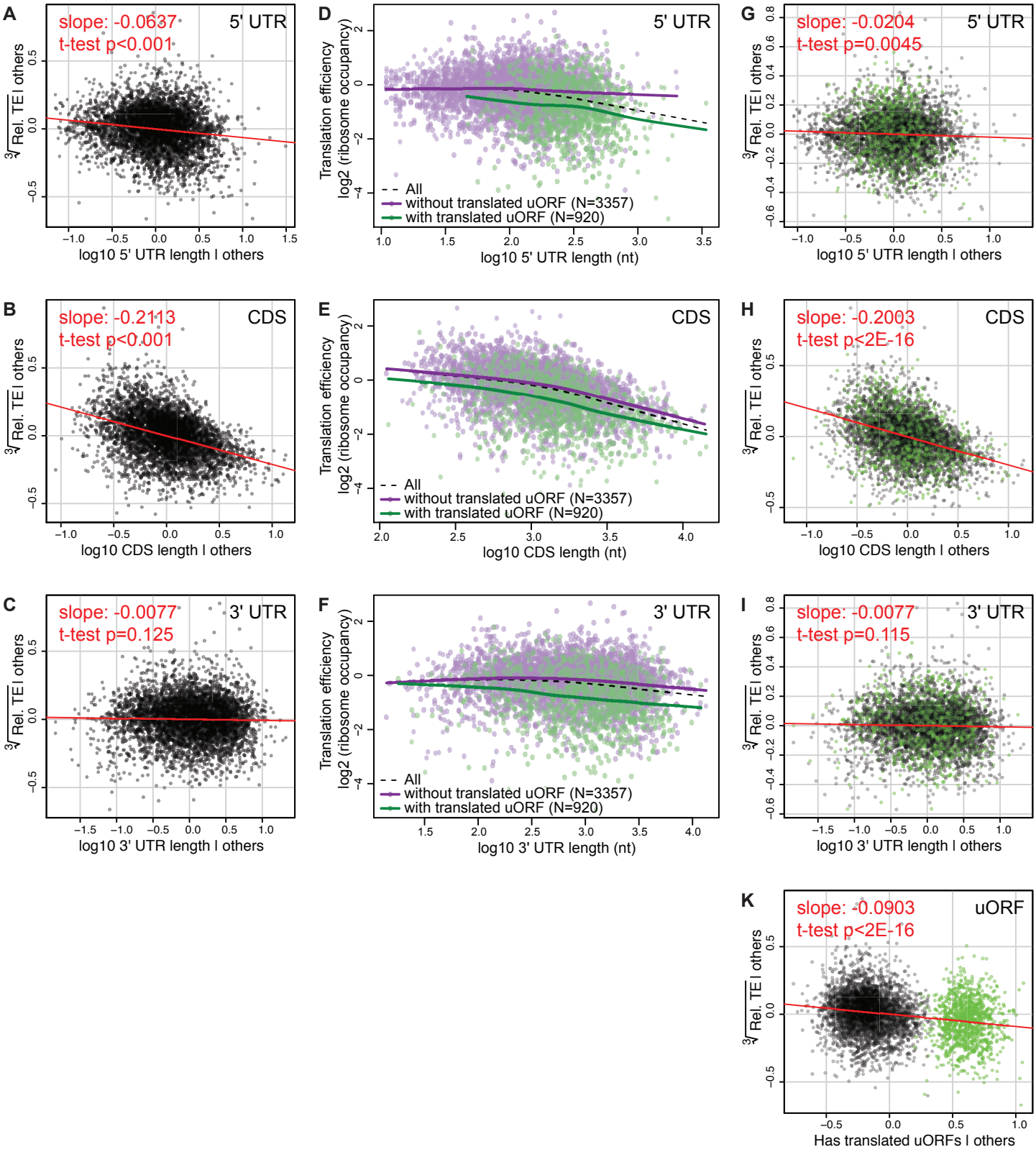
C

RNA-seq

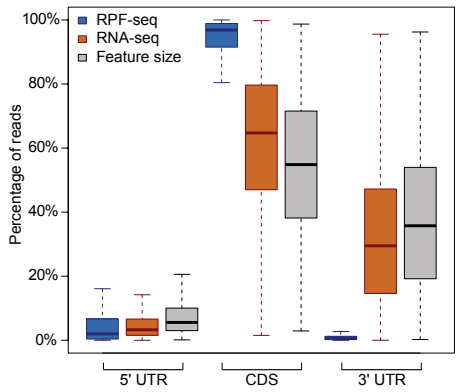


D

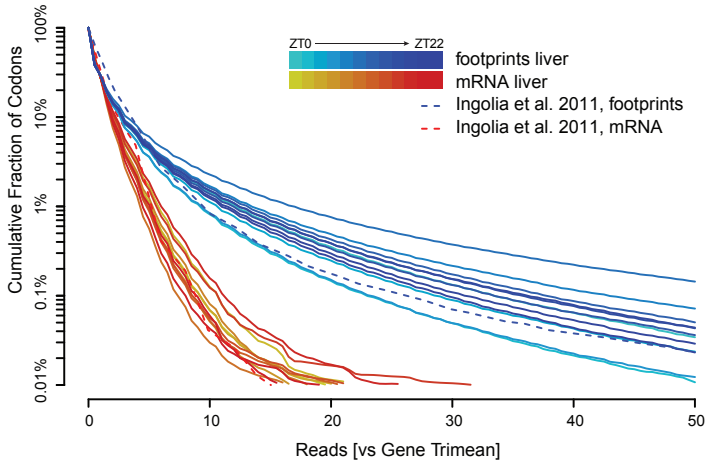



A
slope: -0.0637
t-test p<0.001
5' UTR
Rel. TE | others
log10 5' UTR length | others
D
5' UTR
Translation efficiency
log2 (ribosome occupancy)
All
without translated uORF (N=3357)
with translated uORF (N=920)
log10 5' UTR length (nt)
G
slope: -0.0204
t-test p=0.0045
5' UTR
Rel. TE | others
log10 5' UTR length | others
B
slope: -0.2113
t-test p<0.001
CDS
Rel. TE | others
log10 CDS length | others
E
CDS
Translation efficiency
log2 (ribosome occupancy)
All
without translated uORF (N=3357)
with translated uORF (N=920)
log10 CDS length (nt)
H
slope: -0.2003
t-test p<2E-16
CDS
Rel. TE | others
log10 CDS length | others
C
slope: -0.0077
t-test p=0.125
3' UTR
Rel. TE | others
log10 3' UTR length | others
F
3' UTR
Translation efficiency
log2 (ribosome occupancy)
All
without translated uORF (N=3357)
with translated uORF (N=920)
log10 3' UTR length (nt)
I
slope: -0.0077
t-test p=0.115
3' UTR
Rel. TE | others
log10 3' UTR length | others
K
slope: -0.0903
t-test p<2E-16
uORF
Rel. TE | others
Has translated uORFs | others

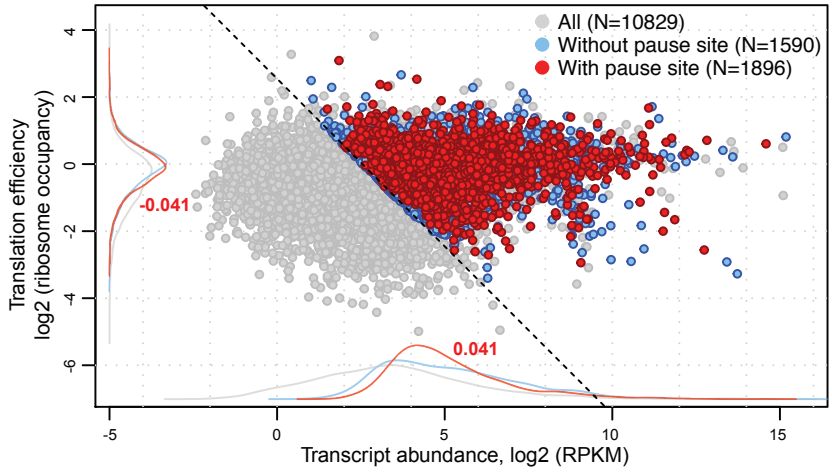
A

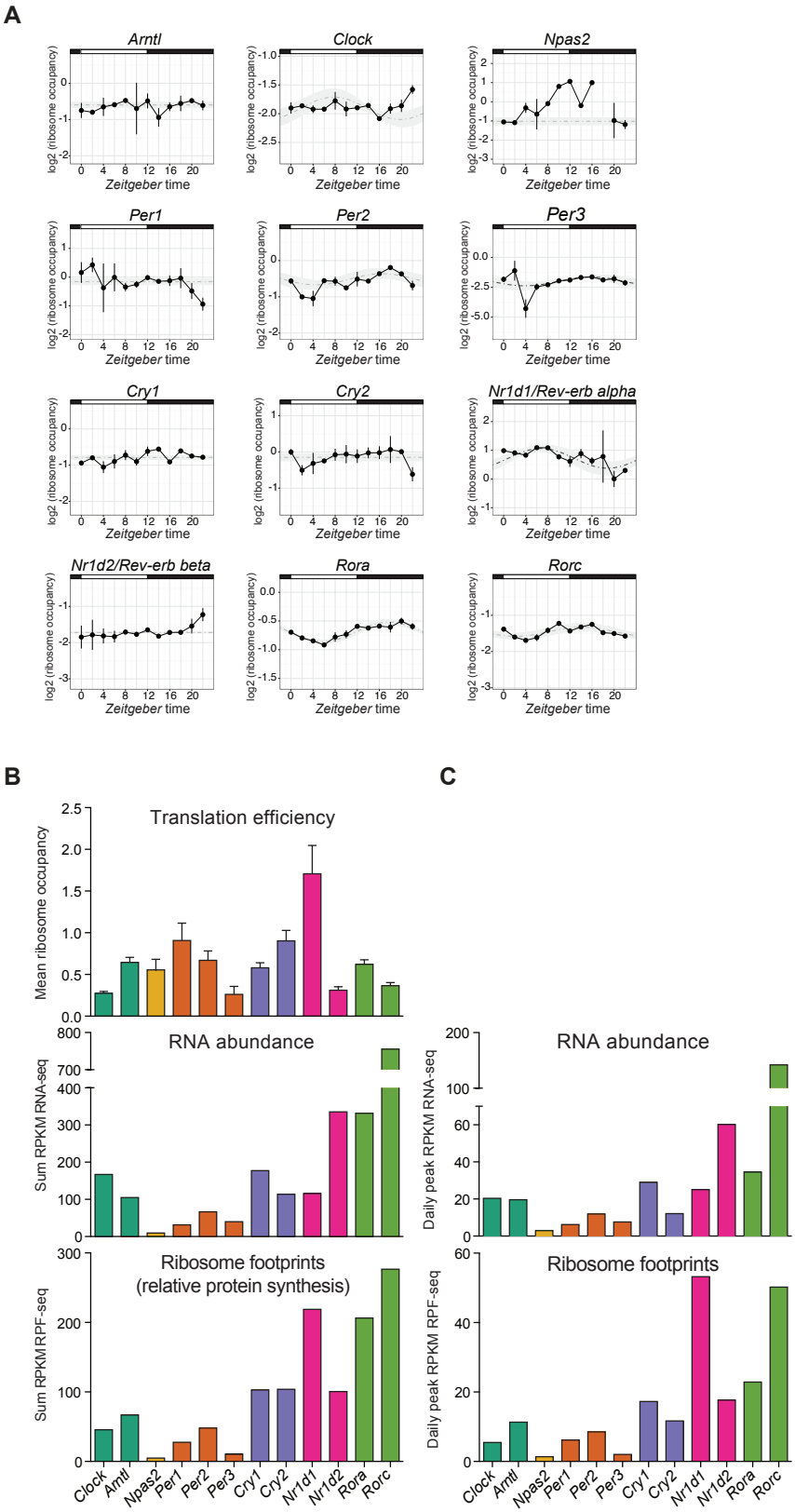

RPF-seq
RNA-seq
Feature size
Percentage of reads
100%
80%
60%
40%
20%
0%
5' UTR
CDS
3' UTR

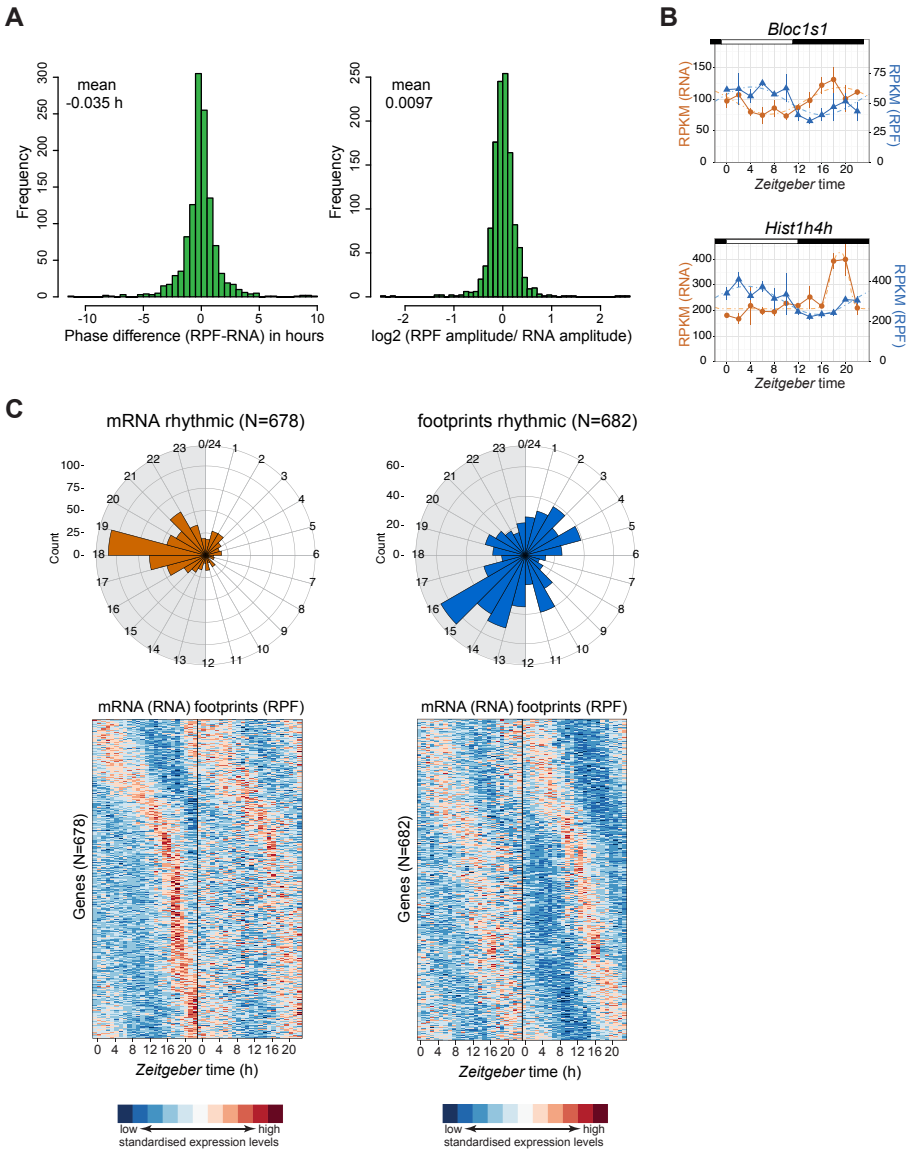
A


ZT0
ZT22
footprints liver
mRNA liver
Ingolia et al. 2011, footprints
Ingolia et al. 2011, mRNA
Cumulative Fraction of Codons
100%
10%
1%
0.1%
0.01%
0
10
20
30
40
50
Reads [vs Gene Trimean]

B


All (N=10829)
Without pause site (N=1590)
With pause site (N=1896)
Translation efficiency
log2 (ribosome occupancy)
4
2
0
-2
-4
-6
-0.041
0.041
-5
0
5
10
15
Transcript abundance, log2 (RPKM)

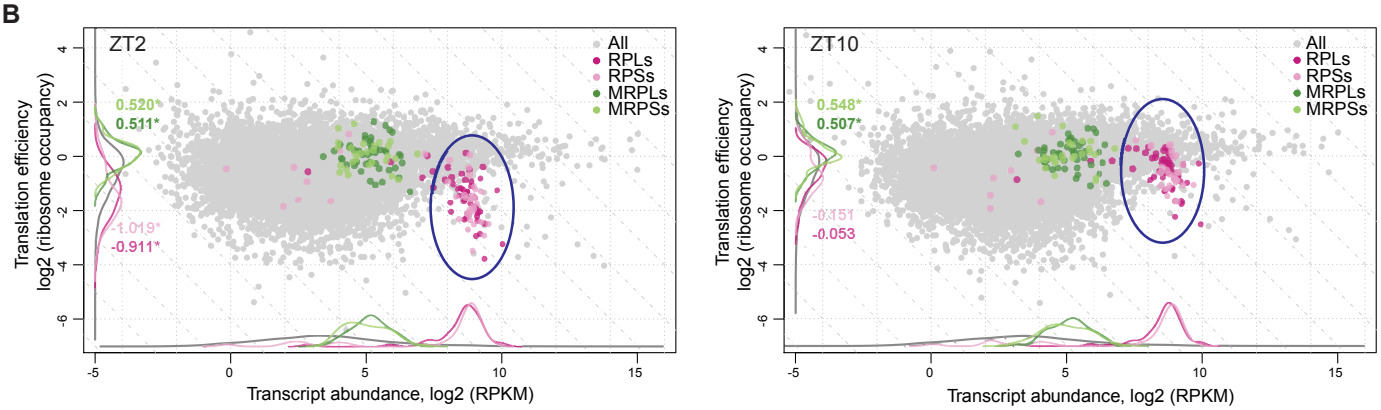

A
Arntl
Clock
Npas2
Per1
Per2
Per3
Cry1
Cry2
Nr1d1/Rev-erb alpha
Nr1d2/Rev-erb beta
Rora
Rorc
log2 (ribosome occupancy)
Zeitgeber time
B
Translation efficiency
Mean ribosome occupancy
RNA abundance
Sum RPKM RNA-seq
Ribosome footprints
(relative protein synthesis)
Sum RPKM RPF-seq
C
RNA abundance
Daily peak RPKM RNA-seq
Ribosome footprints
Daily peak RPKM RPF-seq
Clock
Arntl
Npas2
Per1
Per2
Per3
Cry1
Cry2
Nr1d1
Nr1d2
Rora
Rorc

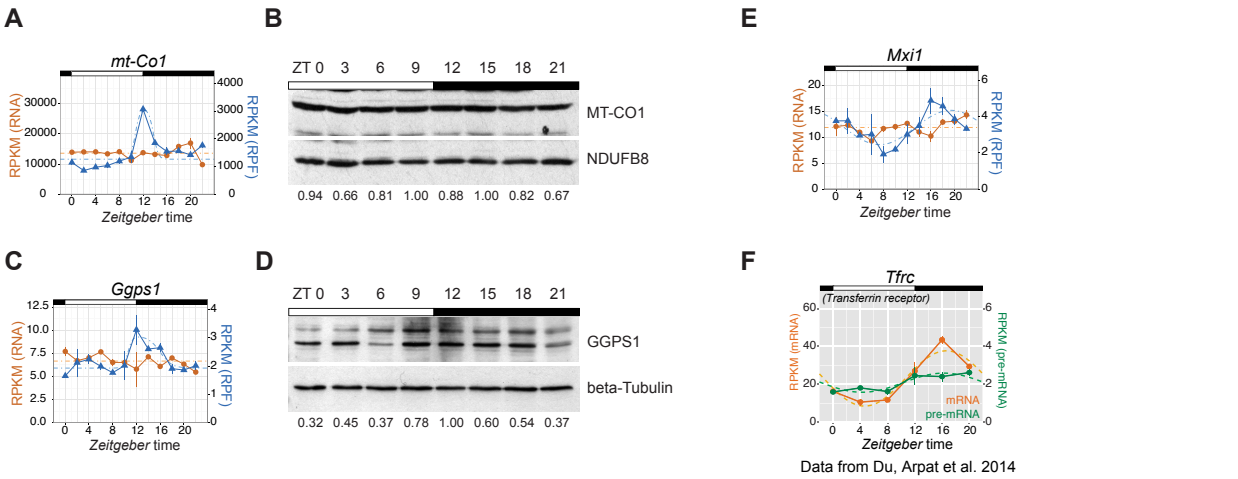

A
mean
-0.035 h
Frequency
Phase difference (RPF-RNA) in hours
mean
0.0097
log2 (RPF amplitude/ RNA amplitude)
B
Bloc1s1
RPKM (RNA)
RPKM (RPF)
Zeitgeber time
Hist1h4h
C
mRNA rhythmic (N=678)
footprints rhythmic (N=682)
Count
mRNA (RNA) footprints (RPF)
Genes (N=678)
Genes (N=682)
Zeitgeber time (h)
low
high
standardised expression levels

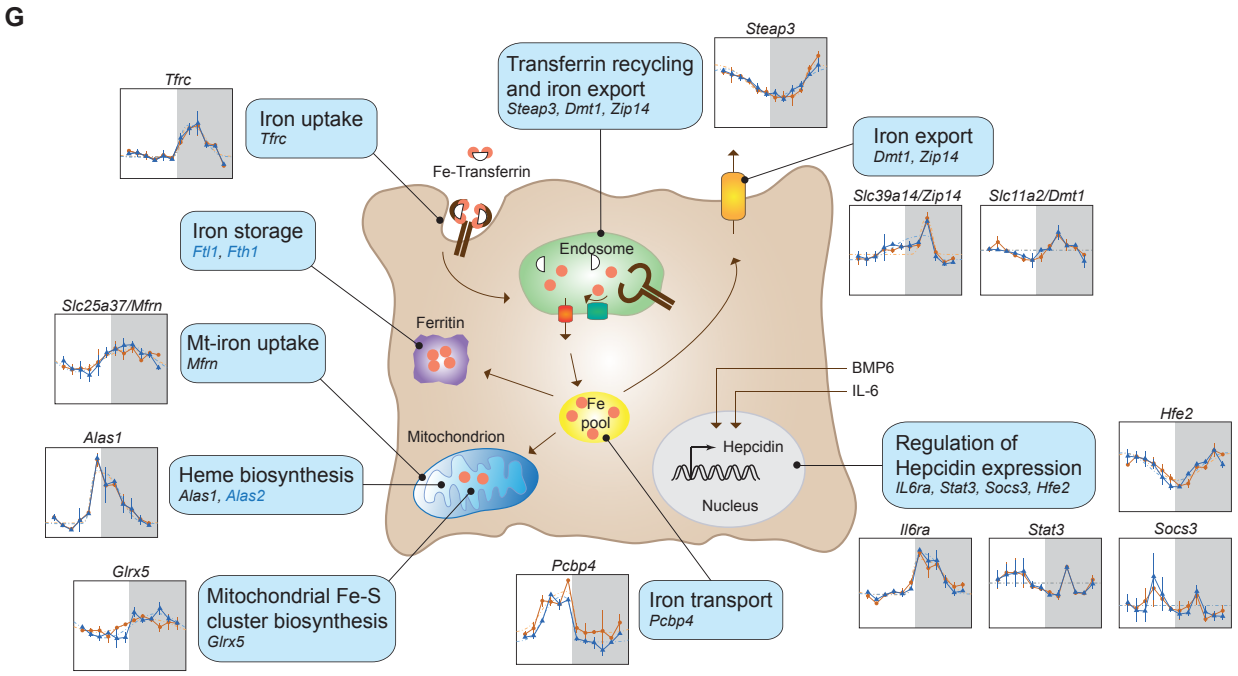
A

Ribosomal subunit	Gene name
40S subunit	*Rpsa, Rps2, Rps3,* ***Rps3a1****,* ***Rps4x****,* ***Rps5****,* ***Rps6****, Rps7,* ***Rps8****, Rps9,* ***Rps10****, Rps11,* ***Rps12****, Rps13, Rps14, Rps15, Rps15a,* ***Rps16****,* ***Rps17****,* ***Rps18****,* ***Rps19****,* ***Rps20****, Rps21, Rps23, Rps24,* ***Rps25****, Rps26, Rps27,* ***Rps27a****,* ***Rps28****, Rps29, Rps30*
60S subunit	*Rpl3, Rpl4,* ***Rpl5****,* ***Rpl6****,* ***Rpl7****,* ***Rpl7a****, Rpl8,* ***Rpl9****,* ***Rpl10****, Rpl10a,* ***Rpl11****, Rpl12, Rpl13, Rpl13a, Rpl14, Rpl15, Rpl17, Rpl18, Rpl18a, Rpl19, Rpl21,* ***Rpl22****,* ***Rpl23****,* ***Rpl23a****,* ***Rpl24****,* ***Rpl26****, Rpl27,* ***Rpl27a****,* ***Rpl28****, Rpl29,* ***Rpl30****,* ***Rpl31****,* ***Rpl32****, Rpl34, Rpl35, Rpl35a, Rpl36,* ***Rpl36a****,* ***Rpl37****, Rpl37a, Rpl38, Rpl39, Rpl40, Rpl41, Rplp0, Rplp1,* ***Rplp2***

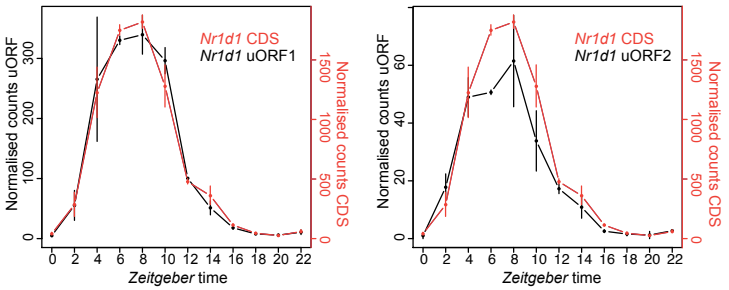
B




A
mt-Co1
RPKM (RNA)
RPKM (RPF)
Zeitgeber time
B
ZT 0 3 6 9 12 15 18 21
MT-CO1
NDUFB8
0.94 0.66 0.81 1.00 0.88 1.00 0.82 0.67
E
Mxi1
RPKM (RNA)
RPKM (RPF)
Zeitgeber time
C
Ggps1
RPKM (RNA)
RPKM (RPF)
Zeitgeber time
D
ZT 0 3 6 9 12 15 18 21
GGPS1
beta-Tubulin
0.32 0.45 0.37 0.78 1.00 0.60 0.54 0.37
F
Tfrc
(Transferrin receptor)
RPKM (mRNA)
RPKM (pre-mRNA)
mRNA
pre-mRNA
Zeitgeber time
Data from Du, Arpat et al. 2014


G
Tfrc
Iron uptake
Tfrc
Transferrin recycling and iron export
Steap3, Dmt1, Zip14
Steap3
Iron export
Dmt1, Zip14
Fe-Transferrin
Slc39a14/Zip14
Slc11a2/Dmt1
Iron storage
Ftl1, Fth1
Endosome
Slc25a37/Mfrn
Ferritin
Mt-iron uptake
Mfrn
BMP6
IL-6
Fe pool
Alas1
Mitochondrion
Heme biosynthesis
Alas1, Alas2
Hepcidin
Nucleus
Regulation of Hepcidin expression
IL6ra, Stat3, Socs3, Hfe2
Hfe2
Il6ra
Stat3
Socs3
Glrx5
Mitochondrial Fe-S cluster biosynthesis
Glrx5
Pcbp4
Iron transport
Pcbp4

A

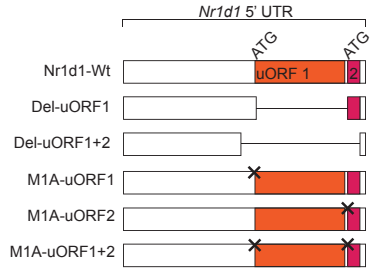


B

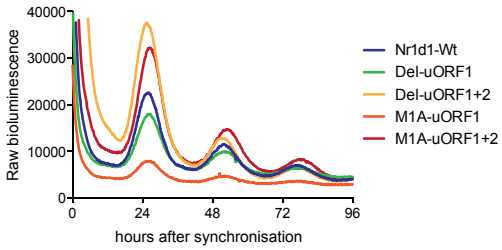
Nr1d1 uORF1

```
M. musculus          MQIRSQVPQRPTLLEGCSIATPRPYCSVPGILIANCGAHSSPSAIAQGAREAITTSSLCQGPV
R. norvegicus        MQIRSQVPQRPTLLEGCSIATPRPYCSVPGILIANCGAHSSPSAIAQGAREAITTSSLCQGPV
H. sapiens           MQSRTPLPQRPTLLQGCSPATPRPCCRVLRILIVSRGVHSPPLASAQGATAAIATSSLSQGPV
P. troglodytes       MQSRTPLPQRPTLLQGCSPATPRPCCRVLRILIVSRGVHSPPLASAQGATAAIATSSLSQGPV
M. fascicularis      MQSRTPLPQRPTLLQGCSPATPRPCCRVLRILIVSRGVLSPPLASAQGATAVIATSSLSQGPV
E. caballus          MQSRPPLPQRPTLLQGCSSATPRPCCRVPRIPIVSRGVHSPPSASAQGATAAIATSSLCQGPV
```

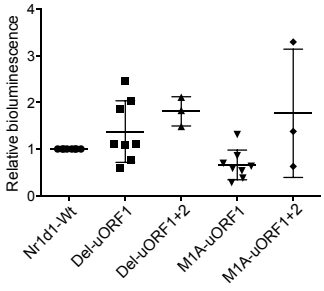
C



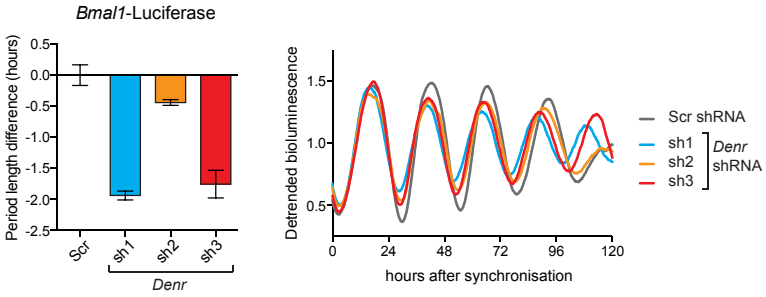
D



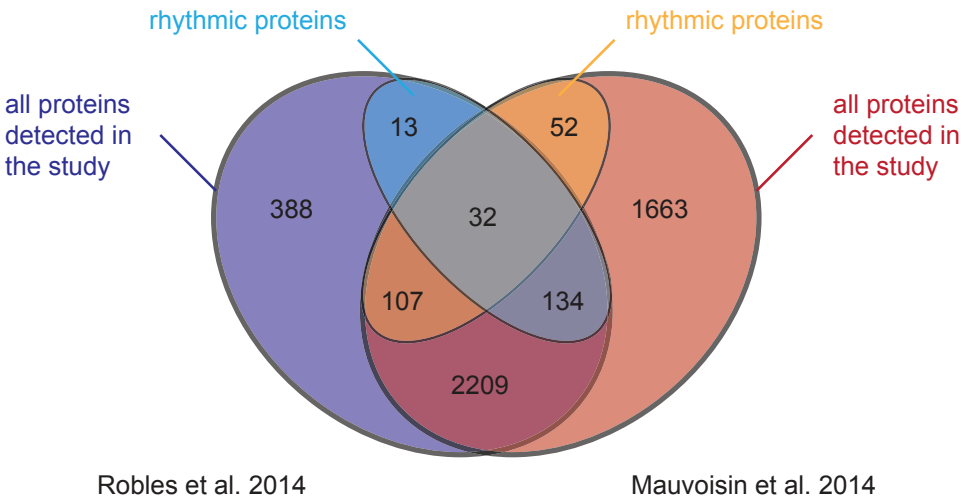
E



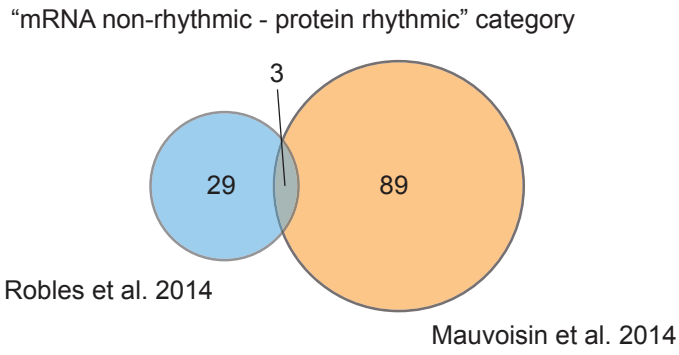
F

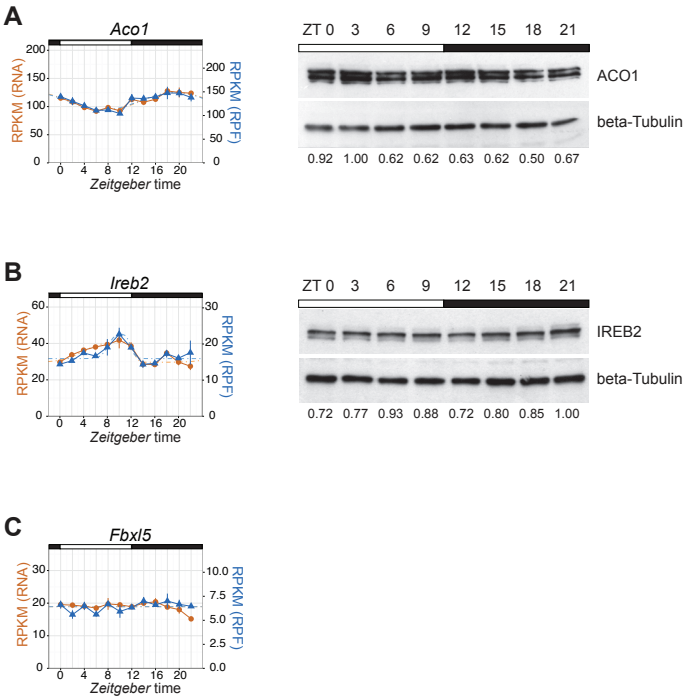


A


rhythmic proteins
rhythmic proteins
all proteins detected in the study
all proteins detected in the study
13
52
388
32
1663
107
134
2209
Robles et al. 2014
Mauvoisin et al. 2014

B


“mRNA non-rhythmic - protein rhythmic” category
3
29
89
Robles et al. 2014
Mauvoisin et al. 2014


A
Aco1
RPKM (RNA)
RPKM (RPF)
Zeitgeber time
ZT 0 3 6 9 12 15 18 21
ACO1
beta-Tubulin
0.92 1.00 0.62 0.62 0.63 0.62 0.50 0.67
B
Ireb2
RPKM (RNA)
RPKM (RPF)
Zeitgeber time
ZT 0 3 6 9 12 15 18 21
IREB2
beta-Tubulin
0.72 0.77 0.93 0.88 0.72 0.80 0.85 1.00
C
Fbxl5
RPKM (RNA)
RPKM (RPF)
Zeitgeber time

SUPPLEMENTAL FIGURE LEGENDS

Supplementary Figure S1. Ribosome profiling and RNA-seq in mouse liver.
A) Insert size distribution of RNA-seq reads across all replicates and timepoints. Box-and-whisker plots: midline, median; box, 25th and 75th percentiles; whiskers, extend to the minimum and maximum values within 1.5 times the interquartile range from the box. While most RNA-seq reads are in the range of around 30 nt (i.e., similar to RPF-seq reads), the distribution is broader, as expected for chemically fragmented RNA. This graph is complementary to the RPF-seq distribution in main Fig. 1E.
B) Intra- and inter-group relationships between normalised CDS counts of RPF-seq libraries are visualised with a heat map of Spearman correlation coefficients. Size and dark shading of disks increase with increasing correlation coefficients; colour correspondence is given in the legend. Average Spearman correlation for biological replicates: 0.987 ± 0.004 (average ± SD).
C) Same as B), for RNA-seq counts. Average Spearman correlation for biological replicates: 0.985 ± 0.004 (average ± SD).
D) Scree plot of the PCA shown in main Fig. 1H. Components beyond the first three PCs shown in Fig. 1H only explain a small percentage of additional variation in the data.

Supplementary Figure S2: Transcript length features and uORF usage as predictors of translation efficiency.
A-C) Partial regression plots for the modelling of TE on the three variables, 5’ UTR (A), CDS (B) and 3’ UTR (C) lengths. Such partial regression plots are useful to investigate the relationship between a predictor and the response variable taking into account the effect of the other predictors. Both 5’ UTR and CDS lengths show significant predictive value for the efficiency at which the transcripts’ main ORFs are translated. The predictive value of 3’ UTR length for TE is, however, not significant in this linear regression analysis. The statistically significant correlation between 3’ UTR length and TE that was originally observed in main Fig. 2B therefore likely results from the correlation between 3’ UTR with CDS length. Analysis was performed on transcripts from single protein isoform genes with an RPKM value >5 and 5’ and 3’ UTR lengths ≥10 nt (N=4277). In red, the slope of the red regression line and the significance level (t-test) are reported.

D) uORF translation correlates with low TE independently of 5’ UTR length. The scatter plot of translation efficiency vs. 5’ UTR length shows that transcripts without translated uORF (violet) have globally shorter 5’ UTRs than those with translated uORFs (green). Nevertheless, the locally weighted fits (Lowess) indicate that even for transcripts with UTRs of similar length, uORF-containing transcripts have lower TEs (i.e., the green line is below the violet line throughout). The effect of low TEs is thus an effect of uORF translation and not merely of 5’ UTR length. Same transcript set as in (A-C) was used.

E) Same analysis as D) for CDS length. The lower TEs of uORF-containing transcripts is seen independently of CDS length (i.e., the green line is below the violet line throughout).

F) Same analysis as D) for 3’ UTR length. The lower TEs of uORF-containing transcripts is seen independently of 3’ UTR length (i.e., the green line is below the violet line throughout).

G-K) Partial regression plots for the modelling of TE on the four variables, 5’ UTR (G), CDS (H), 3’ UTR (I) lengths and uORF presence (K). uORF-containing transcripts are depicted in green (N=920) and other in black (N=3357). uORF presence is thus a strong predictor of reduced TE (K). When compared to the three predictor model (A-C) it becomes clear that with the inclusion of uORF translation as a forth predictor, the 5’ UTR length lost much of its predictive value. This indicates that to a large extent the 5’ UTR effect seen in A) is in reality a uORF effect, whereby longer 5’ UTRs will be more likely uORF-containing than short 5’ UTRs. Given that relatively conservative uORF detection criteria were applied (e.g. only AUG-initiated uORFs with a certain minimal coverage) it is possible that 5’ UTR length *per se* has even lower (or no) predictive value for TE.

Supplementary Figure S3: Read distributions to 5’ UTR, CDS and 3’ UTR per transcript.

A) Read distribution within 5’ UTRs, CDS and 3’ UTRs of RPF-seq (blue) and RNA-seq data (orange) compared to the distribution expected by chance, which is determined by the feature sizes (grey). Box-and-whisker plots: midline, median; box, 25^{th} and 75^{th} percentiles; whiskers, extend to the minimum and maximum values within 1.5 times the interquartile range from the box. Note that 5’ UTR-mapping RPF-seq reads make up on average 6% of all reads (Fig. 1D), but that the actual percentage *per transcript* is variable (N=10829).

Supplementary Figure S4. Pause site occurrence is not correlated with TE.

A) Distribution of per-codon RPF-seq and RNA-seq counts. The plot shows the cumulative distribution of footprint (blue) counts at each codon, relative to the median density across the gene. In analogy to the method previously applied and published by (Ingolia et al. 2011) (see Fig. 2A in the cited publication), stall sites were defined at an arbitrary threshold for individual codons that exceeded read counts that are 25 x the median on the transcript (in our analysis we used the trimean instead of the median; see Supplemental Material). RNA-seq counts (red) control effects of library generation. The twelve blue and red curves represent the means of the 12 timepoints. For B), only pause sites were further considered when they appeared in the RPF-seq counts of both replicates of a timepoint. The dotted lines show the data from (Ingolia et al. 2011) for comparison and indicated that the overall pause site distribution is similar in liver and mESCs.

B) Pause sites are not associated with TE changes. Transcripts that did (red) or did not (blue) contain pause sites detected in A) were plotted according to their mRNA abundance (abscissa) and their TE (ordinate). Density curves for highlighted data points are plotted on the margins with the same colour code. Red numbers on density curves reflect the location shift (log2 values) of the transcripts with pause sites relative to those without pause sites; neither shift was significant at the 0.05 level (Wilcoxon rank sum test). A cut-off on RPF-seq RPKMs was applied – shown by the dotted diagonal – because for very low footprint density the probability to detect 25 x median levels is very low and would lead to false-negative pause site detection due to coverage effects. The applied cut-off corresponds to the minimal RPKM at which pause sites were still detectable.

Supplementary Figure S5. Constant, but widely differing TEs in clock transcripts.

A) Time-resolved TE plots for the core clock transcripts. This graph is complementary to main Fig. 3B and shows the ratios of RPF-seq/RNA-seq RPKMs around-the-clock. Means are plotted with error bars connecting the two biological replicates per timepoint. Grey dotted lines show a harmonic fit to the data points, and grey shading the 95% confidence interval of the fit. A fit was performed irrespective of whether rhythmicity was significant or not (when no fit was possible, e.g. for *Npas2* due to extremely low expression at ZT18, a straight line was plotted). Dark grey curves (e.g. for *Nr1d1*, *Per3*) indicated that the fit itself was significant, whereas light grey curves indicated non-significant harmonic fits or a fit with a <1.5 x peak-to-trough amplitude. Importantly, the meaning of a “significant fit” to TEs should not be

over-interpreted as it does *not* consider if the TEs were significant between timepoints from the start – this important prerequisite was tested using the more sophisticated *Babel* analysis (Olshen et al. 2013). Indeed, none of the core clock transcripts (including *Nr1d1* and *Per3*, for which the harmonic fits themselves were significant) showed significant TE changes according to *Babel* analysis (Supplemental Table S3). It can therefore be concluded that core clock TEs were not subject to time-of-day-dependent control.

B) Bar graph in the top panel shows the mean ± SD of translation efficiencies over the day. Middle and lower panel show the sum of RNA abundance and ribosome footprint RPKMs integrated over-the-day, respectively, for the circadian core clock genes listed below. Ribosome footprints (as RPKMs) are a direct measure of protein biosynthesis rate and can thus be used to calculate the relative daily protein production of the core clock components. Large differences that are seen at the level of RNA abundance (e.g. *Nr1d1* and *Nr1d2*) are thus compensated by TE, leading to protein output that explains some of the core clock biology better than the RNA levels do. See text for details.

C) RNA abundance and ribosome footprint plots as in B), but using the peak expression timepoints only.

Figure S6. Relationships between RPF and RNA rhythmicity parameters.

A) Evaluation of phase differences (left panel) and amplitude differences (right panel) between ribosome occupancy (RPF) and RNA abundance (RNA) of ‘mRNA and footprints rhythmic’ genes. The deviations of the means from 0 are not significant, indicating synchrony in RNA abundance and footprint rhythms for the vast majority of transcripts in the overlapping set.

B) Distinct out-of-phase translation was confined to few exceptional cases, which were initially identified by overlaying the ‘mRNA and footprints rhythmic’ set with the list of transcripts for which the *Babel* analysis had revealed significant TE changes between timepoints (Supplemental Table S3). The resulting 93 genes – many of which showed temporal changes in TE, but similar phase in RNA-seq and RPF-seq profiles – were further examined by visual inspection to identify out-of-phase translation events. Shown are the two most striking cases, Biogenesis of lysosomal organelles complex-1 subunit 1 *(Bloc1s1*) and Histone cluster 1, H4h *(Hist1h4h),* which both showed remarkable out-of-phase peaks in their mRNA and footprint profiles. Means per timepoint are plotted with error bars depicting the replicates.

C) Phase distribution (top panels) and heat maps (bottom panels) of genes identified as rhythmic only at the mRNA level (left) or only at the footprint level (right). mRNA

abundances and footprint levels are standardised within each gene. Note that many of these supposedly non-rhythmic transcripts still had underlying, yet noisier rhythms that apparently escaped the detection algorithm – a problem that prompted us to use the *Babel* method with the aim of obtaining a high-confidence set of "mRNA flat – footprints rhythmic" transcripts (see main text and Fig. 5A).

Figure S7. Rhythmic control of ribosomal protein biosynthesis.

A) Table summarising the genes encoding ribosomal proteins of the 40S and 60S ribosomal subunits. Highlighted are those genes that were either detected as "mRNA flat – footprints rhythmic" in main Fig. 4A (bold print), or that had a similar profile as those in Fig. 4A (as judged by visual inspection) but were below the amplitude cut-off of >1.5 fold (normal print), or were not rhythmic, or not detected in our datasets altogether (grey print). The vast majority of *Rps* and *Rpl* transcripts thus undergoes similar regulation at the translational level.

B) Scatter plot of translation efficiencies and transcript abundances for *Rpl* and *Rps* genes of the cytoplasmic (pink) or mitochondrial (green; used as control) ribosome for timepoints ZT2 and ZT10 (small subunit proteins are shown in lighter shadings; large subunit proteins in darker shadings). Density curves of translation efficiencies and transcript abundances for highlighted data points are plotted on the margins with the same colour code. Numbers on density curves reflect the location shift (log2 values) relative to all genes (grey); a (*) denotes significance at the 0.05 level (Wilcoxon rank sum test). The two plots visualise the shift in translation efficiency of mRNAs of cytoplasmic ribosomal proteins between ZT2 and ZT10. Wilcoxon rank sum test indicated that only RPs (RPLs+ RPSs) changed their TEs (location shift=0.864, p=3.016e-10), whereas MRPs (MRPLs+MRPSs) remained unchanged (location shift=0.009, p=0.916) between ZT2 and ZT10.

See also Supplemental Movie M1.

Figure S8. Rhythmicity in iron metabolic genes and validation of other cases of rhythmic translation.

A) Mitochondria encode 13 protein-coding genes, of which the expression of 12 (exception: *mt-Nd4l*) was detectable in our datasets and all showed a characteristic translational spike at ZT12 (Supplemental Dataset 1). As an example, this graph shows RPKM-normalised mRNA abundance (orange) and footprint (blue) profiles for *mt-Co1* around the 24-hour daily cycle.

B) Western blot analysis for MT-CO1 and NDUFB8 proteins (loading control) in liver mitochondrial extracts. Numbers show relative MT-CO1 protein levels across different

timepoints after normalisation to NDUFB8 protein levels. Data from one representative time series are shown (N=3). Note that despite the translational peak at ZT12, no protein rhythmicity was seen. An explanation may be the reported high stability of mitochondrial proteins (Kim et al. 2012). Nevertheless, one may speculate that the rhythmic translation has a role in assuring the coordinated biosynthesis of the individual proteins, thus enabling an efficient assembly of newly synthesised mitochondrial complexes.

C) RPKM-normalised mRNA abundance (orange) and footprints (blue) profiles for *Ggps1* around the 24-hour daily cycle.

D) Western blot analysis of GGPS1 was performed in total liver extracts (N=2). The sum of the signals from both bands was used for quantification. Beta-Tubulin served as a loading control and for normalisation. Note that rhythmic translation led to rhythmic protein accumulation.

E) Profiles for *Mxi1* similar as in A).

F) Rhythmic regulation of transferrin receptor (*Tfrc*) mRNA stability around the clock. The graph shows the temporal profile of pre-mRNA (a proxy for transcription rate; green) and mRNA abundance (orange) profiles. Note that the mRNA is rhythmic, but not the pre-mRNA, indicating that oscillations are engendered at the level of mRNA stability. Extracted from the RNA-seq data in (Du et al. 2014); error bars represent the two replicates of the time series. Of note, post-transcriptional rhythmicity of *Tfrc* mRNA was independently confirmed by (Le Martelot et al. 2012).

G) Schematic illustrating the various steps involved in the maintenance of cellular iron homeostasis. The small gene expression graphs show mRNA abundance (orange) and footprint profiles (blue) of rhythmically expressed genes associated with key regulatory processes of iron metabolism. Gene names highlighted in blue mark transcripts with rhythmic footprints shown in Fig. 5G. Moreover, marked rhythms in mRNA abundance were thus seen at the level of regulators of hepcidin expression, which is a liver-produced hormone that is crucial for the maintenance of systemic iron homeostasis and that was recently found to oscillate in human blood (Schaap et al. 2013). Moreover, iron uptake *(Tfrc, Slc25a37/Mfrn)*, export *(Slc11a2/Dmt1, Slc39a14/Zip14)*, storage *(Ftl1, Fth1)* and transport *(Pcbp4),* as well as several biosynthetic processes that use iron *(Alas1, Alas2, Glrx5)* all showed pronounced rhythmicity.

Figure S9. Analysis of uORF usage in circadian core clock genes.

A) Normalised footprint read counts for uORF1 (left panel), uORF2 (right panel) and the main ORF (red) for the *Nr1d1* transcript. uORF and main ORF translation thus occur in sync.

B) Sequence alignment of the hypothetical 63 amino acid Nr1d1-uORF1 polypeptide across 6 mammalian species (mouse, rat, human, chimpanzee, macaque, horse). The uORF1 sequence is well conserved even at the amino acid level.

C) Schematic representation of the *Nr1d1* 5' UTR mutants that were tested for influence on main ORF translation. Nr1d1-Wt represents the annotated, full-length *Nr1d1* 5' UTR. In Del-uORF1, the first uORF was deleted as a whole. In Del-uORF1+2 a region encompassing both uORFs was removed. While these deletion mutants can be considered relatively invasive as they not only remove the uORF but also shorten the UTR as a whole, the following point mutants are likely more specific and suitable to measure uORF translation effects. M1A-uORF1 thus carries a point mutation at the predicted AUG start codon of uORF1 (converting it to an alanine codon). M1A-uORF2 is mutated to an alanine codon at the AUG of uORF2, whereas M1A-uORF1+2 is a combination of the two point mutations.

D) Raw (not detrended) data for the Nr1d1-Luc reporter traces shown in main Fig. 6C.

E) Summary of luciferase bioluminescence signal at the first peak of rhythmicity recordings as that in D) from several experiments (relative to Nr1d1-Wt which was set to 1 within each experiment). The signals of the mutants thus showed variability which conceivably originated from several sources: On the one hand, it may represent a true effect of impaired uORF translation, but quantitatively a likely stronger influence are technical parameters that are not controlled for in this assay. These are in particular the efficiency of lentiviral production and resulting lentiviral titres, the transduction efficiency of the NIH3T3 cells, the cellular growth rate and cell density in the transduced cells in the dishes subject to real-time recordings, etc. For these reasons, the UTRs shown in C) were recloned in a different assaying vector containing *Renilla* luciferase as an internal control gene, shown in main Fig. 6D.

F) shRNA-mediated *Denr* knockdown causes a short period phenotype of free-running circadian rhythms in NIH3T3 cells also using the *Bmal1-Luciferase* reporter (Nagoshi et al. 2004). Left panel: Summary of period change engendered by *Denr* shRNAs 1-3 relative to Scr (control) shRNA. Data are from 3 individual assayed dishes per condition from a single experimental series and show period length differences relative to the average of the Scr cells (error bars represent standard deviations). Within the experiment, the period shortening was significant for all three shRNAs at

the p=0.05 level (t-test). Right panel: Representative bioluminescence tracks of Scr (control) and *Denr* shRNA-transduced *Bmal1-Luc* cells.

Figure S10. Comparison of published rhythmic proteome datasets.

A) Euler diagram showing the overlap in detected rhythmic and non-rhythmic proteins in the two publications (Mauvoisin et al. 2014; Robles et al. 2014). Robles et al. analysed a total of 2883 detected proteins, and Mauvoisin et al. 4197 detected proteins. 2482 proteins are present in both datasets, corresponding to 86% (2482/2883) of the Robles datasets and 59% (2482/4197) of the Mauvoisin dataset. 179 proteins are rhythmic according to Robles et al., and 191 proteins according to Mauvoisin et al., but here the overlapping set only consists of 32 proteins (18% and 17% of Robles and Mauvoisin rhythmic proteins, respectively). Only in 13 (Robles) and 52 (Mauvoisin) cases the underlying reason was absence from the other dataset, indicating that the relatively poor overlap is not only the result of proteome coverage differences. A likely cause are differences in rhythmicity algorithms, cut-offs, FDRs, etc., that were applied.

B) Area-proportional Venn diagram of the subset of proteins that Robles et al. (left) and Mauvoisin et al. (right) classified as “mRNA flat – protein rhythmic”. Only 3 proteins are in common, namely Albumin (*Alb*), Apolipoprotein A-I (*Apoa1*) and Progesterone receptor membrane component 1 (*Pgrmc1*).

Figure S11. Known regulators of translation of IRE-containing transcripts do not show high-amplitude rhythmicity.

A) The *Aco1* gene encodes iron regulatory protein 1 (IRP1), one of the main actors of IRE-controlled translational regulation. Neither at the mRNA and footprint level (left panel) nor at the protein level (right panel) does ACO1 show high-amplitude rhythmicity. The fluctuations in ACO1 protein shown in this blot were not consistently found in a second protein series from independent animals and should therefore likely not be considered oscillating protein expression. The western blot shows one out of N=2 series that were analysed.

B) As in A), but for the *Ireb2* gene/IREB2 protein (also known as iron regulatory protein 2, IRP2). A slight rhythm – that did not pass the 1.5x amplitude cut-off applied in our study – can be seen at the level of mRNA and footprint abundance. At the protein level, fluctuations were not consistent between the two western blots that were performed on independent extract series, suggesting that they did not represent robustly regulated rhythmicity.

C) mRNA and footprint quantification for *Fbxl5*, a known regulator of IREB2 protein levels. No rhythm is detectable. Due to the lack in a functioning antibody (several commercial sources tested), confirmation by western blot as in A) and B) was not possible.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell culture

NIH3T3, NIH3T3-Dbp-Luc (Stratmann et al. 2012), NIH3T3-Bmal1-Luc (Nagoshi et al. 2004) and HEK 293FT cells were cultured under standard conditions (DMEM; 10% FCS, 1% penicillin/streptomycin; all from Invitrogen; 37°C; 5% CO_2). Lentiviral particle production (from plasmids prLV1-*Nr1d1* or pLKO.1-shRNA, using envelope and packaging plasmids pMD2.G, psPAX2) and viral transduction of NIH3T3 cells followed published protocols (Salmon and Trono 2007). Puromycin selection in shRNA experiments occurred at 5 µg/ml for 4 days. Dexamethasone synchronisation and recording of circadian bioluminescence rhythms has been described (Stratmann et al. 2012). Detrending of raw bioluminescence data was performed using a 24-hour moving average. For dual luciferase assays, cells were lysed in passive lysis buffer and luciferase activities were determined using the DualGlo Luciferase assay system and a GloMax 96 Microplate luminometer (all Promega). For protein quantification, cells were lysed in RIPA-buffer and analysed by SDS-PAGE and western blotting.

Ribosome profiling

Freshly harvested mouse livers were homogenised in 3 volumes of lysis buffer (150 mM NaCl, 20 mM Tris-HCl pH7.4, 5 mM $MgCl_2$, 5 mM DTT, 100 µg ml^{-1} cycloheximide, 1% Triton X-100, 0.5% Sodium deoxycholate, complete EDTA-free protease inhibitors (Roche) and 40 U ml^{-1} RNasin plus (Promega)) using a Teflon homogeniser. Lysates were incubated for 10 min on ice and cleared by centrifugation at 1000 x g, 4°C for 3 min in a tabletop centrifuge. Supernatants were flash-frozen and stored in liquid nitrogen. Lysates were thawed on ice and the OD_{260} was determined using a Nanodrop spectrophotometer. For each replicate and timepoint equal amounts of OD_{260} lysate from two mice were pooled. From the lysate pool, 15 OD_{260} were incubated with 650 U RNase I (Ambion) and 5 U Turbo DNase (Ambion) for 45 min at room temperature and gentle agitation. Nuclease digestion was stopped through addition of 8.7 µl Superasin (Ambion). Subsequently, lysates were applied to

sephacryl S-400 HR spin columns (GE Healthcare Life Sciences), pre-washed 3 times with 700 µl polysome buffer for 1 min at 600 x g, and centrifuged for 2 min at 600 x g and 4°C. The flow-through was immediately mixed with 1 ml Qiazol and ribosome-protected mRNA fragments were purified using miRNeasy RNA extraction kit (Qiagen) according to the manufacturer's instructions. For each sample 25 µg RNA were separated on a 15% urea-polyacrylamide gel and stained with SYBR-Gold (Invitrogen). Gel slices between 26-34 nucleotides were excised and RNA was extracted using 600 µl gel extraction buffer (0.5 M Ammonium acetate, 1 mM EDTA and 0.1% SDS) for 2 hours at room temperature and gentle agitation. Gel pieces were removed by centrifugation over spin filter tubes for 3 min at 5000 rpm. RNA was precipitated over night at -20°C in the presence of 900 µl isopropanol and 2 µl glycogen. RNA was pelleted and washed with 70% ethanol in a tabletop centrifuge at maximum speed and 4°C. Ribosomal RNA was depleted using the Ribo-Zero magnetic kit (Epicentre) according to the manufacturer's instructions. Sequencing libraries were generated using the ARTseq ribosome profiling kit (Epicentre) and sequenced on an Illumina HiSeq 2500.

RNA sequencing

For each timepoint and replicate the same samples as for RPF-seq (pool of liver lysates from two mice) was used to extract total RNA using miRNeasy RNA extraction kit (Qiagen). Ribosomal RNA was depleted using the Ribo-Zero magnetic kit (Epicentre) according to the manufacturer's instructions. Sequencing libraries were generated using the total RNA preparation protocol of the ARTseq ribosome profiling kit (Epicentre). Libraries were sequenced on an Illumina HiSeq 2500.

RT-qPCR

Reverse transcription was performed with random hexamer primers and the following gene-specific primers were used for real-time PCR. *Denr* forward/reverse: AAAGGCGATACGAAGAACAGTG, CATCCGGCATGTATTCACAGT*; Nudt4* (control gene) forward/reverse: AAGTTCAAGCCCAACCAGACG, TCCTGGGACAATCCATTGGTC*; Firefly luciferase 1* forward/reverse: CCGCCTGAAGTCTCTGATTAAGT, ACACCTGCGTCGAAGATGTTG*; Firefly luciferase 2* forward/reverse: TGCAAAAGATCCTCAACGTG, AATGGGAAGTCACGAAGGTG*; Renilla luciferase* forward/reverse: GGAATTATAATGCTTATCTACGTGC, CTTGCGAAAAATGAAGACCTTTTAC.

Preparation of protein extracts

For the preparation of nuclear extracts, freshly harvested livers were homogenised each in 4 ml nuclear homogenisation buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.3 M sucrose, 0.1 mM EDTA, 0.74 mM spermidine, 1 mM DTT, supplemented with complete protease inhibitor tablets from Roche) for 1 min using a Teflon homogeniser and subsequently mixed with 8 ml cushion buffer (10 mM HEPES pH 7.9, 2.2 M sucrose, 0.1 mM EDTA, 0.74 mM spermidine, 1 mM DTT, complete protease inhibitors). The homogenate was split in two and each part was layered on top of 3 ml cushion buffer and centrifuged at 4°C for 1 h at 25000 rpm in a SW40 rotor. The nuclear pellets were re-suspended in 400 µl RIPA-buffer (150 mM NaCl, 50 mM Tris pH 8.0, 1% NP40, 0.5% DOC, 0.1% SDS and complete protease inhibitors) and incubated for 20 min on ice. The lysates were cleared by centrifugation at 16000 x g for 5 min at 4°C and stored at -80°C.

For the preparation of total cell extracts, livers were lysed in 3 volumes (w/v) RIPA-buffer using a Teflon homogeniser, incubated for 30 min on ice and centrifuged for 10 min at 16000 x g and 4°C. Supernatants were stored at -80°C. Protein concentrations were quantified using a BCA protein assay kit (Pierce).

For the isolation of mitochondria, livers were homogenised in 5 volumes (w/v) Mt-homogenisation buffer (10 mM Tris-MOPS pH 7.4, 1 mM EGTA-Tris pH 7.4, 200 mM sucrose) with 5 strokes in a Teflon homogeniser. The homogenate was centrifuged at 600 x g for 10 min at 4°C. The supernatant was transferred to a new tube and centrifuged at 7000 x g for 10 min at 4°C. Pellets were washed in 1 ml Mt-homogenisation buffer, re-centrifuged at 7000 x g for 10 min and re-suspended in SDS-PAGE sample buffer.

Protein extracts (25 µg) were analysed by SDS-PAGE using standard conditions.

Western blot analysis

Immunoblotting to PVDF or nitrocellulose membranes was performed using a Tris/CAPS discontinuous buffer system with semi-dry transfer (Trans-Blot semi-dry electrophoretic transfer cell, Bio-Rad) according to the manufacturer’s instructions. Membranes were blocked and incubated with primary and secondary antibodies in 5% milk/TBS-T or 5% BSA/TBS-T. Primary and secondary antibodies were used at the following dilutions: anti-FTL1 (1:100, Santa Cruz Biotechnology, sc-14420), anti-FTH1 (1:1000, Cell Signaling, #4393), anti-DENR (1:5000, Abcam, ab108221) anti-ALAS2 (1:1000, Abcam, ab184964), anti-NDUFB8 or anti-mt-CO1 (OxPhos rodent antibody cocktail, 1:250, Abcam, ab110413), anti-GGPS1 (1:1000, Abcam, ab167168), anti-DEAF1 (1:1000, LSBio, LS-C80262), anti-U2AF2 (also known as U2AF65) (1:5000, Sigma, U4758), anti-beta-Tubulin (1:5000, Sigma, T5201), anti-ACO1 (1:2000, Abcam,

ab126595), anti-IREB2 (1:2000, Abcam, ab181153), anti-rabbit-HRP or anti-mouse HRP (1:10000, Promega) and anti-goat-HRP (1:5000, Santa Cruz Biotechnology). Signals were visualised using LumiGLO or LumiGLO-Reserve chemiluminescent substrate kits (KPL). For each timepoint 2-3 animals were analysed; shown are the results from one representative time series.

Cloning, plasmids

A 8.2 kb fragment containing 3.5 kb of the *Nr1d1* promoter region, exon 1, intron 1, the first 14 bp of exon 2 and the luciferase coding region was amplified by PCR from the adenoviral plasmid pCV100-*Nr1d1*-Luc described previously (Saini et al. 2013) using the following primers: forward: ACGTTCGCGACGCGTGTGTGTATGTGTGTG, reverse: TGCACTCGAGTTACACGGCGATCTTTCCGCC. Fragments were cloned between *Xho*I and *Nru*I sites to replace the dual luciferase reporter cassette in the prLV1 lentiviral vector previously described (Du et al. 2014) to generate prLV1-*Nr1d1*-Wt. Subsequently, a 1.1 kb fragment of the *Nr1d1* promoter region was deleted by site-directed mutagenesis using forward: GAAGCCAGAAGCTGGGAGCACACACGCGTCGCGATT, reverse: AATCGCGACGCGTGTGTGCTCCCAGCTTCTGGCTTC to reduce plasmid size. The size-reduced plasmid was then used to generate *Nr1d1*-mutants by means of site-directed mutagenesis using the following primer combinations: M1A-uORF1, mutation of the uORF1 start codon to alanine, (F: CTCTCTGCTCTTCCCGCGCAAATCAGATCTCAGG, R: CCTGAGATCTGATTTGCGCGGGAAGAGCAGAGAG); M1A-uORF2, mutation of the uORF2 start codon to alanine, (F: CAAGGTCCAGTTTGAGCGACCGCTTTCAGCTGG, R: CCAGCTGAAAGCGGTCGCTCAAACTGGACCTTG); Del-uORF1, deletion of uORF1 (F: ACACTCTCTGCTCTTCCCATGACCGCTTTCAGCTGG, R: CCAGCTGAAAGCGGTCATGGGAAGAGCAGAGAGTGT); Del-uORF1+2, deletion of uORF1 and the first 4 codons of uORF2, (F: GGAGCTCCAGATTCATTACCAGCTGGTGAAGACATGACG, R: CGTCATGTCTTCACCAGCTGGTAATGAATCTGGAGCTCC).

For the generation of dual luciferase (Firefly/*Renilla*) reporter plasmids, fragments containing the 5’ UTR and the first 10 amino acids of the *Nr1d1* coding sequence were amplified by PCR from prLV1-*Nr1d1*-Wt and mutant plasmids using the forward: AAAAGGATCCACTAGTGGGAAAGGCTCGGGCAAAAGGCGG and reverse primer: TTTGGATCCTGTGTTGTTATTGGAGTCCAGGGT and cloned into the *BamH*I site of the prLV1 dual luciferase reporter plasmid (Du et al. 2014).

For shRNA vectors, the following hairpins targeting *Denr* were cloned: shRNA1: GTACCACAGAAGGTCACGATA, corresponding to clone TRCN0000308443; shRNA2: GTGCCAAGTTAGATGCGGATT, corresponding to clone TRCN0000098826; shRNA3:

GTACCACAGAAGGTCACGATA, corresponding to clone TRCN0000098827); scramble shRNA (Addgene #1864) served as control.

Initial processing and quality assessment of sequencing data

Initial quality assessment of the sequencing reads was conducted based on the preliminary quality values produced by the Illumina pipeline Casava 1.82. Quality related statistics used were the percentage of clusters passed filtering (%PF clusters) and the mean quality score (PF clusters), as other quality related statistics, such as percentage of reads aligned, mean alignment score and percentage of alignment error were not very useful due to large contributions of adapter sequences. Adapter sequences were removed using cutadapt utility (Martin 2011) with following options: -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC –match-read-wildcards -m 6. The size distribution of insertions was employed to assess the quality of sequencing libraries, especially for ribosome footprints; for samples that did not produce a distinct peak around 29-30 nucleotides, either the sequencing or the library preparation was repeated. Next, trimmed read sequences were filtered by their size using an in-house Python script with following inclusive ranges: [26,35] for footprints, [21,60] for total RNA. Smaller or larger fragments were kept separately and not used in further analyses.

Alignment to the mouse genome

Trimmed and filtered insert sequences were mapped sequentially to following databases: mouse rRNA, human rRNA, mt-tRNA, mouse tRNA, mouse cDNA from Ensembl mouse database release 75 (Flicek et al. 2013) and, finally, mouse genomic sequences (Genome Reference Consortium GRCm38.p2). All but last one were mapped using bowtie version 2.2.1 (Langmead and Salzberg 2012) using the following parameters: -p 2 -L 15 -k 20 –no-unal. Mapping against genomic sequence was performed using tophat v2.0.11 (Trapnell et al. 2009) with the following parameters: –transcriptome-index=Mmusculus.GRCm38.75.dna.ensembl_data -p 2. After each alignment, only reads that were not aligned were used in the following mapping. For further analysis, only alignments against mouse cDNA were used, unless specifically stated otherwise. For each query sequence, only alignments with maximum alignment score (AS) were kept.

Separately from this sequential alignment strategy, trimmed and filtered insert sequences from each sample were also directly aligned against the mouse genome using tophat with similar parameters. The output of this alignment was used to

estimate expressed transcript models out of all models contained in Ensembl mouse database release 75. To this end, we used cufflinks v2.2.1 (Trapnell et al. 2010) to estimate the number of fragments per kilo base of exon per million fragments mapped (FPKM) for each transcript, with the following parameters: –GTF Mus_musculus.GRCm38.75.gtf –frag-len-mean 37 –frag-len-std-dev 8 –compatible-hits-norm –multi-read-correct –upper-quartile-norm –frag-bias-correct Mmusculus.GRCm38.75.dna.ensembl.fa -p 3. Transcript FPKM estimates from all total RNA samples were merged using cuffcompare with following parameters: -r Mus_musculus.GRCm38.75.gtf -R -V. Resulting FPKM tracking information was parsed with an in-house Python script to filter out transcripts which were not found to have an FPKM > 0.1, a LOW95 > 0.05 and a FMI > 2.0 in at least 3 samples. A database of expressed transcripts based on this filtering was used in further analysis.

Quantification of mRNA and ribosome footprint abundance

Abundance of mRNA and RPF (ribosome protected fragment) was estimated per gene. For this quantification, only reads that were uniquely mapped to a single gene and only transcripts that were identified as expressed (see Alignment to the Mouse Genome) were used. An in-house Python script was used to count reads within each annotation feature (5’ UTR, CDS, 3’ UTR) per gene. For genes that were associated with multiple expressed transcripts (multi-isoform genes), reads that were not mapped unambiguously within a single feature were assigned to one with the following preference order: CDS, 5’ UTR, 3’ UTR.

Prior to further analysis, variation captured by the CDS counts was inspected with the principal component analysis (PCA) tools provided in the R package DESeq v1.14.0 (Anders and Huber 2010). The aim of this type of analysis is to visualise the overall effect of experimental covariates and to detect if batch effects or outlying samples are present in the experiment. For this, we followed the steps described in the reference manuals of DESeq. Briefly, a combined matrix of CDS counts for RPF and mRNA was built, where each row contained time-series data with biological replicates per gene. Normalisation factors for library sizes were calculated using upper quartile method and dispersions were blindly estimated using the "pooled" method (ignoring conditions) of DESeq. Prior to PCA, the normalised (division by size factors) count data were transformed using the getVarianceStabilizedData function from DESeq package, which calculates a variance stabilising transformation (VST) from the fitted dispersion-mean relation, yielding a matrix of values which are approximately homoskedastic. The variance-stabilised data matrix was then used for PCA using the plotPCA function of DESeq. Briefly, rows were ranked in decreasing order by their

row-wise variances and the top 4000 genes were then used in PCA. The number of genes included was identified empirically by increasing the initial default number (500) iteratively until the PCA results reached a stable state. The plotPCA function, by default, performs a PCA without centring or further scaling.

Mappable and countable feature lengths (in nucleotides) for each locus were calculated by means of generating all possible 21-, 30-, 45-, and 60-nt long reads *in silico* (faux reads) for each expressed transcript in the database and counting the faux reads through identical mapping and counting work flow used for real experimental reads (Du et al. 2014). Weighted averages of these mappable and countable lengths were later used in RPKM calculations. For weights, proportions of 21-, 30-, 45-, and 60-nt inserts within averaged shifted histograms of insert length distributions were used.

Read counts of total RNA and RPF were normalised with upper quantile method of R package edgeR v3.4.2 (Robinson et al. 2010). Prior to normalisation, transcripts, which did not have at least 10 counts in at least one fourth of the samples, were removed from the datasets. For better comparability between datasets, RPKM values were calculated as the number of counted reads per 1000 mappable and countable bases per geometric mean of normalised read counts per million. The geometric mean of normalised read counts were 11,025,834 and 18,520,239 for total RNA and RPF datasets, respectively. Translational efficiencies (TE) were then calculated as the ratio of RPF-RPKM / mRNA-RPKM for each gene per sample. For most analysis downstream, TEs were log2 transformed and means or replicates per timepoint or grand-means over all timepoints were used as necessary.

Only for comparison of the translational efficiencies obtained from this study to those reported in (Ingolia et al. 2011), we first corrected the differences in location by adjusting the means trimmed 10%. Then, 13 data points at the extremities were removed from Ingolia’s dataset to achieve similar ranges in both datasets.

Analysis of mRNA abundances and translation efficiencies

Changes in the mRNA abundance or TE distributions between different sets of genes were compared by estimating the amount of location shift and its significance using Wilcoxon rank sum test (equivalent to the Mann-Whitney test). For all cases, a two-sided test was performed and if necessary resulting p-values were corrected for multiple testing via BH method (Benjamini and Hochberg 1995). The median of all possible differences between a sample from the first set and a sample from the second set was then reported as an estimate for the difference in location parameters (location shift).

For all other statistical calculations and figure production, the R environment (R Core Team, 2013) with various packages (ggplot2 (Wickham 2009); MASS (Venables 2002)) was used. For production of plots of RNA and RPF densities along transcripts, an in-house Python script was used.

Analysis of significant changes in ribosome occupancy

Significance of changes in translational regulation within and between timepoints were assessed with the *Babel* computational framework (Olshen et al. 2013). To this end, unnormalised count data for both total RNA and RPF were used following the authors' instructions. 10,000,000 permutations were performed within comparisons to achieve precise estimates. The false discovery rate-adjusted p-values were then used to create lists of genes, whose ribosome occupancies were significantly altered (fdr-adjusted p-value < 0.05) within a timepoint or between any two timepoints. A union of these two gene lists was also created to include all genes that are predicted to have alterations in their ribosome occupancies (e.g. translational efficiencies).

Rhythmicity detection in mRNA and RPF profiles

Rhythmic parameter estimation methods that depend on harmonic regression typically fail to produce accurate estimates from abundance profiles of genes that show pronounced deviations from sinusoidal curves. An example is *Per1* (Fig. 3B), where the increasing and decreasing portions of the profile are not symmetrical. To overcome this shortcoming, we have devised an algorithm to detect rhythmicity and estimate rhythmic parameters, which is based on Akaike information criterion (AIC) based model selection (Wagenmakers and Farrell 2004) between three linear models: mean model with 1 parameter, harmonic model with 2 parameters (Symul 2013), and a sigmoid model with 5 parameters (Symul 2013). The sigmoid fit is composed of two sigmoid curves, synthesis-up, SU, which lasts l hours, and synthesis-down, SD, which lasts m hours. The two curves, together with the phase p describe a cyclic curve that runs from the basal level B to max level, B times the fold-change fc. The equation for the sigmoid function $f(B,fc,p,m,l)$ of time t is

$B(1+(fc-1)(SU+SD))),$

where,

For each gene, normalised CDS counts of total RNA and RPF reads were then fit to each model using the method of least squares. For all regressions, the two biological replicates per timepoint were treated as independent replicates and were not combined into a false 48-hour data-series. For sigmoid curve fitting, the minpack.lm

package in R environment was used. Minimum and maximum values allowed for *m* and *l* were 3 and 16 hours. First, for each model, we computed the differences in AICc (AIC corrected for finite sample size) with respect to the AICc of the best candidate model, ΔAICc. Then, ΔAICc were used to calculate the Akaike weight for each model. A dataset was considered to be rhythmic if the evidence for mean-model (constant expression) was less than 0.05. For such genes, the sigmoid model was chosen over the sinusoidal (harmonic regression) only when the evidence for sigmoid model was at least 3.5 times higher than that for sinusoidal to prevent over-fitting. The rhythmic parameters, phase and amplitude-ratio, were then estimated from the selected model. Finally, genes that did not have an amplitude-ratio of at least 1.5 were flagged and not categorised as cyclic.

To enable a comparison of our results to those from previously published circadian proteomics studies (Mauvoisin et al. 2014; Robles et al. 2014), the UniProt and gene name identifiers in these reports were converted into EnsemblIDs using the external references maintained in the Ensembl database.

The Rayleigh test for uniformity and the Watson-Wheeler test for homogeneity were performed using the R package 'circular'.

Analysis of RPF positions on transcripts

For all total RNA and RPF reads that were counted towards genes, we have also tracked the position of the 5’ end of the read relative to the 5’ end of its corresponding transcript and the trimmed size of the read. The putative A-site of ribosomes was then calculated as 5’ position of a RPF read plus 15 if the read was ≤30 nt long or 16 otherwise. For analysis of the 3 nucleotide rhythmicity of RPFs, only a subset of genes that i) had a single transcript isoform with a CDS length of at least 400 nt, ii) were expressed at both mRNA and footprint levels, iii) had a RPF-RPKM > 5, iv) had at least 20 nt in UTRs preceding and following CDS region, and v) had at least 10 reads within both the first and the last 200 nt, were considered (N=3237). For calculation of frame preference in CDS, the last two criteria were not used (N=3793). Upstream ORFs (uORFs) were identified within the 5' UTRs of a restricted set of genes, for which a single protein-coding isoform was found to be expressed (N=6774), with the following criteria: i) start with AUG, ii) are at least 18 nucleotides long, iii) can overlap with main CDS. To identify uORFs that are actively translated we used a similar strategy described elsewhere (Bazzini et al. 2014). Briefly, A-site counts for each reading frame within a predicted uORF were tested for significant deviations from a uniform distribution with a chi-square test. Resulting p-values were adjusted for FDR using BH (Benjamini and Hochberg 1995), and a uORF was

considered as actively translated only then if it had an adjusted p-value <0.05, the preferred reading frame was the first one (relative to uORF's 5'), and it had a RPF coverage >10%. Same analysis was applied to total RNA reads, to validate its specificity and also filter out unspecific cases (likely false positives) which suggested translation events from total RNA reads (N=34).

The effects of feature (5' UTR, CDS, 3' UTR) size and presence of translated uORFs on the translation efficiency (TE) from the main CDS were analysed using linear least squares regression. While log-transformation of TE is widely used elsewhere in the manuscript and in almost all other publications reporting ribosome footprinting, for this analysis we found that cube root transformation was more optimum using the powerTransform tools from the R package 'car' v2.0.25 (Fox and Weisberg 2010). Feature sizes were log10 transformed, and the presence or absence of translated uORFs in a transcript was modelled as a binary predictor. The effects of the predictors were then visualised via partial regression plots using the avPlots function from package 'car'.

Identification of pause-sites

Analysis of ribosome pausing was carried out following previous report (Ingolia et al. 2011). The same restricted set of genes (N=6774) that was used for uORF analysis was used for pausing analysis. Mainly, two modifications to the original method were revised. Firstly, as an estimate of central tendency we devised to employ the trimean, which is the average of the median and the midhinge, instead of the median. The trimean is a more resistant measure and copes better where median can have a value of zero. Secondly, we have observed a small but significant positive correlation (Pearson r = 0.18, p-value < 2.2e-16) in normalised codon counts between RPF and mRNA, suggesting local RPF read densities along transcripts could also be influenced by other factors than ribosome pausing. To overcome this effect we devised a method where RPF codon values were normalised by mRNA codon values before identification of pause-sites. Specifically, we first calculated the pileup profile of mRNA reads along each transcript. At positions (nt), where pileup values were smaller than the first quartile of all pileup values, the values were set equal to the first quartile value. The resulting profile was smoothened with a correlation filter and subsequently used to calculate the trimean normalised codon counts for mRNA reads. Finally, at each codon the RPF count was divided by the mRNA count. This approach provided a conservative means of normalisation where over-representation of pause-sites that coincide with prominent peaks in the mRNA pileup profile were was lessened, without extensively promoting non-pause-sites that coincide with troughs in the mRNA pileup

into novel pause-sites. The correlation in codon counts between RPF and mRNA dropped drastically after this normalisation procedure ($r = 0.08$, p-value $< 2.2e-16$). A site is labelled as pause-site only if both biological replicates had a score > 25 at that site.

SUPPLEMENTAL REFERENCES

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome biology* **11**: R106.

Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC et al. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO journal* **33**: 981-993.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**: 289-300.

Du NH, Arpat AB, De Matos M, Gatfield D. 2014. MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. *eLife* **3**: e02510.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S et al. 2013. Ensembl 2013. *Nucleic acids research* **41**: D48-55.

Fox J, Weisberg S. 2011. *An R Companion to Applied Regression*. SAGE Publications.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**: 789-802.

Kim TY, Wang D, Kim AK, Lau E, Lin AJ, Liem DA, Zhang J, Zong NC, Lam MP, Ping P. 2012. Metabolic labeling reveals proteome dynamics of mouse mitochondria. *Molecular & cellular proteomics : MCP* **11**: 1586-1594.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357-359.

Le Martelot G, Canella D, Symul L, Migliavacca E, Gilardi F, Liechti R, Martin O, Harshman K, Delorenzi M, Desvergne B et al. 2012. Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles. *PLoS biology* **10**: e1001442.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**: 10-12.

Mauvoisin D, Wang J, Jouffe C, Martin E, Atger F, Waridel P, Quadroni M, Gachon F, Naef F. 2014. Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 167-172.

Nagoshi E, Saini C, Bauer C, Laroche T, Naef F, Schibler U. 2004. Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell* **119**: 693-705.

Olshen AB, Hsieh AC, Stumpf CR, Olshen RA, Ruggero D, Taylor BS. 2013. Assessing gene-level translational control from ribosome profiling. *Bioinformatics* **29**: 2995-3002.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Robles MS, Cox J, Mann M. 2014. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS genetics* **10**: e1004047.

Saini C, Liani A, Curie T, Gos P, Kreppel F, Emmenegger Y, Bonacina L, Wolf JP, Poget YA, Franken P et al. 2013. Real-time recording of circadian liver gene expression in freely moving mice reveals the phase-setting behavior of hepatocyte clocks. *Genes & development* **27**: 1526-1536.

Salmon P, Trono D. 2007. Production and titration of lentiviral vectors. *Current protocols in human genetics / editorial board, Jonathan L Haines [et al]* **Chapter 12**: Unit 12 10.

Schaap CC, Hendriks JC, Kortman GA, Klaver SM, Kroot JJ, Laarakkers CM, Wiegerinck ET, Tjalsma H, Janssen MC, Swinkels DW. 2013. Diurnal rhythm rather than dietary iron mediates daily hepcidin variations. *Clinical chemistry* **59**: 527-535.

Stratmann M, Suter DM, Molina N, Naef F, Schibler U. 2012. Circadian Dbp transcription relies on highly dynamic BMAL1-CLOCK interaction with E boxes and requires the proteasome. *Molecular cell* **48**: 277-287.

Symul L. 2013. Kinetic Analysis of Transcriptional and Post-Transcriptional Processes During Circadian Cycles. in *Life Sciences*. EPFL, Lausanne.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**: 511-515.

Venables WN. 2002. *Modern applied statistics with S / W.N. Venables, B.D. Ripley*. Springer, New York.

Wagenmakers EJ, Farrell S. 2004. AIC model selection using Akaike weights. *Psychonomic bulletin & review* **11**: 192-196.

Wickham H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer Publishing Company, Incorporated.

Analyzing the temporal regulation of translation efficiency in mouse liver

Peggy Janich, Alaaddin Bulak Arpat, Violeta Castelo-Szekely and David Gatfield

Genomics Data (2016) 8:41-4

Genomics Data 8 (2016) 41–44

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in brief

Analyzing the temporal regulation of translation efficiency in mouse liver

Peggy Janich [a], Alaaddin Bulak Arpat [a,b], Violeta Castelo-Szekely [a], David Gatfield [a,*]

[a] Center for Integrative Genomics, Génopode, University of Lausanne, 1015 Lausanne, Switzerland
[b] Vital-IT, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 10 March 2016
Accepted 16 March 2016
Available online 18 March 2016

ABSTRACT

Mammalian physiology and behavior follow daily rhythms that are orchestrated by endogenous timekeepers known as circadian clocks. Rhythms in transcription are considered the main mechanism to engender rhythmic gene expression, but important roles for posttranscriptional mechanisms have recently emerged as well (reviewed in Lim and Allada (2013) [1]). We have recently reported on the use of ribosome profiling (RPF-seq), a method based on the high-throughput sequencing of ribosome protected mRNA fragments, to explore the temporal regulation of translation efficiency (Janich et al., 2015 [2]). Through the comparison of around-the-clock RPF-seq and matching RNA-seq data we were able to identify 150 genes, involved in ribosome biogenesis, iron metabolism and other pathways, whose rhythmicity is generated entirely at the level of protein synthesis. The temporal transcriptome and translatome data sets from this study have been deposited in NCBI's Gene Expression Omnibus under the accession number GSE67305. Here we provide additional information on the experimental setup and on important optimization steps pertaining to the ribosome profiling technique in mouse liver and to data analysis.

Specifications	
Organism/cell line/tissue	Mus musculus/liver
Sex	Male
Sequencer or array type	Illumina HiSeq 2500
Data format	Raw and processed data
Experimental factors	Livers were collected every 2 h during the 24-h daily cycle (with 2 replicate time series)
Experimental features	RNA-seq and RPF-seq were performed in parallel on the same liver lysates to identify mRNA subject to rhythmicity at the translational level
Consent	Data are publicly available at NCBI GEO
Sample source location	Lausanne, Switzerland

1. Direct link to deposited data

Direct link to deposited files: http://datalink.elsevier.com/midas/datalink/api/downloadfiles?items=18934-18935-18936

Direct link to deposited genomic data: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=etmbssamttcnzsb&acc=GSE67305

* Corresponding author.
E-mail address: david.gatfield@unil.ch (D. Gatfield).

2. Experimental design, materials and methods

2.1. Experimental design

To investigate daily rhythms in translation, we recently performed ribosome profiling in mouse liver (Janich et al., 2015 [2]), which is the most commonly used peripheral organ in circadian research due to its easy dissectability, its relatively homogenous cellular composition and its abundant, high-amplitude rhythms [3]. Ribosome profiling is based on the deep sequencing of ≈30 nucleotide mRNA fragments that are protected by translating ribosomes upon nuclease digestion [4]. The sequence information contained in the footprints allowed us to perform transcriptome-wide, quantitative analyses of protein synthesis rhythms in mouse liver. Parallel RNA-seq data was used to quantify RNA abundance around-the-clock, allowing the identification of those genes whose rhythmicity was exclusively translational. Livers were collected at 2 h intervals around-the-clock in order to have sufficient temporal resolution for reliable rhythmicity detection. For each time point, two replicate samples were generated. Each replicate consisted of a pool of 2 individual livers.

2.2. Mice

Wild type C57BL/6J mice were purchased from Janvier Labs. Animal housing and experimental procedures were in agreement with the veterinary law of the Canton Vaud, Switzerland (authorization to DG: VD2376). For all experiments, mice were entrained to 12-h-light/12-

http://dx.doi.org/10.1016/j.gdata.2016.03.004

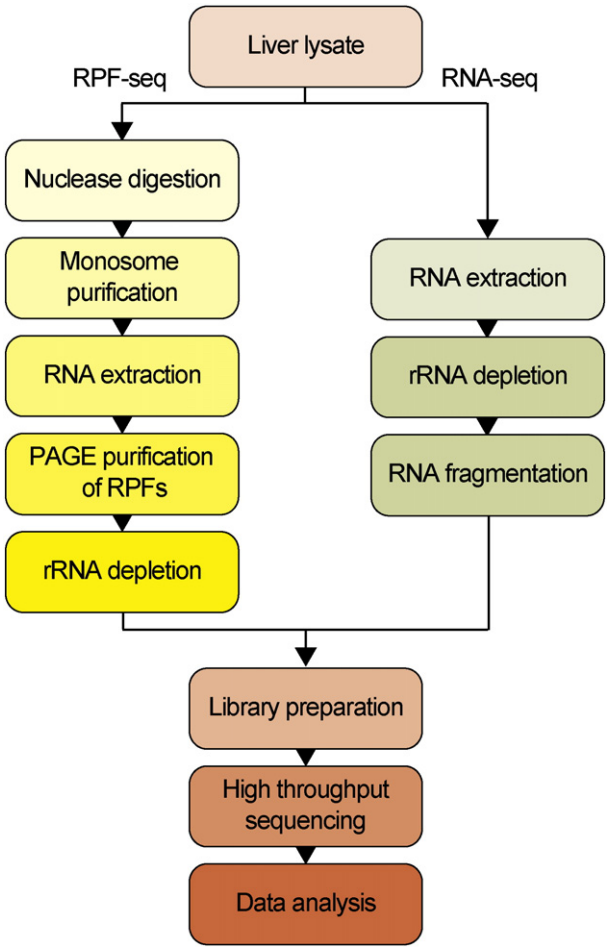


Fig. 1. Overview of the experimental workflow used for ribosome profiling (RPF-seq) and for RNA-seq in mouse liver.

h-dark cycles for 2 weeks with water and normal chow available ad libitum. Prior to organ collection, mice were anesthetized with isoflurane and sacrificed by decapitation. Mice were sacrificed at the indicated Zeitgeber times (ZT), with ZT00 corresponding to "lights on" and ZT12 to "lights off" in the animal housing facility. Livers were rapidly excised and immediately processed to lysate.

2.3. Lysate preparation

Freshly extracted liver tissue from each individual mouse was weighed and subsequently lysed with 8 strokes in a Teflon homogenizer containing 3 volumes of ice-cold lysis buffer (20 mM Tris–HCl pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT, 100 μg/ml cycloheximide, 1% Triton X-100, and 0.5% sodium deoxycholate) supplemented with complete EDTA-free protease inhibitors (Roche) and 40 U/ml RNasin plus (Promega). Of note, RNasin plus inhibits RNase A, B and other RNases present in liver extracts, but not RNase I, which will be used at a later stage of the protocol to generate ribosome protected mRNA fragments (RPFs). The liver homogenates were transferred to microcentrifuge tubes and incubated for 10 min on ice. Cellular debris was pelleted by centrifugation at 1000 ×*g* for 3 min at 4 °C. The supernatant was removed, aliquoted, and snap-frozen and stored under liquid nitrogen until further processing. For absorbance measurements at 260 nm, lysates were gently thawed on ice, diluted 1:10 and 1:20 in water, the absorbance determined by Nanodrop and the average value from the two dilutions was determined. In general, the lysates ranged between 100 and 200 OD260 per 1 ml lysate. Equal amounts of lysate (OD260) from 2 mice, collected at the same time point, were pooled and diluted with lysis buffer to a final concentration of 15 OD260/100 μl. Lysates were processed separately for RPF-seq and RNA-seq (Fig. 1).

2.4. RNA extraction and RNA-seq library preparation

For the isolation of total (cytoplasmic) RNA, 100 μl pooled lysate was mixed with 1 ml Trizol and incubated for 5 min at room temperature. RNA was isolated using the miRNeasy kit (Qiagen) according to the manufacturer's protocol and the concentration determined by Nanodrop. Prior to library preparation, a total of 5 μg RNA was subjected to ribosomal RNA depletion (Ribo-Zero magnetic kit, Epicenter) and subsequently purified using a RNA purification kit (RNA Clean & Concentrator-5, Zymo Research). RNA-seq libraries were generated following the instructions for total RNA library preparation of the ARTseq ribosome profiling kit (Epicenter).

2.5. Preparation of ribosome protected mRNA fragments (RPFs)

In preparation to processing the samples from the large-scale time series, we optimized the conditions for nuclease digest in order to ensure efficient and reproducible generation of RPFs of ≈30 nucleotides in length. To this end, liver lysates were incubated with different amounts of RNase I ranging from 0 to 1000 units (Ambion). The digested mRNA fragments were purified using Trizol extraction and analyzed by northern blot as described previously [5]. Northern blot hybridization was performed using 2 different probes recognizing two highly expressed liver mRNAs (Alb, albumin and Mup, major urinary protein). Analysis of the autoradiographs showed that the optimal concentration for obtaining mRNA fragments of 30 nucleotides in lysates prepared from mouse liver was in the range of 600 to 1000 units RNase I (Fig. 2).

Thus, for our time course experiment, lysates of a concentration of 15 OD260 in a volume of 100 μl were incubated with 650 units RNase I and 2.5 μl DNase I for 45 min at room temperature. After the incubation, samples were placed on ice and 8.7 μl Superasin RNase inhibitor (Ambion) were added to inactivate the RNase I enzyme. In the meantime, size exclusion spin columns (S-400, GE Healthcare Life Sciences), that would subsequently serve to purify the nuclease-generated monosomes, were prepared. To this end, spin columns were washed 3 times with 700 μl lysate buffer containing 20 U/ml Superasin in a microcentrifuge at 600 ×g for 1 min. Between each washing step the matrix of the spin column was gently resuspended by vortexing. After the washing steps, lysates were applied to the matrix and the spin columns were centrifuged for 2 min at 600 ×*g*. 1 ml of Trizol was added

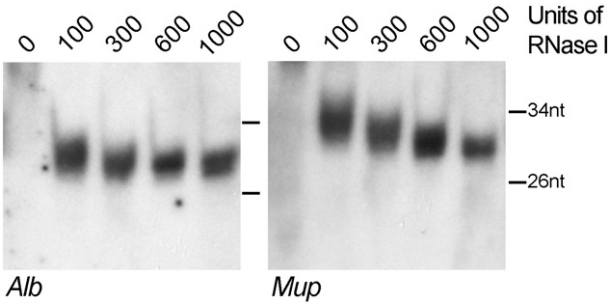


Fig. 2. Optimization of RNase I concentration. Autoradiographs of RNase I-digested liver RPFs probed for two highly expressed liver mRNAs, albumin (Alb, probe: cgatgggcgatctcactcttgtgtgcttctc) and major urinary protein (Mup, probe: gttccttcccgtagaactagcttc). RPFs of 30 nucleotides in length were obtained when 600–1000 units of RNase I were used for digestion.

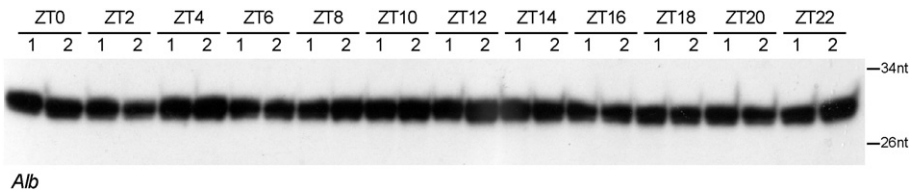


Fig. 3. Quality control of footprints generated from time series. Autoradiograph of liver RPFs from all replicate samples of the time series experiment probed for albumin (Alb, probe: same as in Fig. 2). In all samples RPFs of 30 nucleotides in length were detected, indicating that the nuclease digestion conditions were homogenous and reproducible. ZT, Zeitgeber time.

to the flow-through and RNA was extracted using the miRNeasy kit (Qiagen) according to the instructions of the manufacturer. The concentration of the RNA was determined by Nanodrop and RNA was stored at −80 °C. Before proceeding to library preparation, the quality of the extracted RPFs was verified for all samples of the time course experiment by northern blot analysis (Fig. 3).

2.6. RPF-seq library preparation and sequencing

Library preparation of RPFs was performed according to the ARTseq ribosome profiling kit (Epicenter) with the modification that the steps of ribosomal RNA depletion (Ribo-zero) and polyacrylamide gel electrophoresis (PAGE) purification were inverted. The original protocol uses 5 μg of RPF RNA as starting material, which is the maximum amount of RNA recommended in a Ribo-zero reaction. However, using this protocol we were occasionally unable to obtain sufficiently concentrated libraries for sequencing. The inversion of the 2 steps allowed us to increase the starting material up to 25 μg RPF RNA, without noticing a negative effect on the performance of the Ribo-zero depletion. In our case, the RNA was first separated on a 15% urea-polyacrylamide gel and stained with SYBR-Gold (Invitrogen). Gel pieces between 26 and 34 nucleotides were excised, the RNA extracted, purified and only then used for ribosomal RNA depletion. After the Ribo-zero reaction, the RNA was purified using a RNA purification kit (RNA Clean & Concentrator-5, Zymo Research) and immediately subjected to end-repair and 3′ adaptor ligation according to the instructions of the ARTseq protocol. Reverse transcription, cDNA purification and circularization were done as described in the ARTseq protocol. During the last step of library preparation, i.e. the PCR amplification, 5 μl out of the 20 μl circularized cDNA product was used as template in the PCR reaction. The quality of the final library was assessed by analyzing 5 μl of the PCR reaction on an 8% native PAGE, stained with SYBR Gold. For some libraries, only a faint band was visible, and we then performed up to 3 additional PCR amplifications as described above in order to have sufficient material. In a final step, the identical PCR reactions were pooled and loaded on an 8% native PAGE. The library running at ≈150 bp was gel-purified, precipitated with ammonium acetate and resuspended in water. The concentration and the quality of both RPF-seq and RNA-seq libraries were determined by Qubit (Thermo Fisher) and BioAnalyzer (Agilent). Up to 6 libraries were multiplexed and subjected to 100 cycles of single-end sequencing on an Illumina HiSeq2500 (Illumina).

2.7. Data processing and analysis

Here, we briefly outline the first steps of data analysis. More detailed information on data processing and analysis can be found in the supplementary information of the original publication [2].

In a first step, sequencing reads were de-multiplexed, adapter-trimmed and size-filtered (lengths of 26–35 nucleotides for RPF-seq and 21–60 nucleotides for RNA-seq). Next, we removed from the filtered reads sequentially the mouse and human ribosomal rRNA, mouse mt-tRNA, mouse tRNA, and then finally aligned the reads to mouse cDNA (Ensembl release 75) and genomic sequences (GRCm38.p2), using bowtie version 2.2.1 and Tophat v2.0.11, respectively. In parallel, a database of expressed transcripts was generated from RNA-seq reads (cufflinks v2.2.1) and used for subsequent analyses. Read quantification was done per gene for each annotation feature (5′ UTR, CDS, 3′ UTR). Read counts were normalized with the upper quantile method of R package edgeR v3.4.2 and RPKM values were calculated as the number of counted reads per 1000 mapped and counted bases per geometric mean of normalized read counts per million (RPKM). Translation efficiencies were calculated as the ratio of RPF-RPKM/mRNA RPKM. The Babel framework was used to assess significant changes in translational regulation [6]. We used a mixed model approach for rhythmicity detection (sigmoid and sinusoidal curve fittings). Rhythmic transcripts were defined as those with a peak-to-through amplitude >1.5 and an FDR for rhythmicity detection <0.05.

3. Discussion

We describe here two temporal data sets in mouse liver composed of transcriptome (RNA-seq) and translatome profiling (RPF-seq) with 2-h resolution around the 24-h daily cycle. Rhythmic RNA abundance and rhythmic footprints affected a similar proportion (17%) of the total 10′800 genes detected in both RNA-seq and RPF-seq datasets. The majority of these rhythmic events (≈1200 genes) oscillated both at the mRNA abundance and the translation level. However, we also identified ≈150 genes that showed no daily changes in mRNA abundance but rhythms in translation. These genes are involved in ribosome biogenesis, translation regulation, iron homeostasis and other pathways. Future studies will aim at uncovering the precise mechanisms that engender rhythmicity at the translational level, thus adding a new level of understanding to posttranscriptional control of rhythmic gene expression (Lim and Allada (2013) [1]). In summary, the around-the-clock RPF-seq and matching RNA-seq datasets provide valuable information about the role of translation in generating rhythmic gene outputs, but the data will likely be of high interest to researchers outside the circadian field as well.

Conflict of interest

The authors declare no conflicts of interests.

Acknowledgments

We thank all the staff members at the Lausanne Genomic Technologies Facility and at Vital-IT for sequencing and computational support. DG was supported by Swiss National Science Foundation (professorship grants 128399, 157528); National Center for Competence in Research (NCCR) RNA & Disease; Fondation Pierre Mercier; Fondation Leenaards; Olga Mayenfisch Stiftung; SystemsX.ch StoNets consortium; and the University of Lausanne. PJ was supported by Human Frontiers Science Program long-term fellowship LT000158/2013-L.

References

[1] C. Lim, R. Allada, Emerging roles for post-transcriptional regulation in circadian clocks. Nat. Neurosci. 16 (2013) 1544–1550.
[2] P. Janich, A.B. Arpat, V. Castelo-Szekely, M. Lopes, D. Gatfield, Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res. 25 (2015) 1848–1859.
[3] R. Zhang, N.F. Lahens, H.I. Ballance, M.E. Hughes, J.B. Hogenesch, A circadian gene expression atlas in mammals: implications for biology and medicine. Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 16219–16224.
[4] N.T. Ingolia, S. Ghaemmaghami, J.R. Newman, J.S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324 (2009) 218–223.
[5] D. Gatfield, G. Le Martelot, C.E. Vejnar, D. Gerlach, O. Schaad, F. Fleury-Olela, A.L. Ruskeepaa, M. Oresic, C.C. Esau, E.M. Zdobnov, et al., Integration of microRNA miR-122 in hepatic circadian gene expression. Genes & Development 23 (2009) 1313–1326.
[6] A.B. Olshen, A.C. Hsieh, C.R. Stumpf, R.A. Olshen, D. Ruggero, B.S. Taylor, Assessing gene-level translational control from ribosome profiling. Bioinformatics 29 (2013) 2995–3002.

Translational contributions to tissue specificity in rhythmic and constitutive gene expression

Violeta Castelo-Szekely, Alaaddin Bulak Arpat, Peggy Janich and David Gatfield

Genome Biology (2017) 18:116

Castelo-Szekely et al. Genome Biology (2017) 18:116
DOI 10.1186/s13059-017-1222-2

Genome Biology

RESEARCH Open Access

Translational contributions to tissue specificity in rhythmic and constitutive gene expression

Violeta Castelo-Szekely[1], Alaaddin Bulak Arpat[1,2], Peggy Janich[1] and David Gatfield[1*]

Abstract

Background: The daily gene expression oscillations that underlie mammalian circadian rhythms show striking differences between tissues and involve post-transcriptional regulation. Both aspects remain poorly understood. We have used ribosome profiling to explore the contribution of translation efficiency to temporal gene expression in kidney and contrasted our findings with liver data available from the same mice.

Results: Rhythmic translation of constantly abundant messenger RNAs (mRNAs) affects largely non-overlapping transcript sets with distinct phase clustering in the two organs. Moreover, tissue differences in translation efficiency modulate the timing and amount of protein biosynthesis from rhythmic mRNAs, consistent with organ specificity in clock output gene repertoires and rhythmicity parameters. Our comprehensive datasets provided insights into translational control beyond temporal regulation. Between tissues, many transcripts show differences in translation efficiency, which are, however, of markedly smaller scale than mRNA abundance differences. Tissue-specific changes in translation efficiency are associated with specific transcript features and, intriguingly, globally counteracted and compensated transcript abundance variations, leading to higher similarity at the level of protein biosynthesis between both tissues.

Conclusions: We show that tissue specificity in rhythmic gene expression extends to the translatome and contributes to define the identities, the phases and the expression levels of rhythmic protein biosynthesis. Moreover, translational compensation of transcript abundance divergence leads to overall higher similarity at the level of protein production across organs. The unique resources provided through our study will serve to address fundamental questions of post-transcriptional control and differential gene expression in vivo.

Keywords: Circadian clocks, Translation, Ribosome profiling, Kidney, Liver

Background

Circadian clocks serve organisms to synchronise behaviour, physiology and gene expression according to time of day. The mammalian circadian system consists of a master clock in the brain's suprachiasmatic nuclei (SCN) that receives photic inputs from the retina and synchronises peripheral clocks present in most cells throughout the body. The molecular timekeeping mechanism—the core clock—consists of a network of transcriptional activators and repressors interacting in negative feedback loops (reviewed in [1, 2]). In the core loop, the heterodimeric transcription factor ARNTL:CLOCK (also known as BMAL1:CLOCK) drives the expression of its own repressors, encoded by the *Period* (*Per1*, *Per2*, *Per3*) and *Cryptochrome* (*Cry1*, *Cry2*) genes—a configuration also known as the positive and negative limbs of the oscillator. Additional feedback—in particular, an interconnecting limb involving nuclear receptors of the REV-ERB (encoded by genes *Nr1d1*, *Nr1d2*) and ROR (*Rora*, *Rorb*, *Rorc*) family—intersects with the core loop and numerous post-translational modifications of clock proteins further add to the complexity of the circuitry. The final outcome is a set of robustly cycling transcriptional activities peaking at different phases around the day that drive the rhythmic expression of hundreds to thousands of other genes, termed the clock output or clock-controlled

* Correspondence: david.gatfield@unil.ch
[1]Center for Integrative Genomics, University of Lausanne, Génopode, 1015 Lausanne, Switzerland
Full list of author information is available at the end of the article

genes (CCGs). It is noteworthy that, despite the probably (near-)identical molecular makeup of the core clock across cell types, CCGs show considerable tissue specificity [3]. The co-regulation by core clock and tissue-specific (non-rhythmic) transcription factors may engender such cell type-specific rhythmic expression patterns, as shown to occur in *Drosophila* [4]. Overall, however, the origins of tissue specificity in rhythmic gene output (and even in certain core clock parameters [5]) are poorly understood. Mechanisms that act at the post-transcriptional level and that impact daily messenger RNA (mRNA) and protein accumulation kinetics are plausible players in the generation of cell-type differences as well.

Rhythmic gene expression has been mainly investigated at the transcriptome level, i.e. using mRNA abundances as a primary readout. However, comparison of mRNA levels with datasets of genome-wide transcriptional activity and of protein abundances that have become available recently, has suggested that a surprisingly large fraction of gene expression oscillations may have post-transcriptional origins (reviewed in [6]). The many cases of protein rhythms that are independent of an underlying oscillating transcript (initially reported in a low-throughput mass-spectrometric study from mouse liver ten years ago [7] and recently confirmed at a comprehensive scale [8, 9]) point to important roles for translation, protein degradation and protein secretion in shaping time of day-dependent proteomes. We [10] and others [11] have recently used ribosome profiling, a genome-wide method that assesses translation efficiency through the deep sequencing of ribosome-protected mRNA fragments, to chart the contribution of translational control to daily protein biosynthesis in mouse liver. One conclusion that emerged from the identified cases of translationally generated oscillations was that circadian clock activity and feeding rhythms both contribute to regulating rhythmic gene expression outputs [10, 11]. Notably, the most abundant group of transcripts subject to rhythmic translation, i.e. mRNAs encoding ribosomal proteins and other components of the translation machinery that all contain 5′-terminal oligopyrimidine tract (5′-TOP) sequences regulated by the mammalian target of rapamycin (mTOR) [12], appear to be under the dominant control of feeding [11].

We have now performed ribosome profiling using a second organ from the same cohort of animals, the kidney, which is an emerging circadian model organ with distinct rhythmic functions [13]. By contrasting kidney and liver datasets, we comprehensively assessed commonalities and differences in their translatomes and we evaluated how far the regulation of translation efficiency contributed to tissue specificity in rhythmic and constitutive protein biosynthesis.

Results

Around-the-clock ribosome profiling datasets from two organs

For our recent study of the liver translatome around-the-clock [10], we had used ribosome profiling [14] (RPF-seq) on a time series of organs collected from mice sacrificed every 2 h over the 24-h day (12 timepoints in duplicate; Fig. 1a). To generate a complementary dataset from a second organ, we chose the kidneys from the same cohort of animals. Liver and kidney express thousands of genes in common [3, 15], thus providing a particularly suitable setting for a cross-organ comparison of gene expression.

Applying the same experimental and computational methods as for liver RPF-seq [10, 16], we obtained comparable high-quality data for kidney (see Additional file 1: Figure S1A–C and Additional file 2 for details on sequencing and mapping outcomes). Briefly, ribosome footprints from both organs showed similar enrichment for protein coding sequences (CDS) of mRNAs and depletion of untranslated regions (UTRs) (Fig. 1b). Like the footprints from liver, those from kidney also exhibited excellent reading frame preference, which allowed resolving the 3-nt periodicity of coding sequences transcriptome-wide (Fig. 1c and Additional file 1: Figure S2A, B). Moreover, the high correlation coefficients seen across replicates of the kidney time series for both RNA-seq and RPF-seq data indicated excellent biological and technical reproducibility (Additional file 1: Figure S3A, B). We also used a recently developed tool, termed Ribo-seq Unit Step Transformation (RUST) [17], to confirm high technical similarity of datasets between organs (Additional file 1: Figure S2C, D). Finally, principal component analysis (PCA) on all available datasets (96 libraries, i.e. RPF-seq and RNA-seq from two organs, 12 timepoints, in duplicate) segregated the data according to the main experimental and biological covariates. PC1 (explaining 64.2% of variation) thus separated libraries according to organ, indicating that tissue origin represented the major source of divergence, followed by PC2 (28.4%) that separated RNA-seq (mRNA abundance) and RPF-seq (footprints/translation) (Fig. 1d). The cyclic nature of the data was resolved in the representation PC3 versus PC5 (together 12.5%), in which timepoints assembled to a near-perfect clock (Fig. 1e). The larger circular arrangement of the liver versus kidney time series suggested that rhythmic gene expression from liver contributed more strongly to overall variation than did kidney rhythms. This observation is in line with the notion that there are more and higher amplitude rhythms in liver than in kidney [3]. Taken together, we concluded that the kidney data were of similarly high quality as our previous liver datasets [10]. Together, they would be suitable for comparative analyses of time of day-dependent and constitutive translation across two tissues.

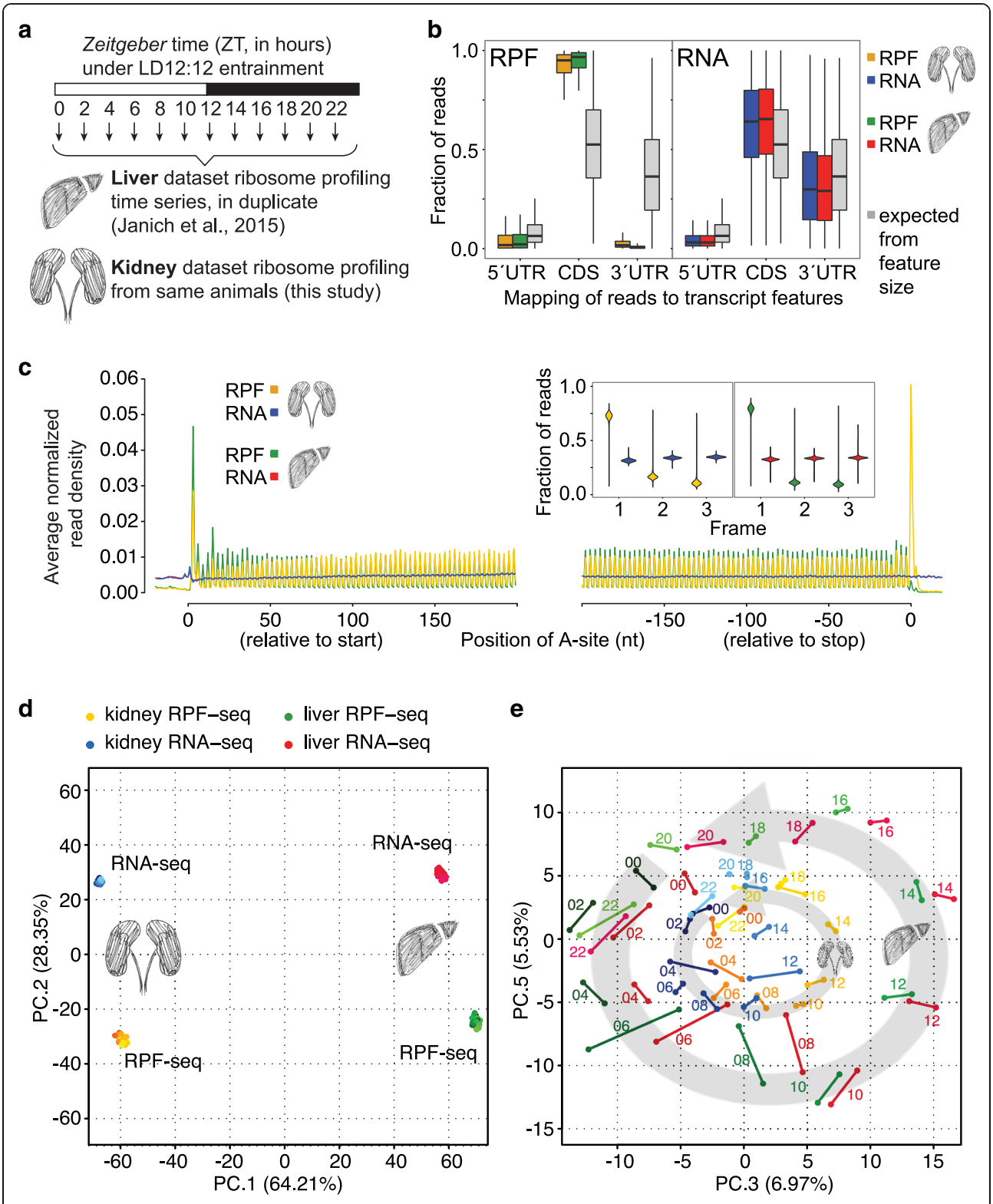


Fig. 1 (See legend on next page.)

(See figure on previous page.)
Fig. 1 Ribosome profiling around-the-clock in mouse liver and kidney. **a** Overview of the experimental design. Livers and kidneys for ribosome profiling were collected every 2 h for two daily cycles. Each timepoint sample was a pool of organs from two animals. Mice were kept under 12 h:12 h light-dark conditions, with *Zeitgeber* times ZT00 corresponding to lights-on and ZT12 to lights-off. **b** Read distribution to transcript features. RPF-seq (*left*; kidney in *orange*, liver in *green*) and RNA sequencing (RNA-seq) (*right*; *blue* and *red* for kidney and liver, respectively) compared with a distribution expected from the relative feature sizes (*grey*; the distributions based on feature sizes were highly similar for both organs, thus only that for kidney is shown). Note that RPF-seq footprints were enriched on the CDS and depleted from UTRs, whereas RNA-seq reads distributed more homogeneously along transcripts, according to feature size. Of note, the higher level of 3′ UTR footprints in kidney resulted mainly from differences in the efficiency with which stop codon footprints were captured, as described in (**c**). **c** Predicted position of the ribosome's aminoacyl tRNA-site (A-site) of reads relative to the CDS start and stop codons. Read density at each position was averaged across single protein coding isoform genes (i.e., genes with one main expressed transcript isoform) that had an average RPF RPKM > 5, a CDS > 400 nt in length and were expressed in both organs ($n = 3037$ genes). This analysis revealed the trinucleotide periodicity of RPF-seq (but not RNA-seq) reads in both organs. *Inset*: frame analysis of CDS reads showed preference for the annotated reading frame (frame 1, the same frame as the start codon) in RPF but not in RNA reads. *Violin plots* extend to the range of the data ($n = 3694$ genes for liver, $n = 4602$ genes for kidney). A separate analysis of the higher level of stop codon footprints in kidney, that also led to the differences in 3′ UTR reads in B, can be found in Additional file 1: Figure S2A, B. **d** Principal component analysis (PCA) of kidney and liver RPF-seq and RNA-seq datasets, using the 4000 most variable genes. The first two components reflected the variability coming from organ (PC1, 64.21%) and from RPF/RNA origin of datasets (PC2, 28.35%). **e** PC3 vs. PC5 (together 12.5% of variation) resolved the factor time within each dataset, leading to a representation that resembled the face of a clock. Each *dot* represents one sample, timepoint replicates are joined by a *line* and timepoints within each dataset are sequentially *coloured*. The circular arrangement was larger for liver than kidney, suggesting a higher contribution of hepatic rhythmic genes to overall variability. Additional file 1: Figure S4 shows the scree plot for the ten first components

Cross-organ differences in translation efficiency are widespread, of moderate scale and partially compensate RNA abundance differences

To what extent do differences in translation efficiency contribute to different gene expression outputs across organs? We addressed this question using the set of 10,289 genes whose expression was detectable in kidney and in liver at both RPF and RNA level (Fig. 2a). From the ratio of CDS-mapping normalised read counts for RPF-seq relative to RNA-seq, we first calculated relative translation efficiencies (TEs) per transcript and for each organ. TEs were overall rather similar between tissues, with 95% of genes falling into a less than threefold range for the kidney/liver TE ratio, as compared with a greater than 100-fold range for the transcript abundance ratio (Fig. 2b). This observation was coherent with the considerably broader spread of mRNA abundances versus TEs across genes within each organ (greater than 500-fold versus just over tenfold, respectively; Additional file 1: Figure S5A, B) and is in line with a dominant role for the regulation of mRNA levels (i.e., transcription and mRNA decay) in controlling quantitative differences in gene output.

Intuitively, we had expected that RNA levels that were widely dissimilar between kidney and liver and subsequently further modulated by organ-specific TEs, would probably give rise to even greater cross-organ divergence at the RPF level. Intriguingly, however, the global correlation between kidney and liver was better for footprint abundances than for transcript abundances (Spearman ρ [RPF]: mean 0.784 vs. ρ [RNA]: mean 0.736; $p < 2.2e\text{-}16$; paired t-test of Fisher-transformed correlation coefficients, n = 24) (Fig. 2c; Additional file 1: Figure S3C, D). This phenomenon was observed irrespective of whether the genes expressed only one dominant protein-coding transcript isoform ('single isoform genes' in the following) that was common to both organs, or whether they gave rise to different (including tissue-specific) mRNA variants ('multiple isoform genes') (Fig. 2d). The observed higher cross-organ concordance of RPFs could have simply had technical reasons, e.g. if the RPF-seq protocol gave more reproducible results than the RNA-seq protocol. We addressed this caveat by comparing measurement errors (MEs) for RNA and RPF data using a similar approach as in a recent publication [18]. We found that MEs scaled inversely with expression levels, as expected, and showed some variation due to organ (Additional file 1: Figure S6A, B, F, G). Especially in liver and among low expressed transcripts, a tendency towards smaller MEs for RPF than for RNA was indeed visible (differences statistically non-significant). In most other cases, however, measurement errors were (in part significantly) higher for the transcripts' RPF counts than for their RNA counts. Of note, the better cross-organ correlation of RPF vs. RNA levels seen in the full transcript set (Fig. 2c) was also evident within various transcript subsets (Additional file 1: Figure S6C–E, H–L), including such subsets for which RPF MEs were higher than RNA MEs (Additional file 1: Figure S6E, L). It is thus unlikely that technical bias was the reason for the higher RPF correlation. Finally, an analysis that we performed on independent ribosome profiling datasets from rat liver and heart [19] allowed us to confirm the phenomenon of higher concordance of RPF versus RNA abundance also between these organs (Additional file 1: Figure S7A–C). Taken together, these findings are suggestive of a potentially broader biological phenomenon that consists in the partial compensation of differences in a gene's mRNA expression through counteracting effects exerted through its TE, resulting in the convergence at

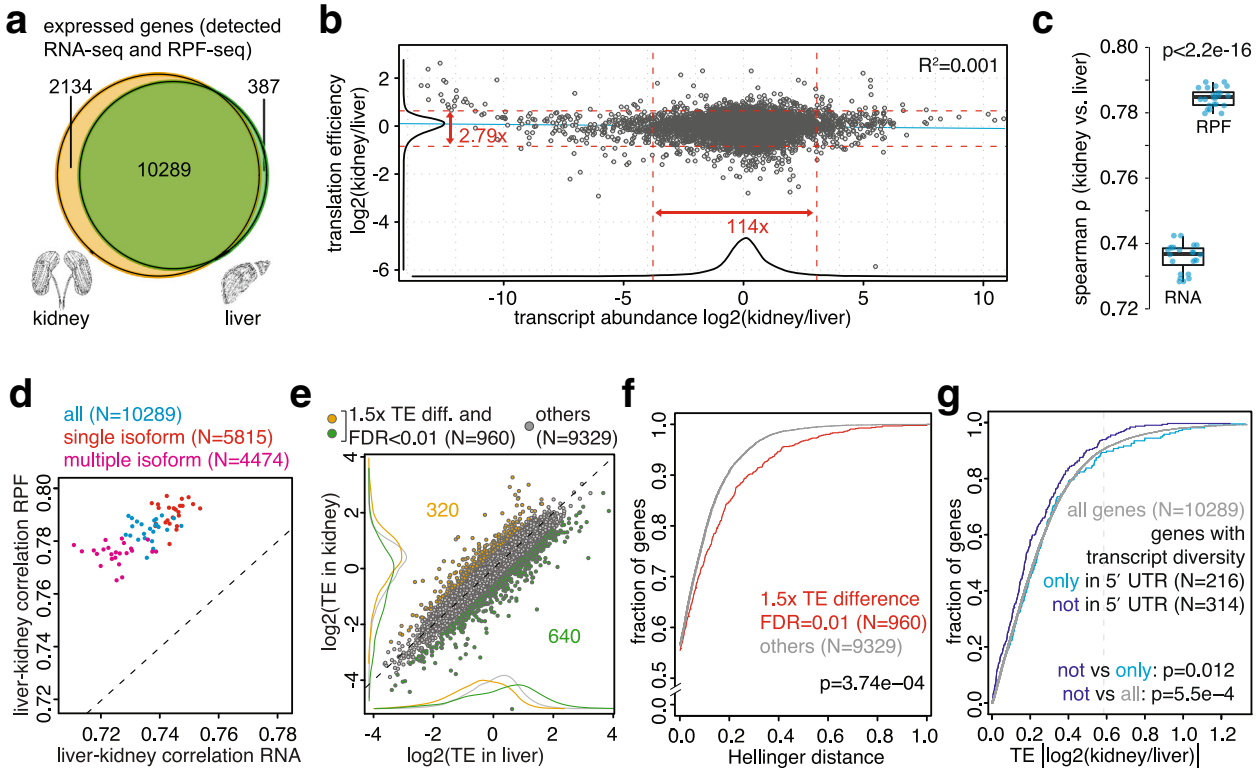


Fig. 2 Cross-organ differences in translation efficiency partially compensate RNA abundance differences and show association with transcript features. **a** *Venn diagram* showing the gene expression overlap (i.e. genes detected at both RPF and RNA level) between kidney (*yellow*, $n = 12{,}423$ genes) and liver (*green*, $n = 10{,}676$ genes). Same cutoffs on RPKM (reads per kilobase of transcript per million mapped reads) were used for both organs. **b** *Scatterplot* of kidney-to-liver ratio of mRNA abundance versus translation efficiency (TE) for all expressed genes ($n = 10{,}289$), averaged over all timepoints. Corresponding density curves are plotted on the margins. *Dashed red lines* represent the 2.5 and 97.5 percentiles of each variable and the corresponding fold-change is indicated. Linear regression line is depicted in blue ($R^2 = 0.0009$, $p = 0.0009$). While 95% of genes spanned a 114-fold range in mRNA abundance differences across organs, the same number of genes changed less than threefold in TE, underlining that transcript abundance was the main contributor to divergent gene expression. **c** Inter-organ Spearman correlation for RNA-seq and RPF-seq samples. Each *dot* represents the correlation coefficient between kidney and liver for a timepoint and replicate sample. Note that RPF-seq samples consistently correlated significantly better than RNA-seq samples ($p < 2.2e\text{-}16$, $n = 24$, paired t-test of Fisher-transformed correlation coefficients). **d** *Scatterplot* of inter-organ RNA vs. RPF correlation coefficients for each sample separately calculated from all (*blue*, $n = 10{,}289$), from single isoform (*red*, $n = 5815$), and from multiple isoform (*pink*, $n = 4474$) genes. Consistently better RPF correlation was evident in all cases. **e** Relative TE in liver vs. kidney (data centred and averaged over all timepoints for all expressed genes, $n = 10{,}289$) showed an overall strong inter-organ correlation. Differential TE—defined as having false discovery rate (FDR)-corrected $p < 0.01$ (Wilcoxon signed rank test on TE) and > 1.5 difference in TE across organs—was apparent for ca. 9% of genes (*yellow* and *green* show cases where TE is higher in kidney and liver, respectively, $n = 960$). **f** Cumulative distribution of Hellinger distances for genes showing differential TE (*red*, $n = 960$), or not (*grey*, $n = 9329$), as detected in (**e**). Hellinger distance was used as a quantitative measure for relative transcript isoform diversity across organs, as described in 'Results' and 'Methods'. The analysis shows that divergent TE correlated with larger diversity in transcript isoform expression ($D = 0.0702$, $p = 3.74e\text{-}04$, two-sample Kolmogorov–Smirnov [KS] test). **g** Cumulative distribution of the kidney-to-liver TE ratio for genes whose transcript diversity originated exclusively from the 5′ UTR (identical CDS and 3' UTR, *light blue*, $n = 216$; these genes show more TE differences across organs) and genes whose transcripts had identical 5′ UTR (and divergent CDS and/or 3′ UTR, *purple*, $n = 314$; these genes show less TE differences across organs). The *vertical dashed grey line* marks the 1.5-fold difference used to define differential TE (as in (**e**)). These results suggested that tissue specificity in TE was partially achieved by expressing transcript isoforms that differed in their 5′ UTRs (note the significant shift towards smaller TE differences for genes with identical 5' UTRs). See also Additional file 1: Figure S9

the level of protein biosynthetic output (footprints, RPF) across tissues.

Transcript features associated with cross-organ differences in translation efficiency

Do particular transcript features have predictive value for organ-specific differences in translation efficiency? To investigate this question, we selected the genes with significantly different TEs between tissues ($n = 5013$; Wilcoxon signed rank test; FDR < 0.01) and implemented a 1.5-fold cutoff on TE ratio between the organs to retrieve the most pronounced cases (n = 960) (Fig. 2e; Additional file 3). Of these, 533 represented 'single isoform genes' with no (or negligible amounts of)

expression of tissue-specific mRNA variants. For these genes, we examined whether a higher TE in kidney (n = 193) or in liver (n = 340) was associated with specific transcript characteristics. Of several features tested, we found that CDS and transcript lengths showed the most significant association with differential TE (Additional file 1: Figure S8A, B). Of note, we had previously seen in liver that shorter coding sequences, i.e., transcripts encoding smaller proteins, are more efficiently translated [10]. Our present analyses suggest that such transcripts are also more prone to tissue-specific regulation at the translational level. Other sequence features showed some bias within the differential TE gene sets as well, although the effects were overall weaker and less consistent. Briefly, the 5′ UTRs of genes with higher TEs in liver were longer and predicted to fold more strongly. By contrast, transcripts with higher kidney TEs were associated with lower 5′ UTR GC content and slightly shorter 3′ UTRs. No association with differential TE was found for the Kozak sequence context score.

We also investigated two functional classes of sequence features, micro RNA (miRNA) binding sites and upstream open reading frames (uORFs), for association with differential TE. Of note, the 960 'TE different' transcripts were not enriched for any predicted miRNA binding sites, making it unlikely that this class of post-transcriptional regulators is a major player in establishing tissue-specific TEs (data not shown). We had previously observed that in the liver the presence of a translated uORF in the 5′ UTR was strongly predictive of low TE at the main ORF [10]. An analogous relationship was also evident in kidney (Additional file 1: Figure S5C). To assess whether uORF translation was associated with TE differences across organs, we compared how the identified uORF-containing transcripts (i.e. single isoform genes showing translated uORFs in at least one organ; n = 1377) distributed to the differential versus non-differential TE gene sets. The group of genes with higher TE in liver was significantly enriched for transcripts with translated uORFs (p = 6.08e-04; Fisher's exact test) and there was slight depletion among genes with higher TE in kidney (not significant) (Additional file 1: Figure S8C). Only few differential TE genes exhibited uORF translation that was exclusive to one organ, but there was a tendency for kidney-specific translation of uORFs to be associated with higher TE on the CDS in liver and vice versa (Additional file 1: Figure S8D). For the genes with uORFs translated in both tissues, we expected that cross-organ differences in the strength of uORF usage would negatively correlate with TE differences at the CDS. However, such a trend was only visible for liver differential TE genes (Additional file 1: Figure S8E); and globally, uORF and CDS TEs even showed slightly positive correlation. In summary, these analyses suggested that uORF translation contributed to some extent (and especially for genes that were more efficiently translated in the liver) to cross-organ differences in TE; however, the overall impact appeared limited (see 'Discussion').

We next included the 'multiple isoform genes' in the analyses and asked whether transcript isoform diversity between the two organs—i.e. the occurrence of tissue-specific mRNA variants generated by alternative transcriptional start sites, splicing and 3′ processing—had any relationship to differential TE. Briefly, using our RNA-seq data we first compiled an inventory of all annotated, protein-coding transcript isoforms and their estimated relative expression levels per gene and tissue. We then used the Hellinger distance [20] as a measure of dissimilarity in isoform expression levels between kidney and liver. A value of 0 for this metric indicates that a gene has identical isoform distribution in both tissues (i.e. these are essentially the 'single isoform genes' described above), while a value of 1 denotes a lack of overlap in expressed isoforms. Globally, the 960 genes with differential TE showed significantly higher Hellinger distances than the remainder of the expressed genes (p = 3.74e-04; Kolmogorov–Smirnov test) (Fig. 2f). Molecularly, the term 'transcript isoform' comprises variations affecting 5′ UTR, CDS and 3′ UTR. By comparing the genes for which all expressed variants affected exclusively one single feature or for which this particular feature was not affected at all, it became apparent that transcript diversity in the 5′ UTR was particularly strongly associated with differential TE (Fig. 2g). By contrast, variation in the CDS showed significantly less association with cross-organ differences in translation efficiency (Additional file 1: Figure S9A, B). Although the low number of available transcripts bearing exclusively 3′ UTR differences precluded a rigorous interpretation, 3′ UTR variation did not appear to be associated with differential TE either (Additional file 1: Figure S9C). Altogether, we thus concluded that TE differences between tissues may, at least in part, have their origin in tissue-specific transcript variants, especially through alternative 5′ UTRs.

Finally, we were interested in whether cross-organ differences in translation efficiency affected specific pathways. For the 640 'TE different' genes that showed increased TE in liver (Fig. 2d), gene ontology (GO) - analyses revealed significant enrichment for categories related to transcription (Additional file 3). Conceivably, tissue-specific translational control of transcriptional regulators may thus impact also on the organs' transcriptomes. The 320 'TE different' genes that were translated better in kidney did not show any significant enrichment.

Translational modulation of phase of oscillation in kidney

We next turned to the analysis of factor time across the datasets. We annotated rhythmic events in kidney with

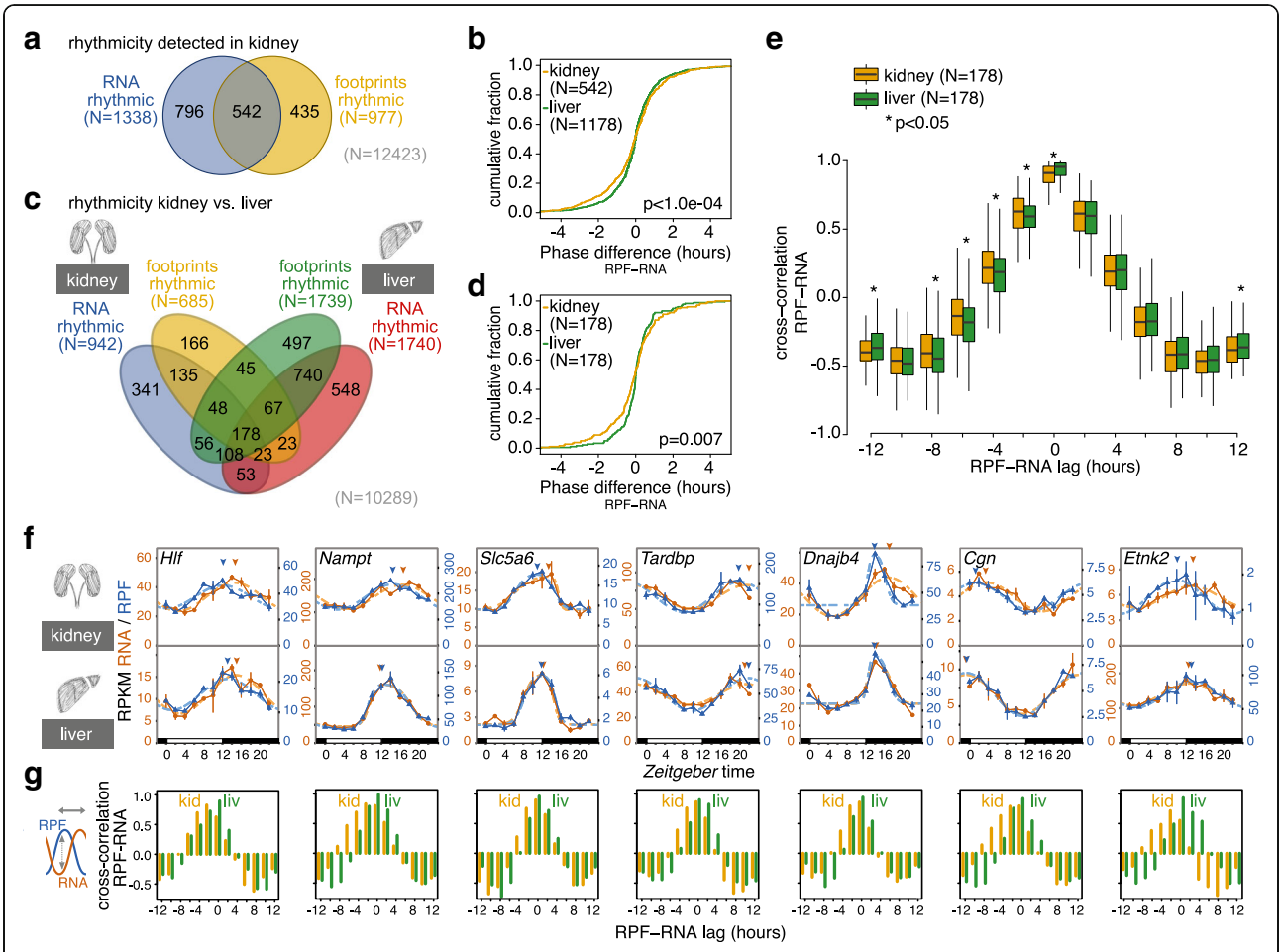


Fig. 3 Rhythmicity analyses across organs reveals phase modulation by translation in kidney. **a** *Venn diagram* showing rhythmic genes in kidney. Of the 12,423 expressed genes, 1338 showed 24-h oscillations of > 1.5-fold amplitude in mRNA abundance (RNA-seq, 10.7%) and 977 in footprint abundance (RPF-seq, 7.9%). A total of 542 genes (4.3%) were identified as rhythmic at both levels. **b** Cumulative distribution of phase differences (RPF peak – RNA peak, in hours) for genes rhythmic at both RNA-seq and RPF-seq in liver (*green*, $n = 1178$) and kidney (*yellow*, $n = 542$). The two distributions were significantly different ($p < 1e\text{-}04$, permutation test) and reflected that maximal footprint abundance frequently preceded mRNA abundance peaks in kidney (note that the two distributions differed mostly in their negative tail). **c** Four-way Venn diagram of rhythmicity sets for genes expressed in both tissues ($n = 10,289$). Of all genes, 364 and 238 were detected as rhythmic in both organs at the RNA-seq and RPF-seq levels, respectively, and 178 genes were detected as rhythmic throughout (i.e. RNA-seq and RPF-seq, in kidney and liver). **d** Cumulative phase difference distribution in liver (*green*) and kidney (*yellow*) for the 178 common rhythmic genes. As in (**b**), the distributions were significantly different ($p = 0.007$, permutation test) and corroborated that even when comparing the same set of genes, footprint peaks frequently preceded mRNA abundance maxima in kidney. **e** Cross-correlation in kidney (*yellow*) and liver (*green*) of time-resolved RPF-seq profiles relative to the RNA-seq profiles of the $n = 178$ common rhythmic genes. The analysis showed that profile correlations for negative lags (i.e. RPF peaking before RNA) were significantly higher in kidney than liver (* indicate $p < 0.05$, Wilcoxon signed rank test). *Boxplots* represent the interquartile range and whiskers extend to the minimum and maximum expression within 1.5 times the interquartile range. **f** Examples for genes with maxima in RPF (*blue*) preceding those in RNA (*orange*) by several hours in kidney (*top*) but not, or less so, in liver (*bottom*). Arrowheads indicate the peaks in footprint and mRNA abundance as estimated from the rhythmic fits. **g** Cross-correlation analyses of RPF-seq relative to RNA-seq profiles (kidney in *yellow*, liver in *green*) for the genes in (**f**). Maximal correlations of the profiles in kidney were found to be shifted to the left (more negative RPF-to-RNA lags) as compared with liver. For liver, there was no case with a maximal correlation value in negative RPF-to-RNA lags

the same methodology as previously for liver, including a 1.5-fold cutoff on peak-to-trough amplitudes [10]. A list of the detected RNA and RPF rhythms and genome-wide gene expression plots are provided in Additional file 4 and Additional file 5 under (https://doi.org/10.6084/m9.figshare.4903193), respectively. Our analyses yielded 1338 and 977 genes that cycled at the RNA abundance and footprint level, respectively, with an overlap of 542 genes (Fig. 3a). As discussed later, this relatively modest overlap (542 genes corresponds to 41% and 55% of all 'RNA rhythmic' and 'footprint rhythmic' cases, respectively) likely underestimates the

full extent of shared rhythmicity and only contains the most robustly oscillating gene expression events, which we further explored in the following.

Interestingly, the analysis of rhythmicity parameters across the 542 genes revealed that the timing of their RPF peaks relative to their RNA peaks had a significantly different and broader distribution than the corresponding set from liver ($p < 1.0e-04$; permutation test) (Fig. 3b). This observation suggested that the phase of protein biosynthesis rhythms was subject to marked translational modulation in kidney. In liver, by contrast, RPF peaks were more tightly gated by RNA abundance peaks. Surprisingly, maximal translation tended to precede maximal RNA abundance in kidney (Additional file 1: Figure S10A), as globally the mean RPF peak phase was advanced (−0.123 h) and also RPF rhythms were enriched for phase advances (282) versus delays (260), albeit neither reaching statistical significance ($p = 0.16$, Wilcoxon rank sum test).

The above analyses used different rhythmic gene sets for kidney than for liver, potentially compromising comparability. The observed differences in the RPF-RNA phase relationships could thus have simply arisen from transcript-specific rather than from tissue-specific differences in the timing of translation. We thus analysed the group of 178 genes whose RNA and RPF profiles were rhythmic in both organs (Fig. 3c; Additional file 6 and Additional file 7 (https://doi.org/10.6084/m9.figshare.4903193)). Again, the distribution of RPF-RNA offsets was significantly broader in kidney than in liver (Fig. 3d; $p = 0.007$, permutation test) with an RPF peak phase advance in kidney (mean −0.143 h) and a phase delay in liver (mean 0.036 h) (Additional file 1: Figure S10B, C). We next calculated the gene-wise RPF-RNA peak phase difference in kidney relative to that in liver. More genes showed their RPF maxima earlier (96) than later (82) in kidney versus liver, with a mean advance of −0.178 h (Additional file 1: Figure S10D), but again without passing statistical significance ($p = 0.152$, Wilcoxon rank sum test).

Conceivably, we lost statistical power and introduced error in the above analyses by restricting the phase comparisons merely to the peaks of the rhythmic curve fits. We thus sought a method that would take into account phase differences between RPF and RNA profiles over all data points. To this end, we used cross-correlation to quantify the similarity between the RPF and RNA time series as a function of sliding one series on the time axis relative to the other. When the time series were not shifted against each other at all (RPF-RNA lag = 0 h), the RPF-RNA cross-correlation values were overall highest, as expected, and they were significantly higher in liver, in line with stronger gating of RPF rhythms relative to RNA oscillations in this organ (Fig. 3e). Importantly, when cross-correlation of RNA was calculated with earlier RPF time points (negative RPF-RNA lags; see in particular lags of −2 h to −8 h in Fig. 3e), kidneys scored significantly higher than livers. Sliding the series in the other direction, however, rather led to overall better correlations in the liver (see lags of +4 to +8 in Fig. 3e; liver–kidney difference was non-significant). Taken together, these analyses underscored that there was asymmetry in the data with RPF rhythms preceding RNA rhythms specifically in the kidney.

We confirmed kidney-specific translational phase advances by visual inspection of individual gene expression profiles. Figure 3f shows the profiles for the genes *Hlf*, *Nampt*, *Slc5a6*, *Tardbp*, *Dnajb4*, *Cgn* and *Etnk2*, which all show an RPF phase advance of up to several hours relative to RNA. Cross-correlation analysis for the individual genes also confirmed kidney-specific, phase-advanced translation (Fig. 3g).

At first sight, translation that is phase-advanced to mRNA abundance is counterintuitive. Conceivably, it may occur when translation efficiency is not constant, but decreases over the lifetime of an mRNA. TEs may be higher on freshly synthesised messages that have long poly(A) tails and decrease as a result of gradual deadenylation even before transcript stability and abundance are affected as well [21]. In keeping with the hypothesis of cross-organ differences in poly(A) kinetics, we have observed that most subunits of the major cytoplasmic deadenylase complex, CCR4-NOT, are significantly more highly expressed in kidney than in liver (Additional file 1: Figure S11A–C). Higher deadenylase activity in kidney could provide an attractive molecular explanation for the observed tissue-specific differences in RPF-RNA phasing and for RPF rhythms that are phase-advanced to RNA oscillations.

High tissue divergence in translationally driven rhythms

Rhythmicity detection algorithms are sensitive to false-negatives, i.e. to classify gene expression profiles as 'non-rhythmic' (for example, because they fail imposed thresholds on amplitude or FDR) although the underlying temporal patterns may still be more similar to, and more likely to be, rhythmic than invariable. Of note, the lack of canonical methods to reliably determine true absence of rhythms is a common problem in the field (see recent review by [6] for discussion). Venn diagrams that simply overlap rhythmic gene sets hence need to be interpreted with caution. For these reasons, the extent of 'RNA only' and of 'footprints only' oscillations in Fig. 3a is likely not reported reliably and subject to overestimation. The heatmaps of the corresponding RNA and RPF profiles support this notion as well (Additional file 1: Figure S12B, D).

In order to identify the true-positive 'translation only' cycling transcripts with higher reliability, we

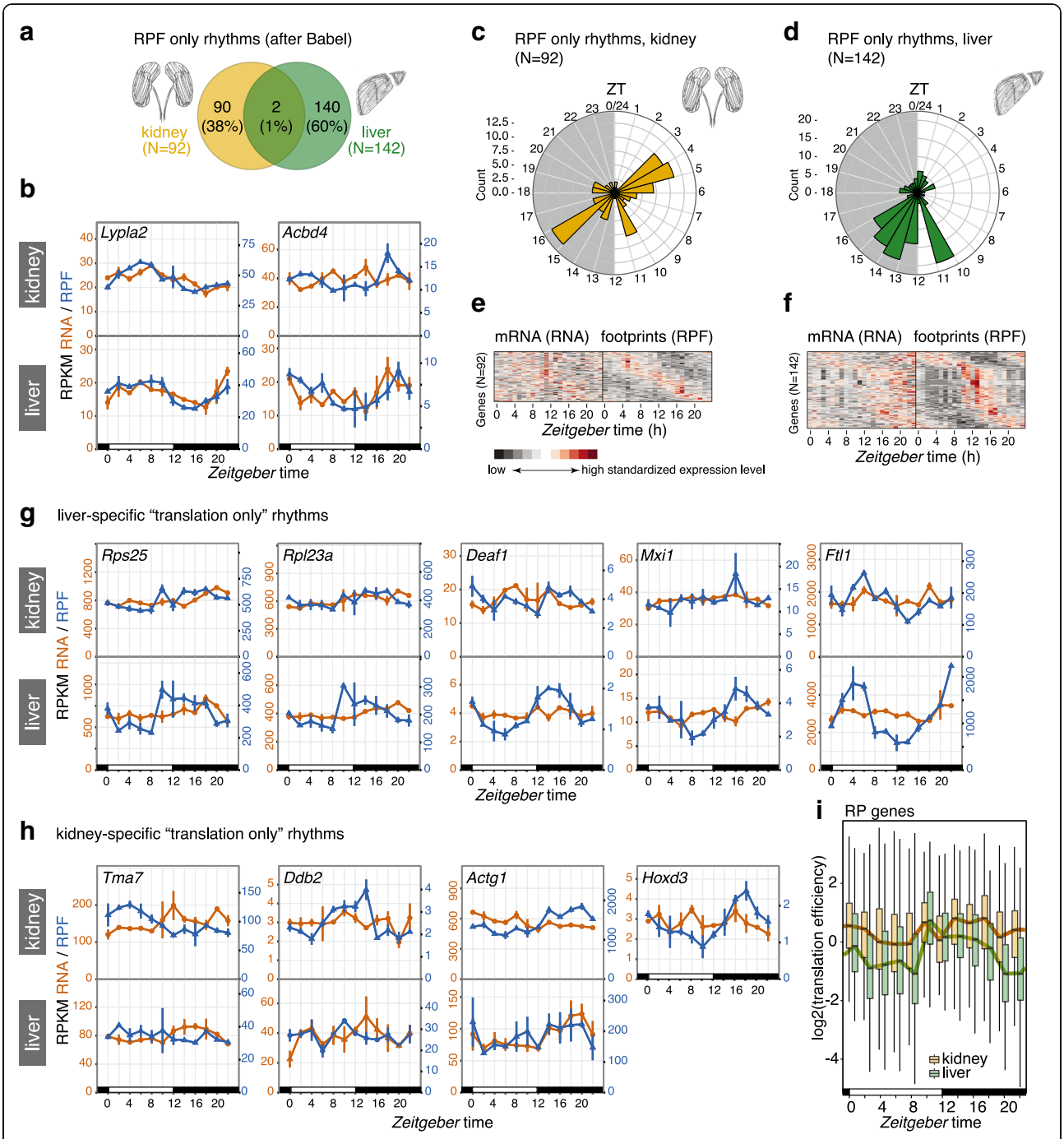


Fig. 4 High tissue divergence in translationally driven rhythms. **a** *Venn diagram* of rhythmic RPF-seq sets in kidney (*yellow*, $n = 92$) and liver (*green*, $n = 142$) after the Babel analysis indicated strong tissue specificity of translational control. **b** Daily profiles of RPF-seq RPKM (*blue*) and RNA-seq RPKM (*orange*) for the two genes detected as translationally regulated in both tissues in (**a**). **c**, **d** Circular phase histogram for the 92 (**c**, kidney) and 142 (**d**, liver) genes showing footprint rhythmicity in the organs. Note that the translational upregulation of transcripts observed at the day-to-night transition in liver was absent in kidney. **e**, **f** *Heatmaps* of RNA (*left panels*) and RPF (*right panels*) rhythms for the 92 and 142 genes translationally regulated in kidney (**e**) and in liver (**f**), respectively. Genes are sorted by footprint phase and expression levels are standardized by row (gene). These sets of genes showed rhythmicity in footprint abundance but no oscillation in mRNA. **g**, **h** Daily profiles of RPF-seq RPKM (*blue*) and RNA-seq RPKM (*orange*) for representative examples of translationally generated rhythms specific for liver (**g**) and kidney (**h**). For each gene, the *upper panel* shows the kidney data and the *lower panel* the liver data. *Hoxd3* was not expressed in liver. **i** Translation efficiency (TE) around-the-clock for ribosomal protein (RP) genes expressed in liver (*green*, $n = 86$) and in kidney (*yellow*, $n = 89$). For each timepoint (ZT) *boxplots* represent the interquartile range and *whiskers* extend to the minimum and maximum TE within 1.5 times the interquantile range. *Lines* connect the median of each *boxplot* to ease visualization. Note the global TE upregulation at ZT10 in liver, whereas TEs in kidney remain high over the day

implemented the same methodology as in our previous study [10]. Briefly, we used the analytical framework *Babel* [22] to preselect all transcripts whose translation efficiency changed significantly over the day (and/or whose TEs deviated significantly from the global transcript population). Rhythmicity analyses were then performed on this gene subset and yielded 92 cases with the sought-after temporal profiles of rhythmic translation on non-rhythmic mRNAs (Fig. 4a). Comparison with the 142 genes of the analogous set from liver revealed near-perfect tissue specificity of translationally driven oscillations. Only two genes, *Abcd4* and *Lypla2*, were shared between the organs; they were both among the least compelling cases of 'translation only rhythms' that our method had identified, as judged by visual inspection (Fig. 4b).

Interestingly, not only the identity of rhythmically translated genes, but also the time-of-day at which the majority of rhythmic translation events occurred, was highly tissue-specific. The phase histograms thus showed striking differences in the peak time distributions between the organs (Fig. 4c, d; difference in distributions: $p = 1.66e\text{-}04$; $W = 17.403$, $df = 2$; Watson–Wheeler test for homogeneity of angles). Of note, the enrichment for translational maxima at the light–dark transition (*Zeitgeber* time, ZT10-16; ZT00 corresponds to lights-on and ZT12 to lights-off) that dominated the distribution in liver (Fig. 4d, f) was virtually absent from kidney (Fig. 4c, e). Instead, kidney showed enrichment for transcripts with maximal translation occurring around ZT4 and ZT16. Visual inspection of individual examples confirmed the organ specificity of RPF rhythms. The cases of robust translational oscillations that we [10] and others [11, 12] had previously identified in liver were thus absent or severely blunted in kidney. This included mRNAs encoding ribosomal proteins (RPs), which make up the bulk of genes showing a translational surge at the light–dark transition (e.g. *Rps25*, *Rpl23a*), as well as transcripts encoding the transcription factors *Deaf1* (deformed epidermal autoregulatory factor 1) and *Mxi1* (MAX interactor 1), and mRNAs containing iron-responsive elements in their 5′ UTRs (e.g. Ferritin light chain 1, *Ftl1*) (Fig. 4g), all of which we had previously reported as translationally rhythmic in liver [10]. Rhythmic translation exclusive to kidney was not significantly enriched for particular pathways (data not shown) and the temporal profiles were overall of lower amplitude than those seen for liver; *Tma7* (translational machinery associated 7 homolog), *Ddb2* (damage-specific DNA binding protein 2), *Actg1* (actin, gamma, cytoplasmic 1) and *Hoxd3* (homeobox D3; not expressed in liver) were among the most distinct examples (Fig. 4h).

In summary, we concluded that temporal changes in TE were strikingly tissue-specific and overall relatively rare in kidney. Specifically for transcripts encoding RPs and other components of the translation machinery, which are the most prominent group of TE rhythmic genes in liver, it has been suggested that feeding-dependent mTOR-signalling underlies the translational upsurge at the light–dark transition via a mechanism involving the 5′-terminal oligopyrimidine (5′-TOP) motifs that these transcripts carry [11, 12]. Interestingly, the TE comparison between both tissues revealed that kidney RP translation occurred at a relatively high level throughout the day (Fig. 4i). The lack in rhythmicity for RP genes in this organ may thus result from an absence of translational repression during the light phase rather than a lack in activation in the dark phase. It may indicate that the kidney is less sensitive to systemic cues engendered by feeding and fasting (see 'Discussion').

Different degrees of tissue specificity in core clock gene expression at the level of RNA abundance and protein biosynthesis

Clocks exhibit functional differences across cell types and organs, for example at the level of rhythmicity parameters (e.g. free-running period and phase [5]), of clock output gene repertoires [3], of oscillator strength and robustness [23, 24] or with regard to clock gene loss-of-function phenotypes [25]. Conceivably, the precise timing and level at which the various clock proteins are produced may modulate properties of the clock circuitry and underlie some of the abovementioned functional variations. In order to investigate these possibilities, we compared the expression of core clock components in both organs.

We first investigated transcript and footprint RPKMs as averages over timepoints to assess the cumulative daily production of clock RNAs and proteins. Most core clock genes showed a considerable degree of organ specificity in their expression levels that was readily appreciable in the footprint versus transcript abundance representation with both organs overlaid in a single graph (Fig. 5a). Two tissue differences caught our particular attention. First, the balance between the transcriptional activators *Rora/Rorc* and repressors *Nr1d1/Nr1d2* differed markedly between organs and was skewed towards repression in kidney (i.e. higher *Nr1d1/2* and lower *Rora/c* RPKMs in kidney, Fig. 5a, b). These transcriptional regulators bind to shared sequence elements on DNA and form the 'interconnecting limb' within the rhythm-generating clock circuitry. In addition, they also control an output branch of the oscillator [1, 2]. It is hence conceivable that adjusting the relative levels of NR1D1/2 versus RORs tailors clock-controlled gene expression in a tissue-specific fashion. Our observation of an active state of this output branch in liver and a more repressed state in kidney is fully consistent with the current knowledge of its target genes and knockout phenotypes, which point to a prominent role in the regulation

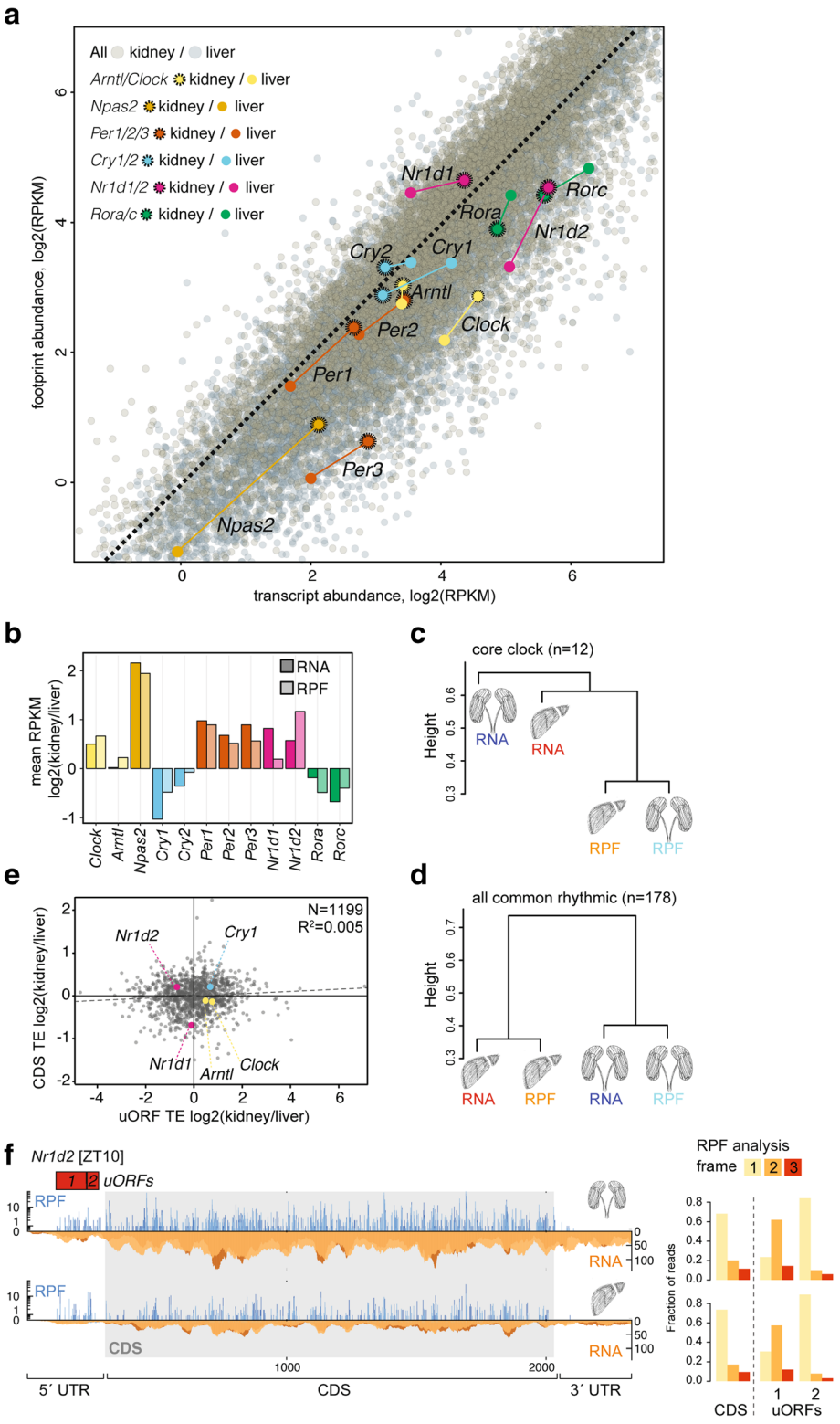


Fig. 5 (See legend on next page.)

(See figure on previous page.)
Fig. 5 Tissue specificity in core clock gene expression at the level of RNA abundance and translation. **a** *Scatterplot* of transcript abundance (RNA-seq) vs. footprint abundance (RPF-seq) for liver (*grey*) and kidney (*sepia*) ($n = 10{,}289$), where core clock components are highlighted (kidney, *dots with dashed circles*). *Coloured dashed lines* join the relative locations of each core clock gene between organs. **b** *Bar graph* of the average RPKM ratio between kidney and liver for the main circadian core clock genes, at the level of mRNA abundance (*dark shades*) and ribosome footprints (*light shades*) suggested that translational compensation led to higher similarity at the level of protein biosynthesis (RPF) for several core clock genes. **c** Hierarchical clustering of the organs' RNA and RPF profiles based on the similarities of the core clock genes expression patterns ($n = 12$, genes shown in B). The height of the branches represents weighted average distances over the considered genes (see 'Methods'). Note that RPF rhythms in two organs were more similar than RNA and RPF rhythms within an organ. **d** Hierarchical clustering as in (**c**) based on the genes detected as rhythmic throughout ($n = 178$, see Fig. 3c). When compared to the clustering based on core clock gene expression patterns in (**c**), this rhythmic gene set showed an organ-based clustering. **e** *Scatterplot* of kidney/liver ratios of uORF vs. CDS translation efficiencies for genes containing AUG-initiated translated uORFs in both organs ($n = 1199$). uORF-containing core clock genes are highlighted. As also shown in Additional file 1: Figure S8E, differential uORF usage could not globally explain differences in CDS TE across organs (note the lack of negative correlation between the two variables, $R^2 = 0.005$, $p = 0.008$). As an exception, the lower uORF TE of *Nr1d2* might have a role in setting relatively higher CDS TE in kidney. **f** RPF (*blue*) and RNA (*orange*) reads mapping along the *Nr1d2* transcript in kidney (*top*) and liver (*bottom*) for the timepoint of maximal CDS translation (ZT10). 5′ UTR and CDS are shown in full, but for better visualization only a portion of the 3′ UTR (the same length as the 5′ UTR) is shown. *Red boxes* indicate the predicted AUG-initiated translated uORFs. *Right panels* show that, similar to the CDS, the uORFs showed clear frame preference, indicative of active translation

of hepatic pathways such as lipid, cholesterol and bile acid metabolism [26].

A second tissue difference concerned the main constituents of the negative limb, the Period (*Per*) and Cryptochrome (*Cry*) genes. Heterotypic PER:CRY protein complexes inhibit ARNTL:CLOCK-driven transcriptional activity and thus lie at the core of the oscillator's principal negative feedback loop. We observed a shift to more PER (and slightly less CRY) biosynthesis in kidney (Fig. 5b). PERs (and in particular PER2) are considered stoichiometrically rate-limiting components of the PER:CRY complex and increased PER2 dosage engenders long periods [27, 28]. Interestingly, tissue explant experiments have shown that kidney clocks free-run with almost 1.5-h longer periods than liver clocks [5], as would be predicted from the increased PER biosynthesis that our analyses revealed. As a more general concept, we deem it conceivable that the modulation of biosynthesis levels for individual clock proteins may be a more general mechanism to engender distinct differences in clock parameters across cell types.

We noted that for the majority of clock genes (*Npas2*, *Cry1*, *Cry2*, *Per1*, *Per2*, *Per3*, *Nr1d1*, *Rorc*) the tissue differences were less pronounced at the RPF than at the RNA level (Fig. 5b), indicating that translation efficiencies partially counteracted RNA expression differences. Only in four cases (*Clock*, *Arntl*, *Nr1d2*, *Rora*) TEs exacerbated transcript abundance differences and led to higher tissue differences at the RPF level. Interestingly, this observation could also be made in the time-resolved data. As a measure of similarity between expression profiles that takes into account profile shape and expression level, we used the Euclidean distances calculated between the four rhythmic traces of each individual gene (i.e. RNA and RPF in kidney and liver; Additional file 1: Figure S13). Hierarchical clustering of the similarities for the ensemble of the 12 main core clock genes showed that RPF profiles from the two organs grouped together (Fig. 5c). The temporal profiles of clock protein biosynthesis between organs were thus more similar than RNA and RPF expression profiles within organs. By contrast, the 178 common rhythmic genes identified in Fig. 3c—serving as a control set for this analysis—revealed within-organ clustering (Fig. 5d). These findings underscored that translational compensation was occurring within the core clock, where it led to more similar expression profiles in clock protein biosynthesis than would have been predicted from the rhythmic RNA abundance. This phenomenon was, however, not a general feature of all rhythmic gene expression.

The transcriptome-wide analyses described further above had shown only weak signs of association between cross-organ differences in TE and in uORF usage (Additional file 1: Figure S8C–E). However, we knew from our previous work in liver that at least five core clock transcripts (*Nr1d1*, *Nr1d2*, *Cry1*, *Clock*, *Arntl*) contained translated, potentially regulatory, AUG-initiated uORFs [10]. We therefore examined whether for any of these concrete cases there was evidence for a connection between uORF translation and cross-organ TE differences. The read distribution along the transcripts (Additional file 1: Figure S14A) and the marked frame preference of RPF reads (Additional file 1: Figure S14B) confirmed that the footprints mapping to our annotated uORFs likely reflected active translation. However, only in one case, *Nr1d2*, there was a distinct anticorrelation between uORF usage and TE differences on the CDS (Fig. 5e). *Nr1d2* contains two translated uORFs in the 5′ UTR (Fig. 5f), whose decreased usage in kidney was accompanied with higher TE on the CDS in this organ (Fig. 5e). For *Nr1d2*, differential uORF usage could thus represent a plausible mechanism that contributes to regulating organ-specific gene expression output at the translational level, keeping NR1D2 biosynthesis low in liver and high in kidney.

Discussion

Given that the functionally relevant output of most gene expression is the protein, quantitative and genome-wide analyses of protein biosynthesis are of high interest to complement the wealth of transcriptomics data that are already available. Of note, only the fairly recent development of the ribosome profiling technique [29] has made it possible to analyse translational events in a quantitative, high-throughput fashion. Our study of two paradigms of differential gene expression, i.e. its tissue-dependence and its time of day-dependence, is among the first of its kind and, together with the associated datasets and resources, will likely be of wide interest and utility to researchers working in the chronobiology and gene expression fields.

We have addressed several, rather fundamental questions that go beyond the chronobiological focus of the study: How does the dynamic range of translation efficiency compare to that of transcript abundances across two distinct organs of an animal? Is translation efficiency a default transcript property and comparable across two tissues or do TEs become reinterpreted depending on cell type or organ? Does cross-tissue variability of TEs come with any direction, i.e. is there a global tendency to either reinforce or to counteract transcriptomal differences?

To our knowledge, only one previous study has reported on ribosome profiling datasets from two complementary mammalian tissues: rat liver and heart [19]. This study also included animals with different genetic backgrounds as covariates in the experimental design and its main focus was on strain differences in translation rather than on tissue differences. Our analyses based on more than 10,000 genes commonly expressed in liver and kidney show that cross-organ TE differences are widespread, but of limited magnitude. Across genes in a tissue and for individual genes between tissues, the dynamic range of translation efficiencies is thus about 30–50-fold narrower than that of transcript abundances. These findings are coherent with the view that major differences in gene expression are set up at the level of transcription (possibly with some influence coming from RNA stability as well), whereas differences in translation rate have more of a modulatory role. It is intriguing that this modulation is overall characterised by directionality, with TE differences between tissues globally counteracting some of the mRNA abundance differences. Such translational compensation has previously been observed for divergent transcript expression levels across yeast species [30] and across different rat strains [19]; our study now extends this observation to gene expression across organs. Moreover, the idea of translational compensation is conceptually similar to findings that proteomes are evolutionarily more highly conserved than transcriptomes [31, 32]. As an underlying common principle, these cases may indicate that selective pressure on precise gene expression levels likely acts on protein abundances, whereas a certain degree of variability (even noise) in RNA levels may be tolerated without further consequences. It will be exciting to study the underpinnings of translational compensation further, across tissues and across species.

Maybe not unexpectedly, there was no dominant, distinct sequence feature that could serve as a predictor for cross-organ TE differences. Rather, we found several associations with a number of transcript characteristics. Conceivably, these contribute collectively to modulating TEs in concert with the specific cellular and tissue environment and possible cell-type differences in the translation machinery including its regulators and trans-acting factors. While our ribosome profiling studies have allowed us to record the outcome of such regulation at high resolution, understanding its causes represents an exciting challenge for the future. For now, we can only infer that an overarching theme of the identified associations is a connection to 5′ UTRs, which is in also in line with the notion that initiation is rate-limiting for most translation events. We thus observed associations of cross-organ TE differences with 5′ UTR length, with uORF usage, with GC content and folding potential, as well as with transcript isoform diversity that affected the 5′ UTR. We would like to point out that comprehensive uORF annotations remain a bioinformatics challenge that is far from resolved. We have therefore restricted our analyses to AUG-initiated ORFs, inevitably leading to a bias towards false-negatives in uORF annotation. As we will learn how to annotate uORFs more comprehensively and more precisely in the future, it may be worth revisiting the relationship between differential TE and uORF translation in our datasets in order to evaluate whether a clearer role for these regulatory sequence elements will emerge.

Our study has led to novel insights into rhythmic gene expression. The extent to which rhythmicity is generated by the temporal regulation of translation has been the subject of speculation ever since the first report of rhythmic proteins encoded by non-rhythmic mRNAs [7]. Our kidney datasets complement recent time-resolved ribosome profiling data from liver [10, 11] and from a cell line [33]. As compared to liver, the number of transcripts subject to translational rhythms in kidney is slightly lower, but overall in a similar order of magnitude with around 1% of the transcriptome affected. It came as a surprise that translational rhythms were essentially tissue-specific in terms of the affected genes and the phase distributions. A possible explanation could be that these rhythms are driven by rhythmic systemic cues to which tissues do not respond equally. The effects of feeding and mTOR signalling, for example, may be more pronounced in liver due to the dedicated role that this organ plays in energy homeostasis and fasting responses,

thus explaining the differences in translational oscillations for RP genes. Beyond the role that translation has in generating rhythms, our analyses have pointed to an additional, rhythmicity-modulating role that appears to affect gene expression quite broadly, i.e. the timing of the phase of protein biosynthesis oscillations relative to that of mRNA abundance rhythms. Consistent with work by the Green lab that showed interactions between polyadenylation status of mRNAs and rhythmic protein expression in the liver [21], it is tempting to speculate that related mechanisms are operative across organs, with tissue-specific deadenylation kinetics tuning the timing of rhythmic protein biosynthesis. Finally, our study is a first attempt to resolve tissue differences in core clock gene expression as a factor contributing to functional differences of the oscillator. It is interesting that the core clock mechanism has a long-standing history of being referred to as a 'transcription-translation feedback loop' [34], although the actual feedback occurs at the transcriptional level and possible mechanistic functions of translational regulation have not been much investigated. Our cross-organ comparison of core clock protein biosynthesis suggests that translational control—including through the activity of uORFs [10, 33]—is of regulatory interest and represents a way by which the identical set of core clock genes could form circuitries with different stoichiometry of its main components. As a result, both clock parameters and output gene repertoires may be organ-specifically tuned.

Conclusions

How translational differences contribute to overall gene expression diversity is still poorly understood. Our study uncovered translational changes that occur across two paradigms of regulated gene expression, i.e. around-the-clock and between tissues. Daily gene expression rhythms generated at the translational level were strongly organ-specific with regard to the identities and phase distributions of affected genes. Moreover, our data indicate that translation efficiency differences between organs can adjust the timing of protein production from rhythmic mRNAs and the levels of core clock protein production, in agreement with the tissue specificities observed in clock output gene sets and clock parameters. Together, these results are consistent with an important role of post-transcriptional mechanisms in mammalian circadian gene expression regulation. Beyond the temporal dimension, we have explored constitutive protein biosynthesis across organs. Our quantitative analyses underscore that gene expression divergence is largely programmed at the transcript abundance level. Interestingly, the widespread differences in translational efficiency that we detected between organs even serve to achieve higher concordance in protein production between tissues. Conceivably, such translational compensation reflects a selective pressure to maintain precise protein levels rather than mRNA levels. The high-resolution genome-wide translatome datasets generated in this study will allow further explorations into the mechanisms of post-transcriptional control and differential gene expression in vivo.

Methods

Animals

Twelve-week-old male mice (C57BL/6 J; Janvier Labs) were entrained for two weeks to light:dark 12:12 cycles with ad libitum access to food and water and were anesthetized (isoflurane) and sacrificed every 2 h (ZT0–ZT22, with ZT0 corresponding to 'lights-on') for two daily cycles. Livers and kidneys were removed and processed either directly or flash-frozen in liquid N_2.

Ribosome profiling

Generation of liver RPF-seq and RNA-seq libraries using the ARTseq ribosome profiling kit (Epicentre) was described recently [10, 16]. Kidney libraries were prepared in the same manner, with a single modification to the order of steps in RPF library preparation. After RNase treatment and recovery of ribosome-protected fragments, 5 μg of material was first ribosomal RNA (rRNA)-depleted (Ribo-Zero magnetic kit, Epicentre) and then purified by 15% PAGE. In the formerly prepared liver libraries, Ribo-Zero treatment and PAGE purification had been inverted because at the time we had found that changing the order had a beneficial effect on obtaining highly concentrated libraries. For the kidney samples, however, we noted that this modified order led to higher contamination with reverse-strand rRNA probes bleeding from the Ribo-Zero kit and we thus reverted to ARTseq's original order. All other steps and materials were identical between liver and kidney samples and followed the ARTseq ribosome profiling kit instructions. RPF and RNA libraries were sequenced on an Illumina HiSeq 2500.

Sequencing data processing, alignment and quantification

Processing, quality assessment, alignment and quantification of sequencing data were performed as described previously [10, 16]. Briefly, after adapter trimming using Cutadapt [35], the length distribution of trimmed reads was used to assess the quality of nuclease digestion and size-selection, which is particularly important for RPF libraries (Additional file 1: Figure S1B). Trimmed reads were filtered by size (26–35 nt for RPF; 21–60 nt for RNA) using an in-house Python script and sequentially mapped to mouse rRNA, human rRNA, mitochondrial tRNA, mouse tRNA, mouse cDNA (Ensemble mouse database release 75) using Bowtie v2.2.1 [36] and mouse genome (GRCm38.p2) using Tophat v2.0.11 [37]. Trimmed and filtered sequences were also directly mapped against the mouse genome (Tophat

v2.0.11) in order to estimate expressed transcript models in each organ (using Cufflinks v2.2.1 [38]). Transcriptome-mapping reads in the sequential alignment were counted towards their location into the 5′ UTR, CDS or 3′ UTR of the transcript, based on feature annotation (Ensemble mouse release 75). Mappable and countable feature lengths were not calculated for this study (see 'faux reads analysis' in the 'Quantification of mRNA and ribosome footprint abundance' section of Supplemental Experimental Procedures of previous study [10]) as its contribution was negligible for further analyses. Therefore, RPKM calculations in this study were not corrected with such factor. Read counts in RNA-seq and RPF-seq datasets were normalised with upper quantile method of edgeR [39] and RPKM values were calculated as the number of reads per 1000 bases per geometric mean of normalised read counts per million. Relative translation efficiencies (TE) were calculated as the ratio of RPF-RPKM to RNA-RPKM per gene per sample. Reading frame and nucleotide periodicity analyses were performed as in [10]. PCA relied on a combined matrix of CDS counts for RPF and RNA from both liver and kidney and following the same approach as before [10]. Ribo-seq Unit Step Transformation (RUST) analysis was used to assess whether the sequencing libraries were globally of similar quality in terms of their local footprint densities [17]. RUST is a simple normalisation method that reduces the heterogeneous noise in the data and allows identification of mRNA sequence features that affect footprint densities globally. We used the version 1.2 of the published rust_codon.py standalone python script with minor modifications to reflect the experimental settings as closely as possible (i.e. A-site offsetting). RUST codon profile and corresponding Kullback–Leibler (K–L) divergence for each library (RPF and RNA) was generated against a database of 8012 single protein isoform transcript sequences using all mapped reads with a length of 28–32 nt. The K-L divergences from all samples for each combination of tissue (kidney or liver) and read type (RPF or RNA) were used to generate K-L profiles at the 0, 10, 25, 50, 75, 90 and 100th quantiles.

Correlation analyses and assessment of translational compensation across organs

Correlation of RNA-seq and RPF-seq across organs: kidney versus liver correlations at the levels of RNA-seq and RPF-seq (i.e. Fig. 2b; Additional file 1: Figure S6C–E, H–L) were calculated in a pairwise fashion for each of the 24 samples (12 timepoints, two replicates/timepoint), as livers and kidneys of each replicate originated from the same animals. Significance of the difference in the Spearman coefficients between both distributions was assessed by paired t-test on Fisher z-transformed coefficients. Heart versus liver correlation at the levels of RNA and RPF-seq (Additional file 1: Figure S7A) was calculated from the study [19], using the BN-Lx reference rat strain data. Since the five heart and liver replicates in this study did not come from the same animals, we calculated all possible pairwise correlation coefficients between heart and liver (i.e. 25) and compared all possible combinations of five coefficients between RNA-seq and RPF-seq (paired t-test on Fisher z-transformed coefficients).

Measurement error: measurement errors (Additional file 1: Figure S6A, B, F, G) were calculated similarly to [18] using the meas.est() function from smatr R package [40]. Genes were first binned according to average expression level (calculated as the fourth root of the product of liver RNA-seq, liver RPF-seq, kidney RNA-seq and kidney RPF-seq) into ten groups, each containing 10% of all genes. Within each bin, the measurement error was calculated separately for RNA-seq and RPF-seq and for liver and kidney, using the two replicates (log of normalised CDS counts) to estimate the error and the 12 timepoint samples to estimate its variability. For the analyses using a filtered gene set (Additional file 1: Figure S6F-G), genes that showed a mean expression ratio (either between organs or between RNA-seq and RPF-seq) greater than 2 for all timepoints were excluded (9236 genes used in analysis).

Analyses of differential translation efficiency

To test for differential translation efficiency (TE) between liver and kidney we used the Wilcoxon-signed rank paired test, using all 24 samples (12 timepoints; two replicates/timepoint) as replicates; resulting p values were FDR-corrected. A gene was defined as having differential TE when FDR < 0.01 and the inter-organ difference in TE was at least 1.5-fold (Fig. 2e).

Analysis of transcript usage diversity across organs: for each gene g, P(g) = (p 1,…,p n) is the vector of the relative expression proportions of its n protein-coding transcripts, as estimated from our RNA-seq analysis (see 'Sequencing data processing, alignment and quantification'). To quantify the dissimilarity in relative transcript isoform expression between liver L and kidney K, the Hellinger distance H is defined as:

$$H(P_L(g), P_K(g)) = 1/\sqrt{2}\sqrt{\sum_{i=1}^{n}\left(\sqrt{p_{Li}}-\sqrt{p_{Ki}}\right)^2} \quad (1)$$

In order to detect the transcript features that were associated with tissue specificity in TE, we selected genes whose transcript diversity between both organs originated from or was excluded from 5′ UTR, CDS, or 3′ UTR, based on feature annotation information for the detected protein-coding transcripts (Fig. 2g and Additional file 1: Figure S9).

Study of transcript characteristics: for single-isoform genes, we investigated whether a particular transcript characteristic (length, GC content, Kozak context, structure) could be predictive of differential TE. Length and GC content were determined directly on the whole transcript and/or on the region of interest (5′ UTR, CDS, 3′ UTR). Kozak context was scored according to the consensus sequence GccA/Gcc**AUG**G, where upper-case letters denote highly conserved bases (scored +3), lower-case letters indicate the most common nucleotides (scored +1) and bold is the start codon (not scored), giving a maximum score of 13. Minimum free energy secondary structures on the 5′ UTR were predicted with RNAfold from ViennaRNA package with default parameters [41].

Detection and translation efficiency calculation for uORFs

To assess the impact of differential uORF usage on TE differences across organs, uORFs were identified as in our previous study [10]. Briefly, genes expressing a single protein-coding isoform in both organs were used for this analysis ($n = 5815$). We selected uORFs with an AUG start codon and a length of at least 18 nt to the first in-frame stop codon and considered them as translated if the reads showed significant frame bias towards the reading frame of the uORF start codon and if coverage was > 10%. uORF translation efficiency was calculated from the ratio of RPF-seq to RNA-seq reads whose predicted A-sites mapped to the annotated uORF regions. If several uORFs partially of completely overlapped on a given 5′ UTR, a composite uORF was considered for read counting. uORFs overlapping with the CDS in the same frame were not considered. When they overlapped in different frames, only reads mapping to the 5′ UTR-specific uORF sequence (but not the overlapping sequence) was considered for quantifications.

Rhythmicity analyses

Rhythmicity detection and rhythmic parameter estimations in each dataset (RNA-seq and RPF-seq, liver and kidney) were done based on Akaike information criterion (AIC) model selection as in our previous study [10]. The Babel computational framework [22] was used to detect rhythmically translated genes from constantly expressed mRNAs within each organ. For cross-correlation of time series to compare the daily profiles of rhythmic genes beyond their peak differences, we used the ccf function in R. As we computed the correlations of the RPF-seq with respect to the RNA-seq profiles, negative lag values correspond to RPF leading RNA.

Hierarchical clustering of rhythmic genes

To evaluate the similarity of the expression profiles for rhythmic genes, a dissimilarity matrix was computed for each gene of interest, based on the Euclidean distance between the RNA-seq and RPF-seq expression profiles within and across organs. A hierarchical clustering tree was constructed on the weighted average of the dissimilarity matrices of genes under consideration (core clock genes in Fig. 5c or all rhythmic genes in Fig. 5d), using the 'average' clustering method. The R functions {packages} dist {stats}, fuse {analogue} and hclust {stats} were used for computing the individual dissimilarity matrices, the weighted mean dissimilarity matrix and the hierarchical clustering, respectively.

Additional files

Additional file 1: Supplementary Figures. This file contains the Supplementary **Figures S1–S14** and Supplementary Figure legends. (PDF 32661 kb)

Additional file 2: Mapping outcome summary. This file contains information on the deep-sequencing data from kidney (raw read counts, mapping summary etc.) (XLSX 23 kb)

Additional file 3: Differential TE analysis. This file contains details of the GO-term analysis on the differential TE gene set of Fig. 2. (XLSX 228 kb)

Additional file 4: Rhythmicity parameters in kidney datasets. This file contains the outcome of the transcriptome-wide rhythmicity analyses on the kidney datasets (related to Fig. 3a). (XLSX 163 kb)

Additional file 5: Rhythmicity parameters of 178 common rhythmic genes. This file contains the outcome of the rhythmicity analyses in kidney and liver for the 178 commonly rhythmic genes (RNA and RPF in kidney and liver; related to Fig. 3c). (XLSX 35 kb)

Additional file 6: Transcriptome-wide kidney RPF (blue) and RNA (orange) levels in the left panels (with "error bars" connecting the two replicates of each timepoint) and TE in the right panels. (ZIP 116896 kb)

Additional file 7: Expression plots for kidney and liver for the 178 common rhythmic genes of Fig. 3c. (ZIP 3338.28 kb)

Acknowledgments

We thank Lausanne Genomic Technologies Facility and Vital-IT staff for high-throughput sequencing and computational support.

Funding

DG acknowledges support by: Swiss National Science Foundation (grants 128399, 157528); National Centre of Competence in Research (NCCR) RNA & Disease; Fondation Pierre Mercier; Fondation Leenaards; Olga Mayenfisch Stiftung; SystemsX.ch StoNets consortium; University of Lausanne. PJ was supported by Human Frontiers Science Program long-term fellowship LT000158/2013-L.

Availability of data and materials

The sequencing data are available at NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE81283.
All the source code used in this study is released under a GNU General Public License and available at Github at https://github.com/gatfieldlab/cross-organ_riboprof and Zenodo repository at http://doi.org/10.5281/zenodo.521199 [42].
Time-resolved gene expression plots can be found under https://doi.org/10.6084/m9.figshare.4903193. A ZIP-file entitled Additional_File_6.zip (104.25 MB) contains transcriptome-wide kidney RPF (blue) and RNA (orange) levels in the left panels (with "error bars" connecting the two replicates of each timepoint) and TE in the right panels; a ZIP-file entitled Additional_File_7.zip (2.97 MB) contains expression plots for kidney and liver for the 178 common rhythmic genes of Figure 3C.

Authors' contributions

VCS and PJ performed the experiments. VCS and ABA analysed the data. DG and VCS wrote the manuscript with inputs from the other authors. DG designed the project. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

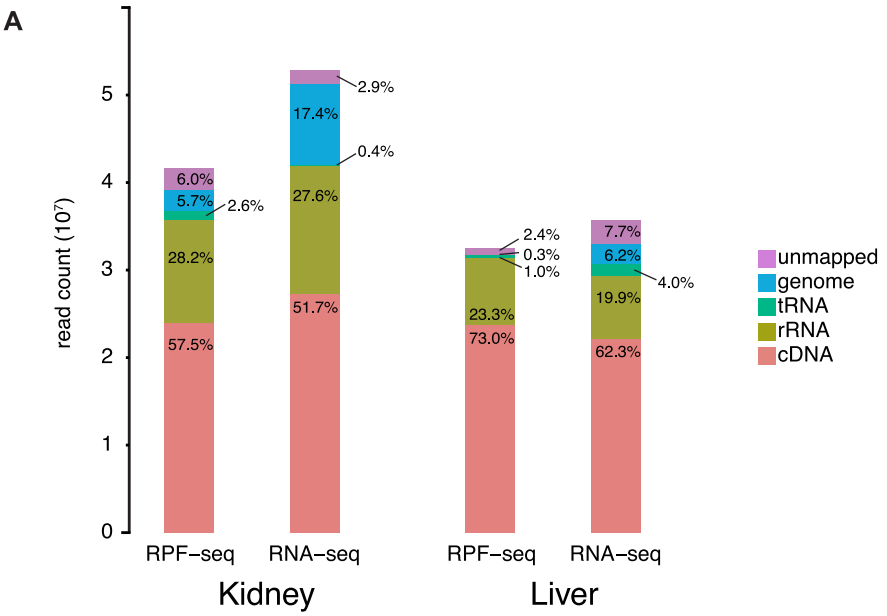
Ethics approval
Animal experimental procedures were approved by the Veterinary Office of the Canton Vaud (authorisation VD2376).

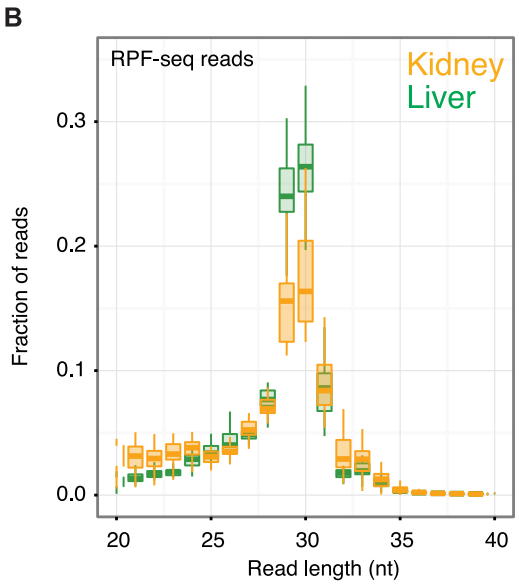
Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

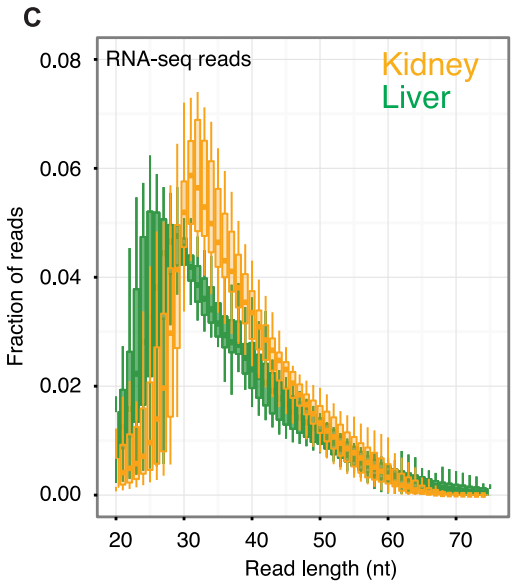
Author details
[1]Center for Integrative Genomics, University of Lausanne, Génopode, 1015 Lausanne, Switzerland. [2]Vital-IT, Swiss Institute of Bioinformatics, Génopode, 1015 Lausanne, Switzerland.

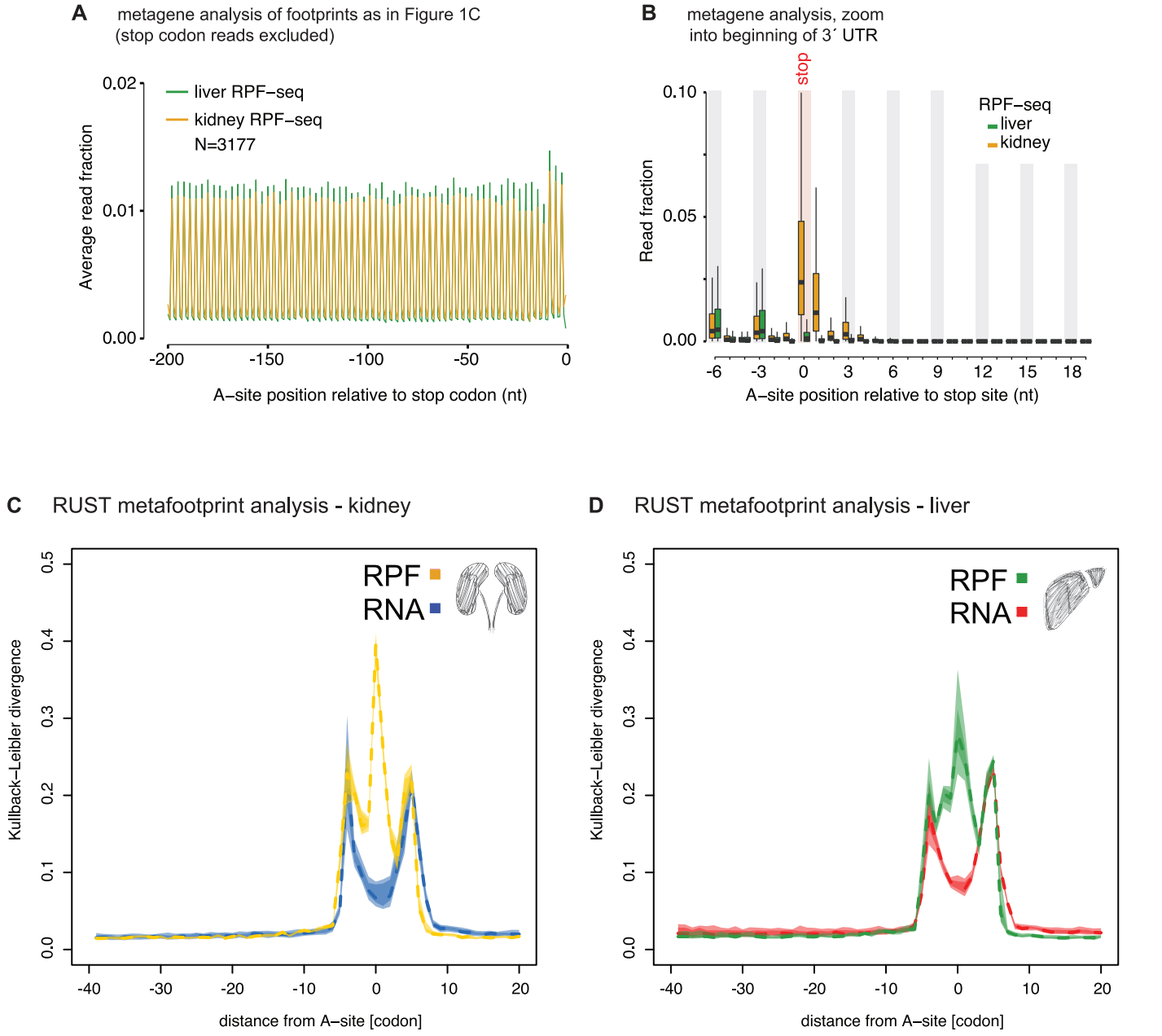
Received: 7 February 2017 Accepted: 26 April 2017
Published online: 16 June 2017

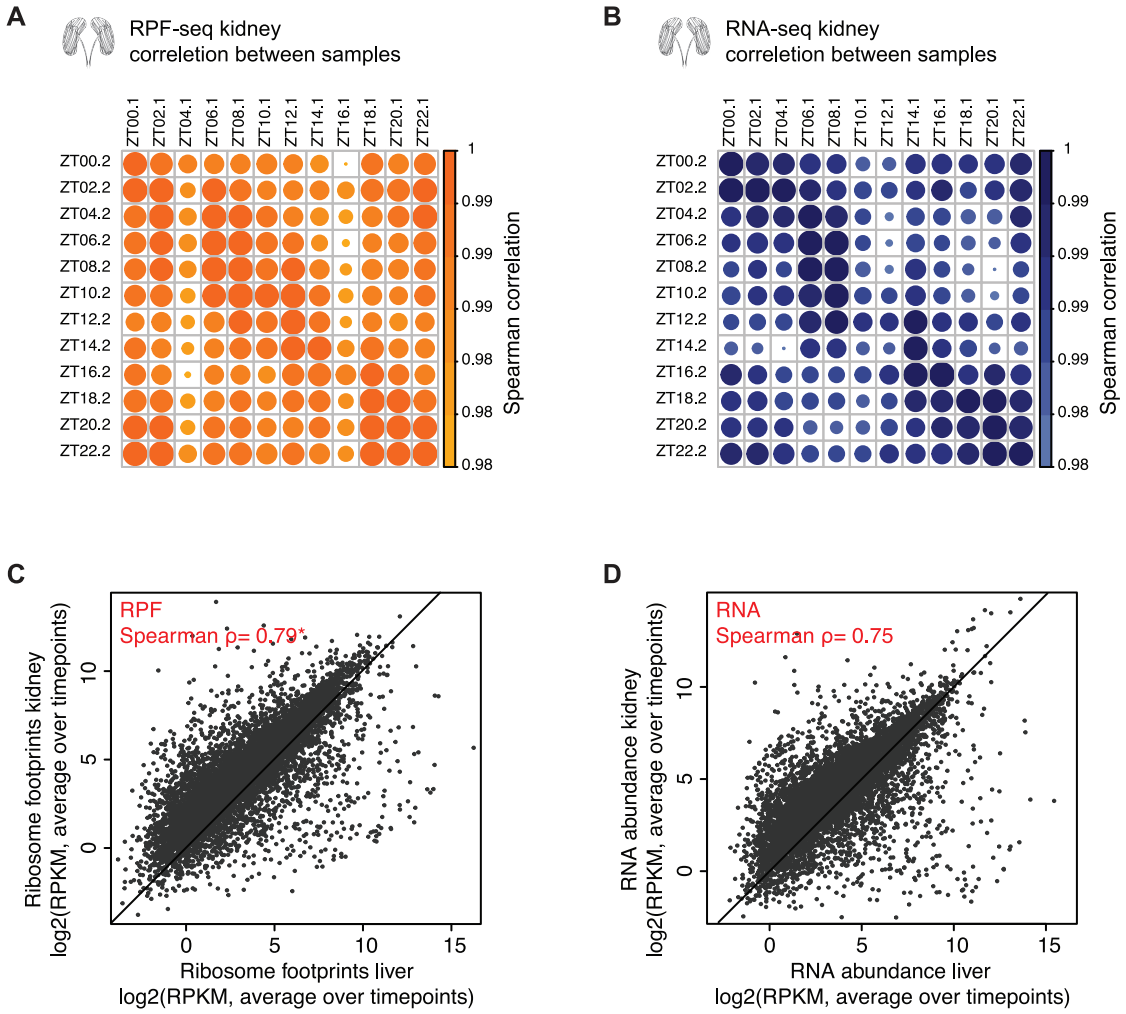
References
1. Dibner C, Schibler U, Albrecht U. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. Annu Rev Physiol. 2010;72:517–49.
2. Partch CL, Green CB, Takahashi JS. Molecular architecture of the mammalian circadian clock. Trends Cell Biol. 2014;24:90–9.
3. Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: Implications for biology and medicine. Proc Natl Acad Sci U S A. 2014;111:16219–24.
4. Meireles-Filho AC, Bardet AF, Yanez-Cuna JO, Stampfel G, Stark A. cis-regulatory requirements for tissue-specific programs of the circadian clock. Curr Biol. 2014;24:1–10.
5. Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, et al. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A. 2004;101:5339–46.
6. Luck S, Westermark PO. Circadian mRNA expression: insights from modeling and transcriptomics. Cell Mol Life Sci. 2016;73:497–521.
7. Reddy AB, Karp NA, Maywood ES, Sage EA, Deery M, O'Neill JS, et al. Circadian orchestration of the hepatic proteome. Curr Biol. 2006;16:1107–15.
8. Mauvoisin D, Wang J, Jouffe C, Martin E, Atger F, Waridel P, et al. Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. Proc Natl Acad Sci U S A. 2014;111:167–72.
9. Robles MS, Cox J, Mann M. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. PLoS Genet. 2014;10:e1004047.
10. Janich P, Arpat AB, Castelo-Szekely V, Lopes M, Gatfield D. Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res. 2015;25:1848–59.
11. Atger F, Gobet C, Marquis J, Martin E, Wang J, Weger B, et al. Circadian and feeding rhythms differentially affect rhythmic mRNA transcription and translation in mouse liver. Proc Natl Acad Sci U S A. 2015;112:E6579–6588.
12. Jouffe C, Cretenet G, Symul L, Martin E, Atger F, Naef F, et al. The circadian clock coordinates ribosome biogenesis. PLoS Biol. 2013;11:e1001455.
13. Bonny O, Vinciguerra M, Gumz ML, Mazzoccoli G. Molecular bases of circadian rhythmicity in renal physiology and pathology. Nephrol Dial Transplant. 2013;28:2421–31.
14. Ingolia NT. Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet. 2014;15:205–13.
15. Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011;478:343–8.
16. Janich P, Arpat AB, Castelo-Szekely V, Gatfield D. Analyzing the temporal regulation of translation efficiency in mouse liver. Genom Data. 2016;8:41–4.
17. O'Connor PB, Andreev DE, Baranov PV. Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. Nat Commun. 2016;7:12915.
18. Albert FW, Muzzey D, Weissman JS, Kruglyak L. Genetic influences on translation in yeast. PLoS Genet. 2014;10:e1004692.
19. Schafer S, Adami E, Heinig M, Rodrigues KE, Kreuchwig F, Silhavy J, et al. Translational regulation shapes the molecular landscape of complex disease phenotypes. Nat Commun. 2015;6:7200.
20. Gonzalez-Porta M, Calvo M, Sammeth M, Guigo R. Estimation of alternative splicing variability in human populations. Genome Res. 2012;22:528–38.
21. Kojima S, Sher-Chen EL, Green CB. Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. Genes Dev. 2012;26:2724–36.
22. Olshen AB, Hsieh AC, Stumpf CR, Olshen RA, Ruggero D, Taylor BS. Assessing gene-level translational control from ribosome profiling. Bioinformatics. 2013;29:2995–3002.
23. Lee Y, Chen R, Lee HM, Lee C. Stoichiometric relationship among clock proteins determines robustness of circadian rhythms. J Biol Chem. 2011;286:7033–42.
24. Yagita K, Horie K, Koinuma S, Nakamura W, Yamanaka I, Urasaki A, et al. Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. Proc Natl Acad Sci U S A. 2010;107:3846–51.
25. Landgraf D, Wang LL, Diemer T, Welsh DK. NPAS2 compensates for loss of CLOCK in peripheral circadian oscillators. PLoS Genet. 2016;12:e1005882.
26. Fang B, Lazar MA. Dissecting the Rev-erbalpha cistrome and the mechanisms controlling circadian transcription in liver. Cold Spring Harb Symp Quant Biol. 2015;80:233–8.
27. D'Alessandro M, Beesley S, Kim JK, Chen R, Abich E, Cheng W, et al. A tunable artificial circadian clock in clock-defective mice. Nat Commun. 2015;6:8587.
28. Gu X, Xing L, Shi G, Liu Z, Wang X, Qu Z, et al. The circadian mutation PER2(S662G) is linked to cell cycle progression and tumorigenesis. Cell Death Differ. 2012;19:397–405.
29. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009;324:218–23.
30. McManus CJ, May GE, Spealman P, Shteyman A. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. 2014;24:422–30.
31. Khan Z, Ford MJ, Cusanovich DA, Mitrano A, Pritchard JK, Gilad Y. Primate transcript and protein expression levels evolve under compensatory selection pressures. Science. 2013;342:1100–4.
32. Schrimpf SP, Weiss M, Reiter L, Ahrens CH, Jovanovic M, Malmstrom J, et al. Comparative functional analysis of the Caenorhabditis elegans and Drosophila melanogaster proteomes. PLoS Biol. 2009;7:e48.
33. Jang C, Lahens NF, Hogenesch JB, Sehgal A. Ribosome profiling reveals an important role for translational control in circadian gene expression. Genome Res. 2015;25:1836–47.
34. Dunlap JC. Genetics and molecular analysis of circadian rhythms. Annu Rev Genet. 1996;30:579–601.
35. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17:10–2.
36. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
37. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–11.
38. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
39. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25.
40. Warton DI, Wright IJ, Falster DS, Westoby M. Bivariate line-fitting methods for allometry. Biol Rev Camb Philos Soc. 2006;81:259–91.
41. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;6:26.
42. Castelo-Szekely V. gatfieldlab/cross-organ_riboprof cross-organ_riboprof_v1. Zenodo 2017.


A
read count (10^7)
0
1
2
3
4
5
6.0%
5.7%
2.6%
28.2%
57.5%
2.9%
17.4%
0.4%
27.6%
51.7%
2.4%
0.3%
1.0%
23.3%
73.0%
7.7%
6.2%
4.0%
19.9%
62.3%
RPF−seq
RNA−seq
RPF−seq
RNA−seq
Kidney
Liver
unmapped
genome
tRNA
rRNA
cDNA


B
RPF-seq reads
Kidney
Liver
Fraction of reads
0.0
0.1
0.2
0.3
20
25
30
35
40
Read length (nt)

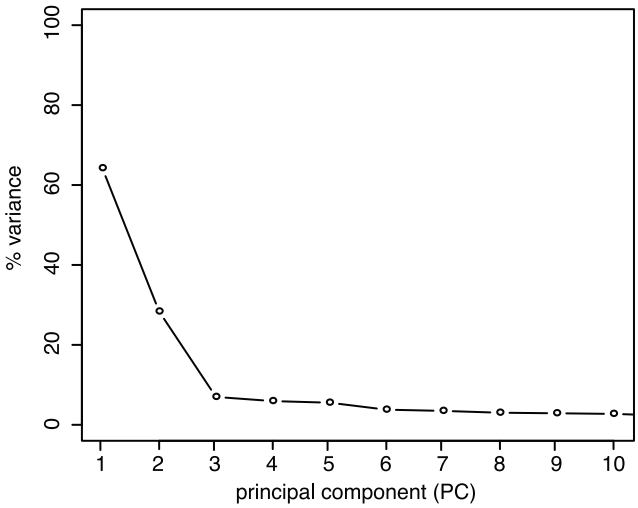

C
RNA-seq reads
Kidney
Liver
Fraction of reads
0.00
0.02
0.04
0.06
0.08
20
30
40
50
60
70
Read length (nt)

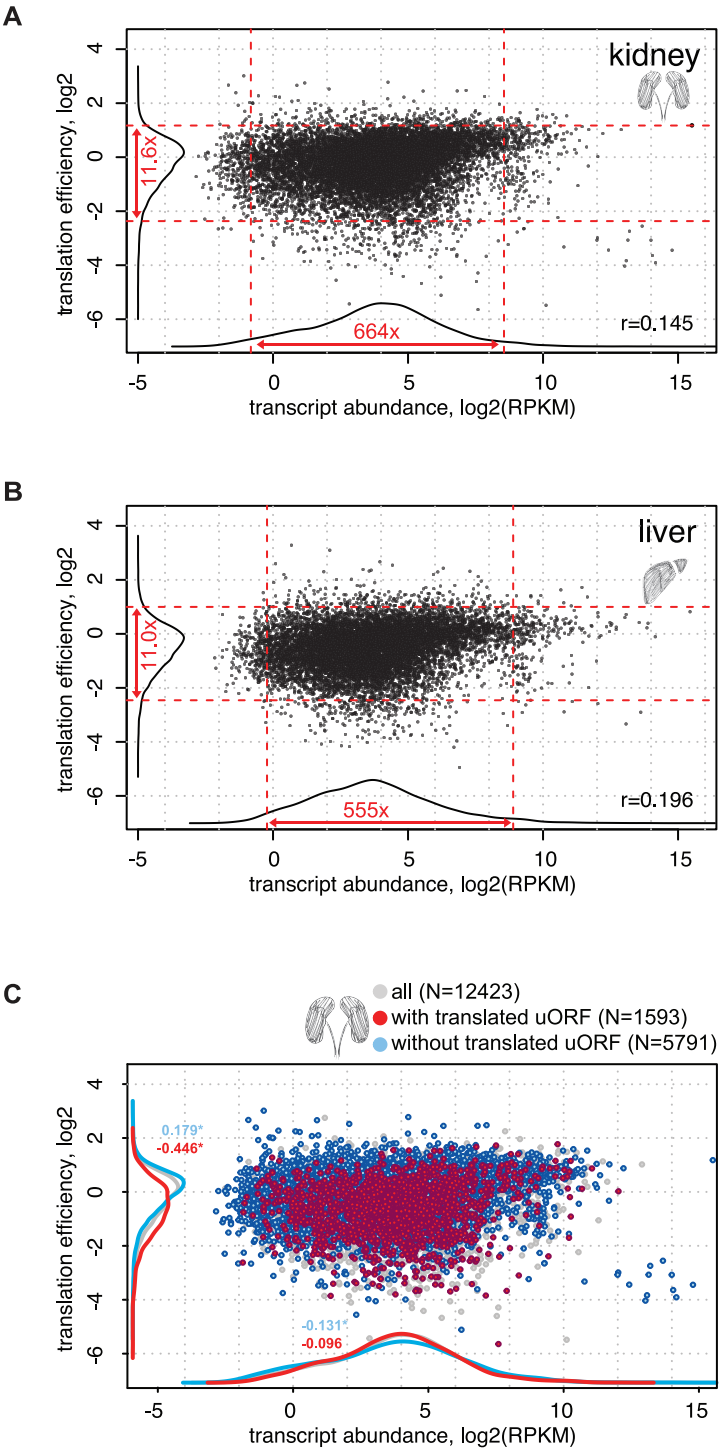

A
metagene analysis of footprints as in Figure 1C
(stop codon reads excluded)
liver RPF−seq
kidney RPF−seq
N=3177
Average read fraction
0.02
0.01
0.00
-200
-150
-100
-50
0
A−site position relative to stop codon (nt)
B
metagene analysis, zoom
into beginning of 3´ UTR
stop
RPF−seq
liver
kidney
Read fraction
0.10
0.05
0.00
-6
-3
0
3
6
9
12
15
18
A−site position relative to stop site (nt)
C
RUST metafootprint analysis - kidney
RPF
RNA
Kullback−Leibler divergence
0.5
0.4
0.3
0.2
0.1
0.0
-40
-30
-20
-10
0
10
20
distance from A−site [codon]
D
RUST metafootprint analysis - liver
RPF
RNA
Kullback−Leibler divergence
0.5
0.4
0.3
0.2
0.1
0.0
-40
-30
-20
-10
0
10
20
distance from A−site [codon]



* correlations are significantly different
Steiger test for difference between 2 dependent correlations, Z value=-22.23, p=8.7e-110 (n=10289)

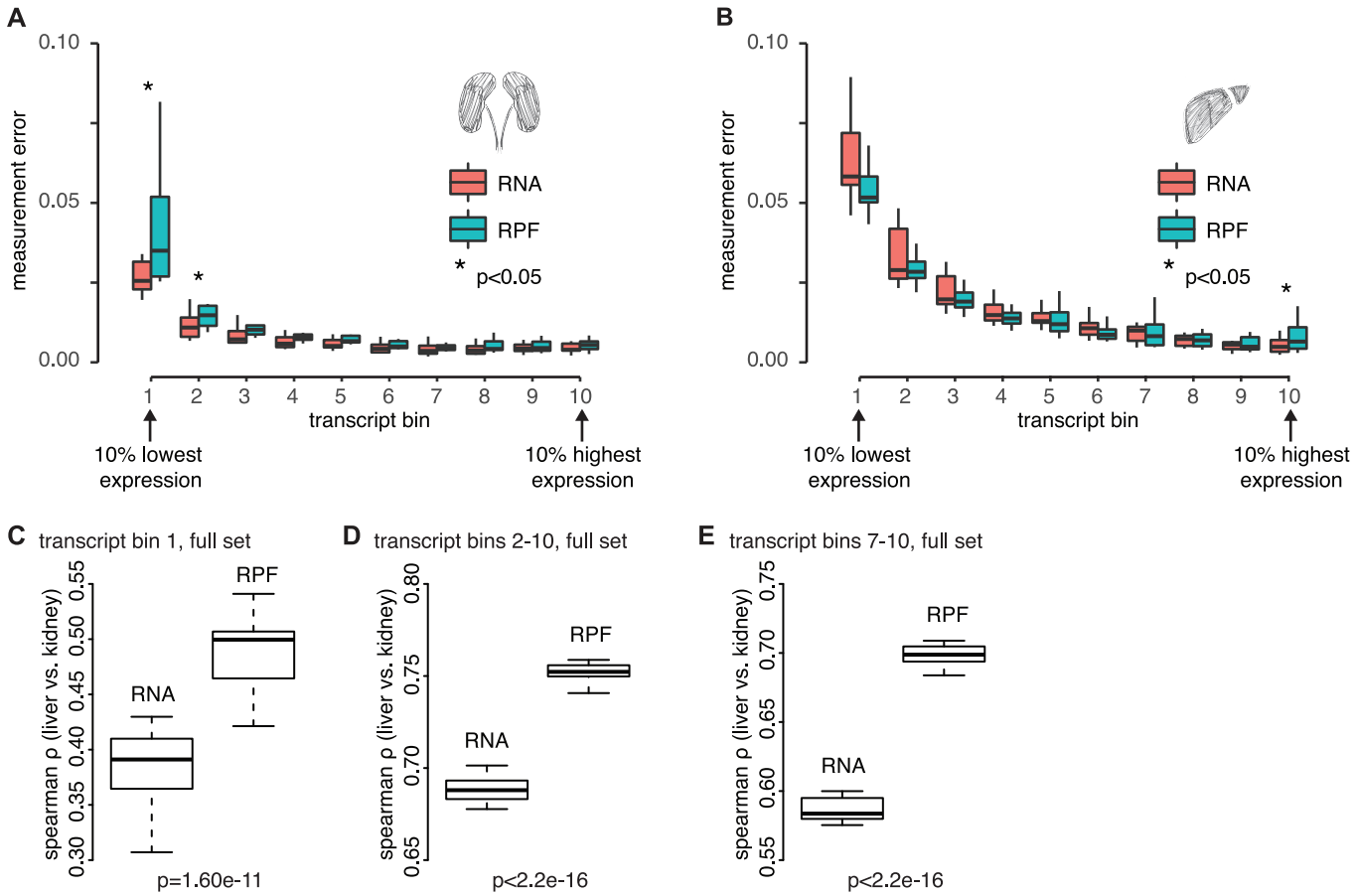
Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological Bulletin, 87, 245-251.


% variance
0
20
40
60
80
100
1
2
3
4
5
6
7
8
9
10
principal component (PC)

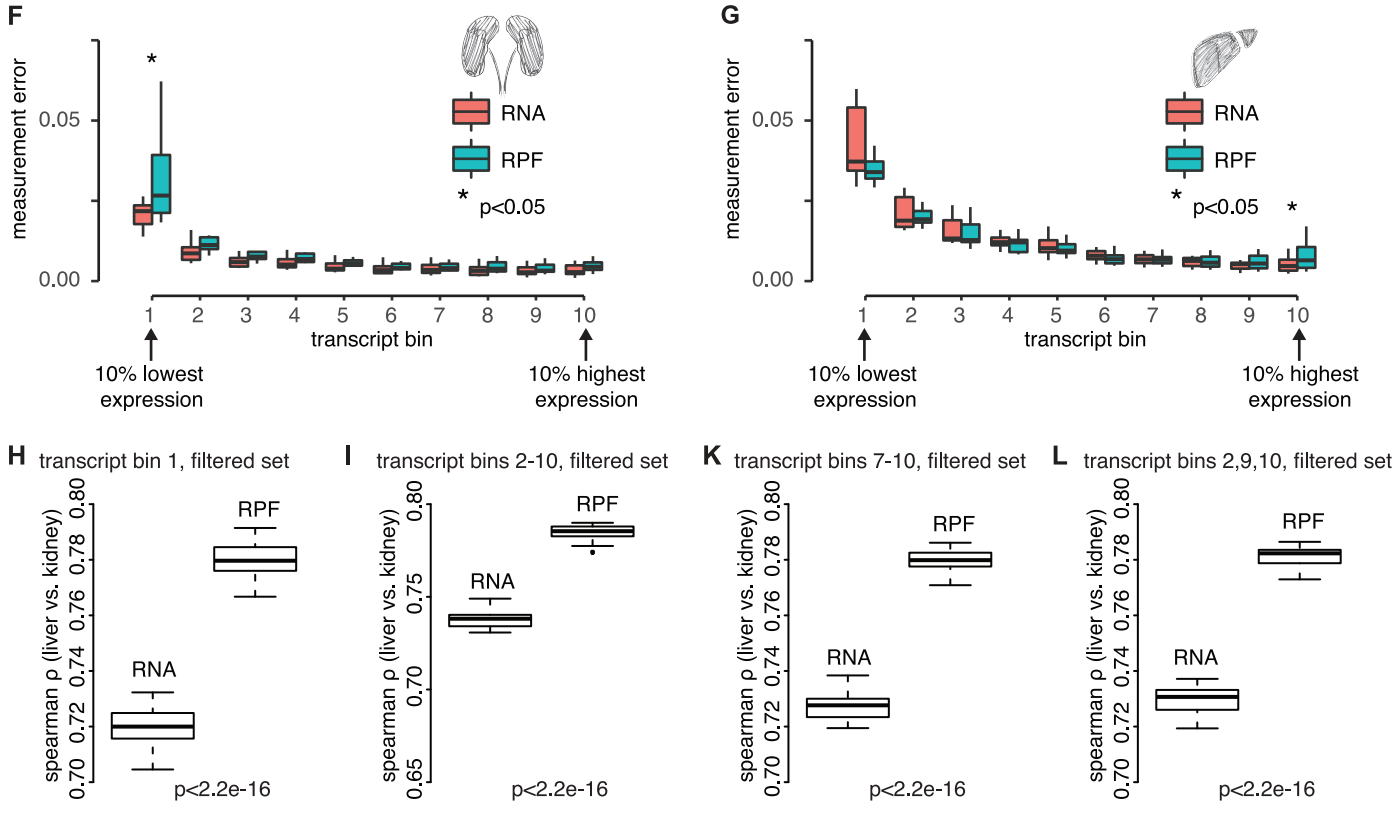

A
kidney
11.6x
664x
r=0.145
translation efficiency, log2
transcript abundance, log2(RPKM)
B
liver
11.0x
555x
r=0.196
translation efficiency, log2
transcript abundance, log2(RPKM)
C
all (N=12423)
with translated uORF (N=1593)
without translated uORF (N=5791)
0.179*
-0.446*
-0.131*
-0.096
translation efficiency, log2
transcript abundance, log2(RPKM)

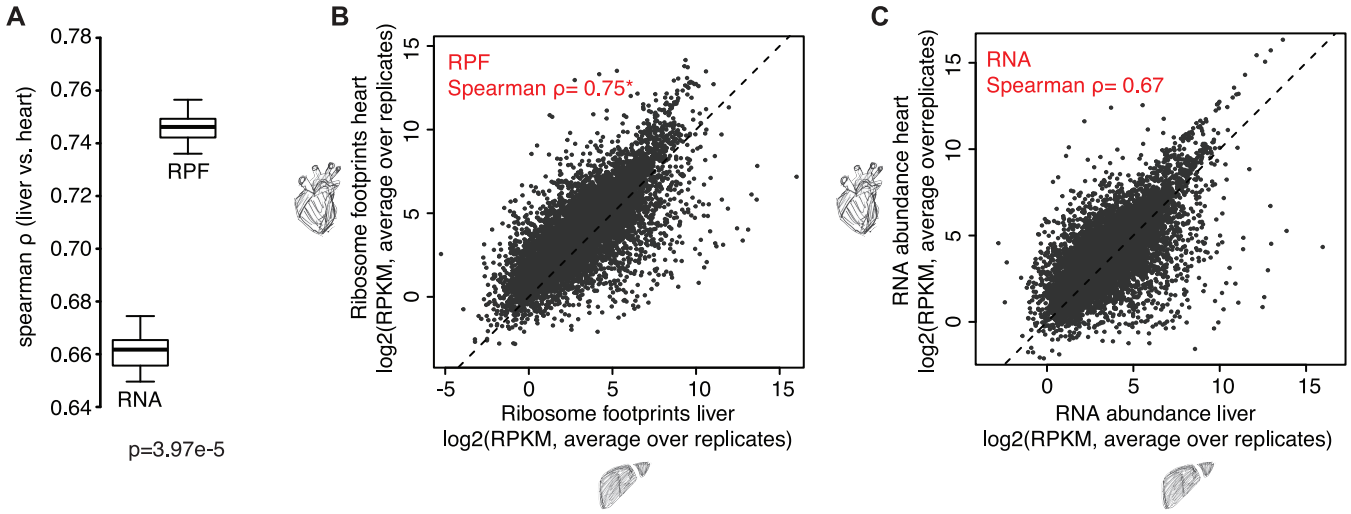
Measurement errors (MEs), genes binned by expression level

based on all genes (full set, N=10289):



after removal of genes that were most variable in their expression (filtered set, N=9236):





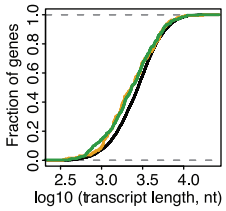
* correlations are significantly different
Steiger test for difference between 2 dependent correlations, Z value=-30.25, p=2.3e-201 (n=9325)

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. Psychological Bulletin, 87, 245-251.

A Cumulative distribution plots

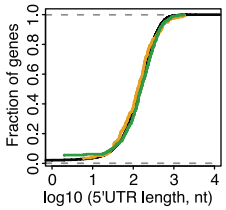
Transcript Length

All (N=5278)
kid.diff.TE (N=193) p=7.4e−4
liv.diff.TE (N=340) p=4.1e−4



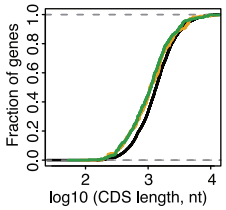
5' UTR Length

All (N=5278)
kid.diff.TE (N=193) p=0.5169
liv.diff.TE (N=340) p=0.1045



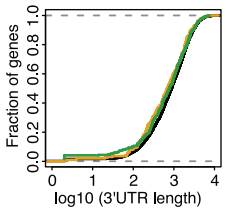
CDS Length

All (N=5278)
kid.diff.TE (N=193) p=0.0021
liv.diff.TE (N=340) p=2.41e−5



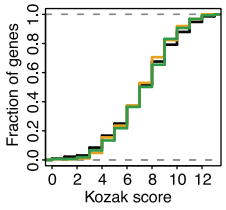
3' UTR Length

All (N=5278)
kid.diff.TE (N=193) p=0.073
liv.diff.TE (N=340) p=0.15



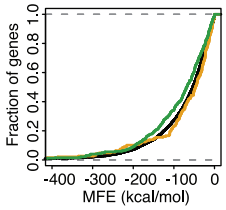
Kozak Context Score

All (N=5278)
kid.diff.TE (N=193) p=0.93
liv.diff.TE (N=340) p=0.72



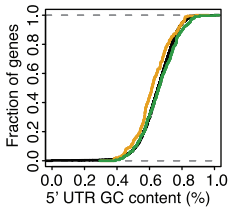
5' UTR Minimum Folding Energy (MFE)

All (N=5278)
kid.diff.TE (N=193) p=0.199
liv.diff.TE (N=340) p=0.00517

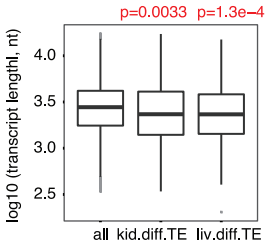


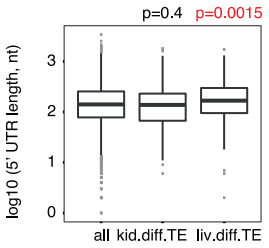
5' UTR GC Content

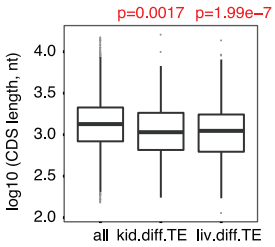
All (N=5278)
kid.diff.TE (N=193) p=3.12e−4
liv.diff.TE (N=340) p=0.138

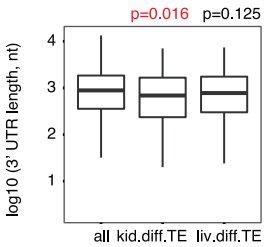


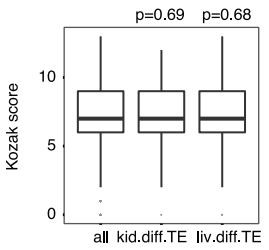
B Box plots

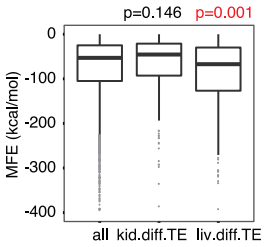


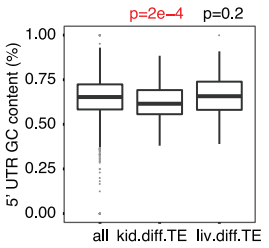




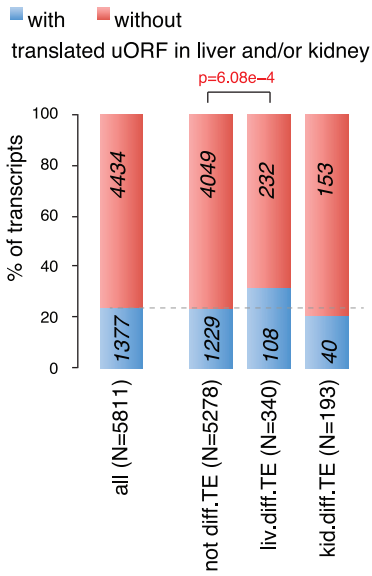








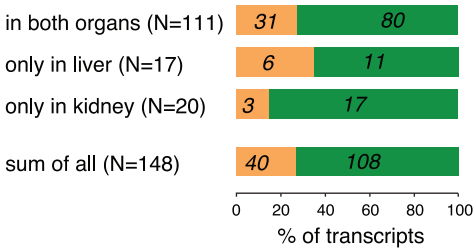
C all uORFs - distribution



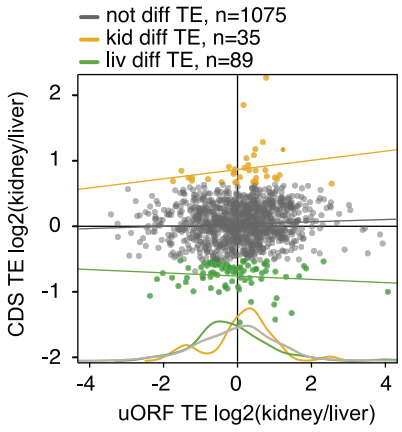
D differential TE transcripts with uORF translation
uORF usage according to organ

kid.diff.TE (higher TE in kidney)
liv.diff.TE (higher TE in liver)

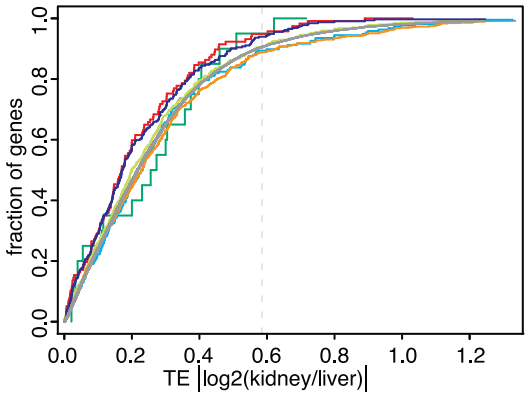
uORFs translation detected:



E uORF TE - CDS TE correlation



A

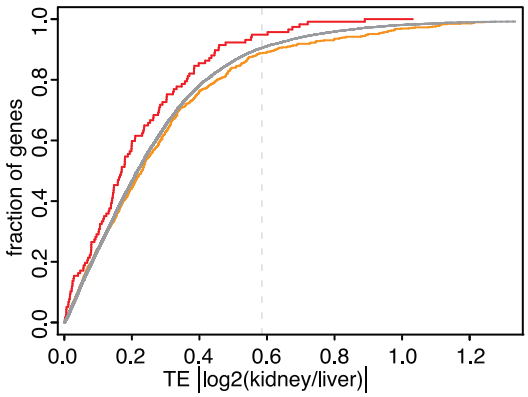


genes with transcript diversity
only in 5′ UTR (N=216), not in 5′ UTR (N=314)
only in CDS (N=117), not in CDS (N=774)
only in 3′ UTR (N=20), not in 3′ UTR (N=497)

all genes (N=10289)

Kolmogorov-Smirnov test
not in 5' UTR vs all: p=5.5e−4
only in CDS vs all: p=0.036
not in 5' UTR vs only in 5′ UTR: p=0.012
only in CDS vs not in CDS: p=0.015

B

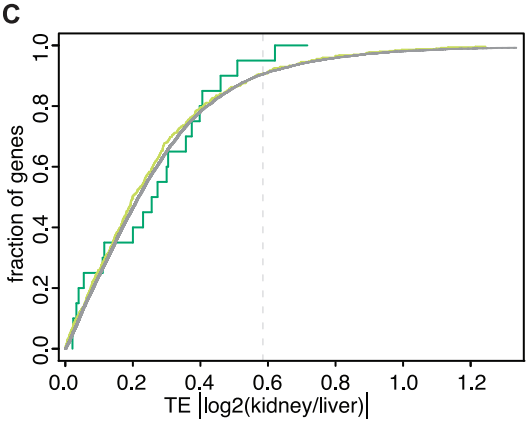


genes with transcript diversity
only in CDS (N=117), not in CDS (N=774)

all genes (N=10289)

Kolmogorov-Smirnov test
only in CDS vs not in CDS: p=0.015
only in CDS vs all: p=0.036

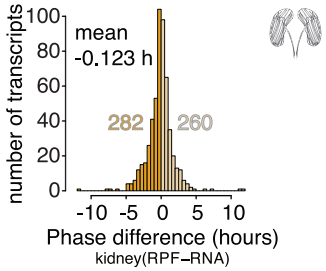
C



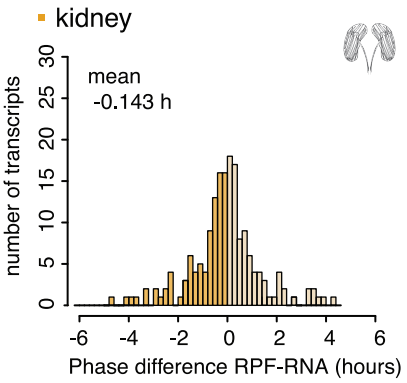
genes with transcript diversity
only in 3′ UTR (N=20), not in 3′ UTR (N=497)

all genes (N=10289)

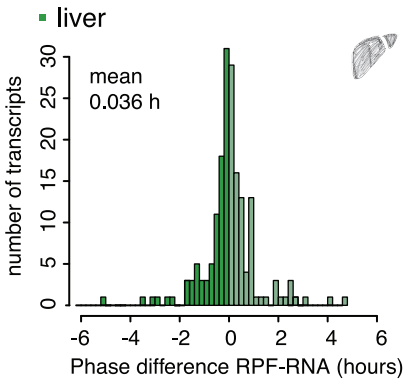
A Peak phase difference RPF-RNA of rhythmic genes in kidney (N=542)



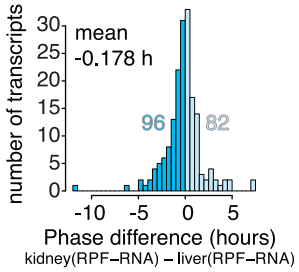
B Peak phase difference RPF-RNA of common rhythmic genes (N=178)

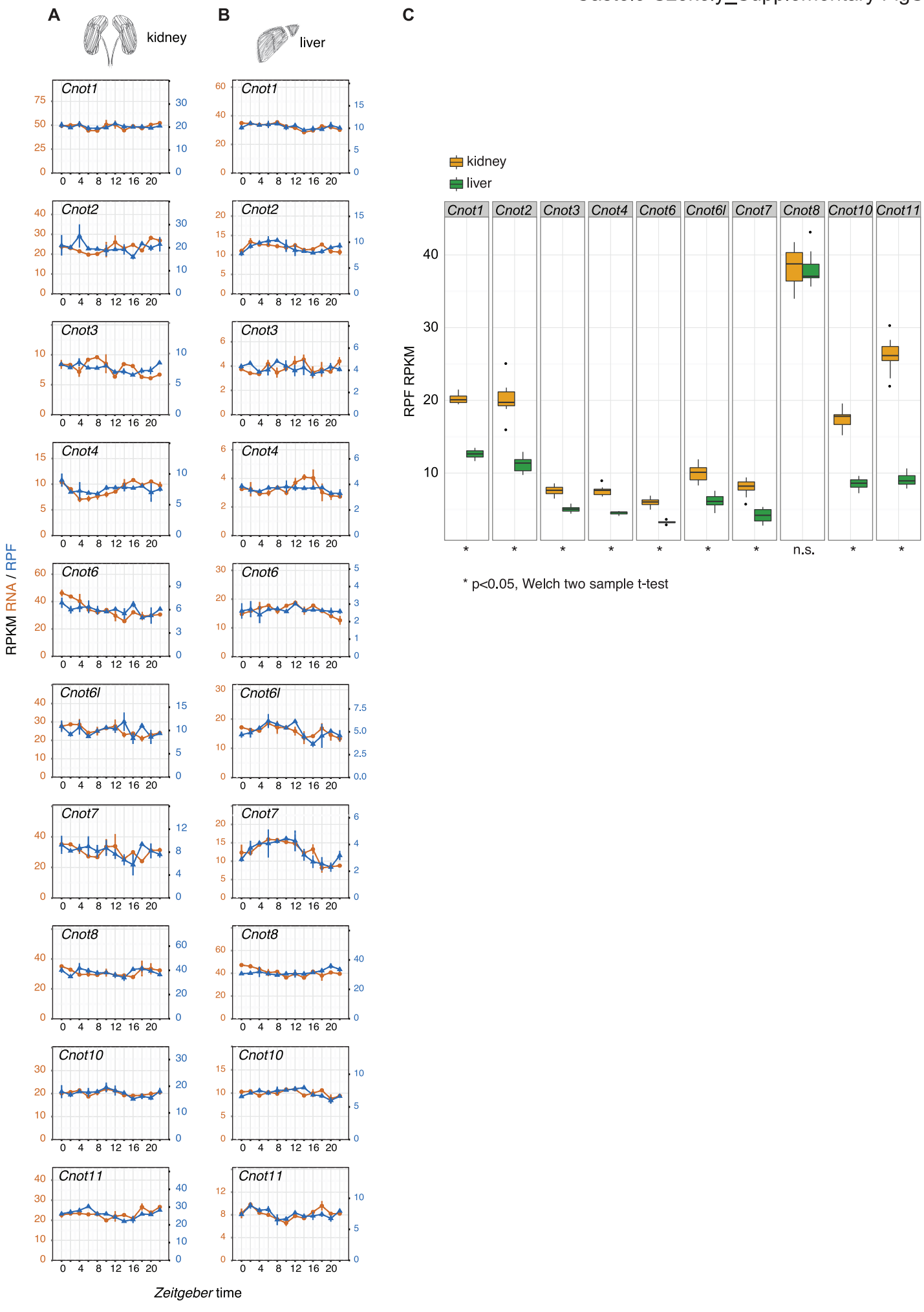


C Peak phase difference RPF-RNA of common rhythmic genes (N=178)

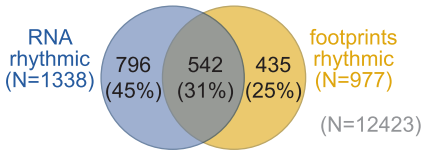


D Peak phase difference (RPF-RNA) in kidney relative to that in liver for the common rhythmic gene set (N=178)

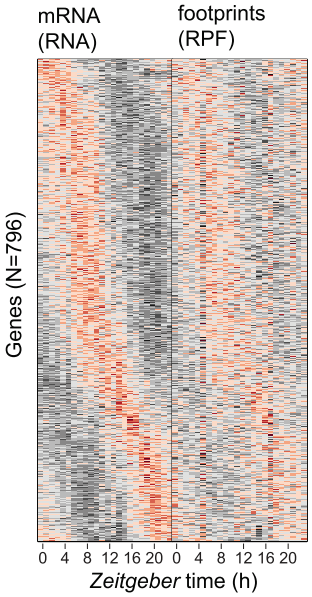



A
kidney
B
liver
C
Cnot1
Cnot2
Cnot3
Cnot4
Cnot6
Cnot6l
Cnot7
Cnot8
Cnot10
Cnot11
RPKM RNA / RPF
Zeitgeber time
kidney
liver
RPF RPKM
n.s.
* p<0.05, Welch two sample t-test

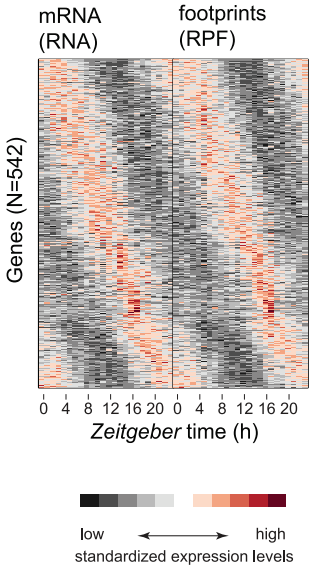
A rhythmic in kidney



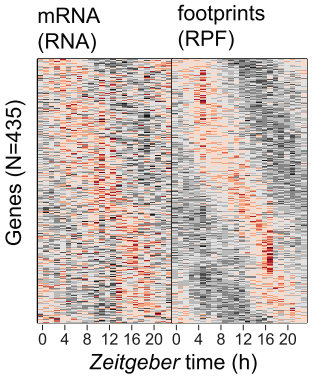
B "RNA only" rhythmic (796 genes)

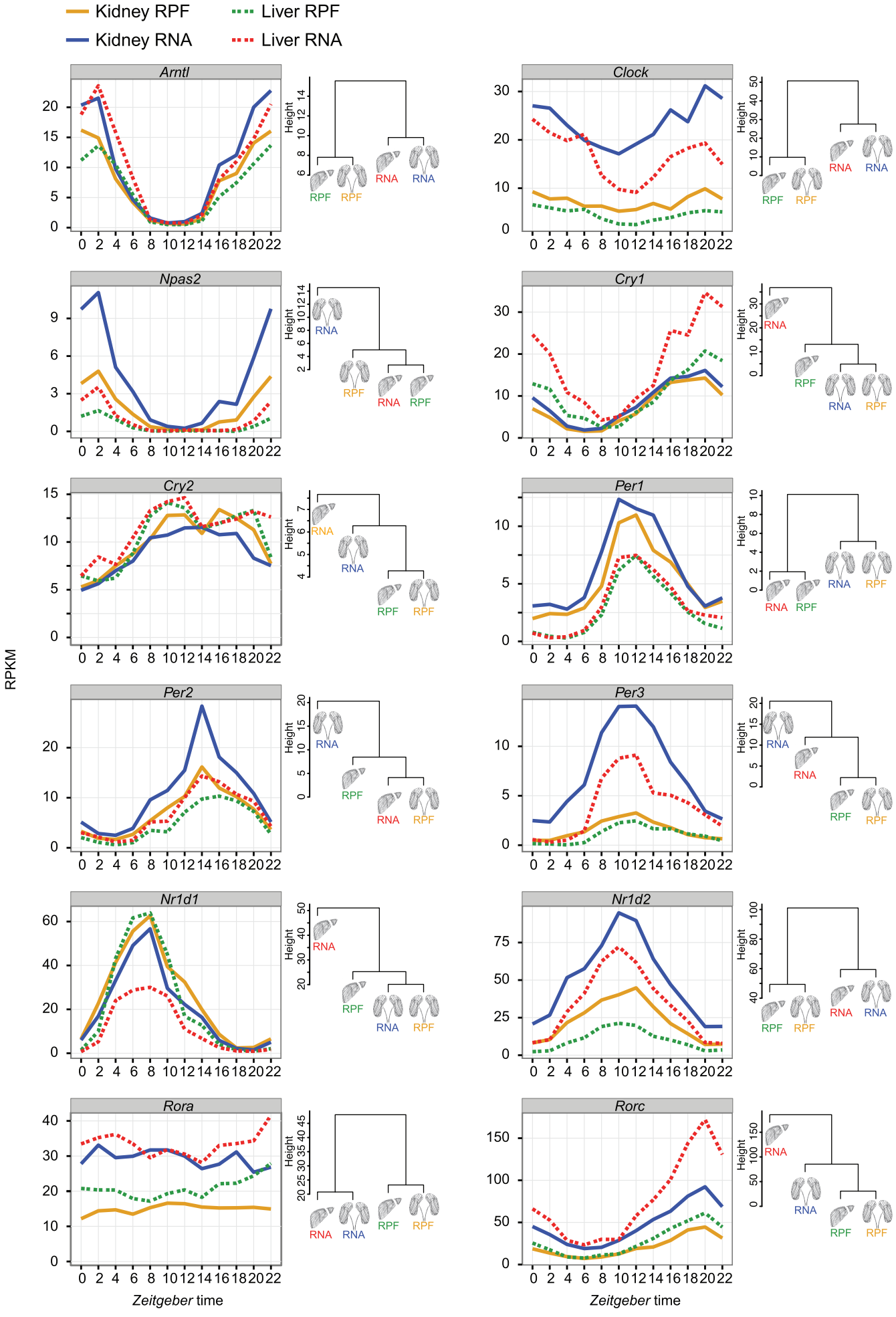


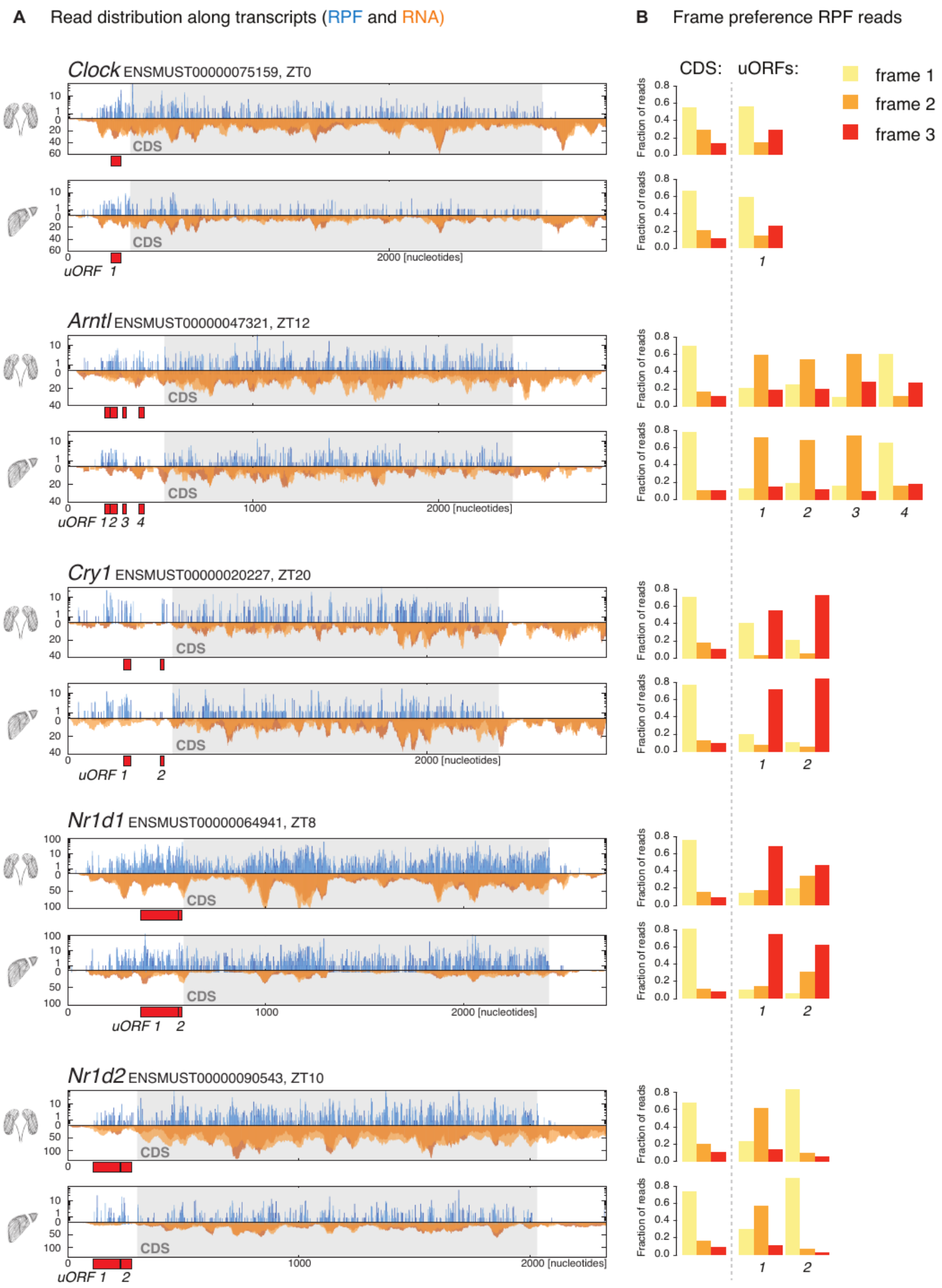
C RNA and RPF rhythmic (542 genes)



D "footprints only" rhythmic (435 genes)




Kidney RPF
Liver RPF
Kidney RNA
Liver RNA
Arntl
Clock
Npas2
Cry1
Cry2
Per1
Per2
Per3
Nr1d1
Nr1d2
Rora
Rorc
RPKM
Zeitgeber time
Zeitgeber time
Height
RNA
RPF


A
Read distribution along transcripts (RPF and RNA)
B
Frame preference RPF reads
CDS:
uORFs:
frame 1
frame 2
frame 3
Clock ENSMUST00000075159, ZT0
Arntl ENSMUST00000047321, ZT12
Cry1 ENSMUST00000020227, ZT20
Nr1d1 ENSMUST00000064941, ZT8
Nr1d2 ENSMUST00000090543, ZT10
CDS
uORF
Fraction of reads
2000 [nucleotides]

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Overview of sequencing outcome and read length distribution for RPF- and RNA-seq data used in the study.
A Summary of outcome of the sequential mapping pipeline, indicating the number (y-axis) and percentage (within bars) of reads mapping to each database, averaged over all timepoints. For each sample of the four datasets an average of more than 20 million reads mapped to the protein-coding transcriptome (cDNA) and was used for the study.
B and C RPF-seq (B) and RNA-seq (C) read length after trimming of adaptors showed that most RPF-seq reads had a length of 29-30 nucleotides in both organs, whereas RNA-seq fragments showed a broader distribution as expected from chemical RNA fragmentation. Boxplots represent the interquartile range and whiskers extend to the minimum and maximum values within 1.5 times the interquartile range.

Figure S2. RPF reads at the stop codon and quality control.
A A-site position of RPF-seq reads in the last 200 nt of the CDS, excluding stop codon reads. Read density along the CDS was similar in liver and kidney; thus the higher read density observed in Fig. 1C at the stop codon does not a ect CDS-based calculations.
B Zooming into the footprint read density at the end of the CDS, stop codon, and beginning of the 3′ UTR indicated read di erences between organs for the stop codon itself and up to 4 nt downstream, which were increased in kidney. Stop codon reads are counted towards the 3′ UTR and their higher level in kidney thus also explains why our analysis in Fig. 1B shows more 3′ UTR reads for this organ. The remainder of the 3′ UTR shows a similar depletion of reads in both organs.
C and D Ribo-seq Unit Step Transformation (RUST) metafootprint analyses to evaluate the contribution of local mRNA positions to the density of footprints. Light and dark coloured polygonal areas denote the 10%-90% and 25%-75% percentiles, respectively, and the dash-line denotes the median of the Kullback–Leibler divergence (K–L) pro les of all samples within kidney (C) and liver (D) separately for RPF and RNA reads (colour code in inset). The K-L pro le for each sample was calculated from the RUST ratio values of 61 sense codons across a moving window of 40 triplet codons upstream to 20 triplet codons downstream of predicted A-sites. High K-L divergence maxima (lowest relative entropy or highest information gain) are thus found in the vicinity of the A-site in RPF libraries, and the 5′ and 3′ termini of the reads in RNA libraries. Importantly, the pro les are similar for kidney and liver, indicating overall similar footprint quality in the two independent datasets.

Figure S3. High technical and biological reproducibility of datasets.

A and B Spearman correlation of normalised CDS read counts between timepoints and between replicates for kidney RPF-seq (A) and kidney RNA-seq (B) datasets. The correlation coe cient is indicated by the size and shading of the disks. Biological replicates thus show excellent correlation; moreover, the correlation coe cients of di erent timepoints re ect the rhythmic nature of the data.
C and D Normalised CDS read counts (RPKM) in liver vs. kidney at the RNA (C) and RPF level (D). In these graphs, the averages over all timepoints were compared for the set of commonly expressed genes (N=10289; see Fig. 2A). Note the overall higher Spearman correlation at the footprint level than transcript level. The di erence between correlations is highly signi cant with p=8.7e-110, Z value=-22.23; Steiger test for di erence between 2 dependent correlations (Reference: Steiger JH, 1980. Psychological Bulletin, 87, 245-251).

Figure S4. Additional information for Principal Component Analysis.
Scree plot showing the rst 10 components of the PCA in Fig. 1D-E. Components 1 and 2 explained most variance, followed by PC3 to PC5, which explained a closely similar proportion of variance in the data; the plateau was apparent from the sixth component.

Figure S5. Contribution of translation e ciency to overall gene expression variation within organs.
A and B Scatterplot of mRNA abundance vs. translation e ciency (TE) in kidney (A; N=12423 genes) and liver (B; N=10676 genes), averaged over all timepoints. Corresponding density lines are plotted on the margins. Dotted red lines represent the 2.5 and 97.5 percentiles, and the corresponding fold change is indicated. The transcript abundance range for 95% of genes thus spanned two orders of magnitude (>500-fold range in either organ), whereas TE dynamic range was less than 12-fold in either organ. Transcript abundance di erences can thus be considered the main source of gene expression variability in the tissues. Moreover, Pearson's r values of 0.145 (kidney) and 0.196 (liver) indicate weak positive correlation of transcript abundance and TE.
C Scatterplot of transcript abundance (TA) vs. translation e ciency (TE) in main CDS for kidney and as averages over all timepoints. Highlighted are single protein-coding genes that contain (red) or do not contain (blue) translated uORFs. Corresponding density lines are plotted on the margins. uORF translation is thus clearly associated with signi cantly reduced translation e ciency. Numbers on the density curves indicate the location shift relative to all transcripts. Genes with translated uORFs: TA, p=0.16; TE, p<2.2e-16 (Wilcoxon rank sum test). Genes without translated uORFs: TA, p=8.7e-5; TE, p<2.2e-16 (Wilcoxon rank sum test).

Figure S6. Translational compensation is independent of technical biases in the datasets.
A and B Measurement error (ME) of all genes in the dataset (n=10289) was

calculated separately for RNA-seq (red) and RPF-seq (turquoise) for kidney (A) and liver (B), and plotted as a function of increasing average expression levels. Brie y, measurement errors within each bin were calculated as in Albert et al., 2015 (Albert FW, Muzzey D, Weissman JS, Kruglyak L, 2014. PLoS Genetics, 10, e1004692) using the 12 timepoints as replicates for the measurement estimates (see Methods).
C to E Spearman correlation between liver and kidney for RNA and RPF data for (C) genes showing the highest measurement errors (bin 1 in A, B), (D) bins 2-10, and (E) genes that have higher measurement error in RPF-seq than in RNA-seq samples in both organs (bins 7-10). Each boxplot contains the correlation coe cients between organs for each timepoint and replicate sample. Together these analyses showed that RPF-seq samples have a higher degree of similarity across organs (indicated p-values are from paired t-tests on Fisher-transformed correlation coe cients), even when considering lowly expressed genes with higher associated experimental error (C), or when considering genes with higher RPF-seq than RNA-seq measurement errors (E). These results thus ruled out that a systematic lower measurement error in RPF-seq experiments could have been the underlying cause of the higher correlation in RPF-seq than RNA-seq observed in Fig. 2C.
F and G Same as A and B, but with a ltered gene set in which speci cally those genes that showed very di erent expression levels/high variability between organs or between datasets (RPF-seq, RNA-seq) were removed (see Methods). The reason to also analyse such a ltered set was that we wished to be sure that genes that were widely di erent in their gene expression level were not distorting the analyses (e.g. speci cally causing extreme measurement errors under a condition where expression was very low). Moreover, because the binning into the groups was based on expression level across all sets (calculated as the fourth root of the product of liver RNA-seq, liver RPF-seq, kidney RNA-seq and kidney RPF-seq), the highly variable genes made binning inaccurate. This ltered set thus contained genes with overall better comparability across datasets; of note, the distribution of ME di erences using the ltered set was very similar to the full set in A-B.
H to L Inter-organ Spearman correlation in RNA-seq and RPF-seq samples for various gene bins as indicated, using the ltered set. As in C-E, even when considering for example the genes with the highest overall measurement error (H), or the genes with higher RPF-seq than RNA-seq measurement errors in both organs (L), a signi cantly higher correlation is observed in RPF-seq samples (paired t-test on Fisher-transformed correlation coe cients).

Figure S7. Translational compensation detected in rat liver and heart.
A) Spearman correlation coe cient between rat heart and liver samples calculated from the data of Schafer et al., 2015 (Schafer S, Adami E, Heinig M, Rodrigues KE, Kreuchwig F, Silhavy J, van Heesch S, Simaite D, Rajewsky N, Cuppen E, Pravenec M, Vingron M, Cook SA, Hubner N, 2015. Nature Communications 8, 7200). Each boxplot contains the correlation coe cients of all possible pairwise comparisons

between heart and liver replicates (remark: in this study, organs in each of the ve replicates did not necessarily come from the same animals, thus precluding de ned pairwise comparisons between same animals). The indicated p-value is the 95th percentile of the ensemble of p-values resulting from all possible comparisons between RPF-seq and RNA-seq correlation coe cients (paired t-test of Fisher-transformed coe cients). This analysis extended our observation of a globally higher conservation between organs at the level of translational output (protein production) than at the level of transcript abundance.

B and C Normalised CDS read counts (RPKM) in rat liver vs. heart at the RPF-seq (B) and RNA-seq level (C), averaged over the ve replicates used in the study of Schafer et al., 2015. Note the overall higher Spearman correlation at the footprint level as compared to the mRNA level. The di erence between correlations is highly signi cant with p=2.3e-201, Z value=-30.25; Steiger test for di erence between 2 dependent correlations (Reference: Steiger JH, 1980. Psychological Bulletin, 87, 245-251).

Figure S8. Analysis of transcript features with predictive value for di erential TE.

A Cumulative distribution of the indicated transcript features for single isoform genes that do not show di erential TE (black, n= 5278), or that show di erential TE and either higher TE in kidney (yellow, N=193) or in liver (green, N=340). The indicated p-values are Kolmogorov-Smirnov test results of each group vs. 'all'. Statistically signi cant comparisons marked in red.

B Same as (A), but in the form of boxplots and using Wilcoxon rank-sum test for the di erences between group means (again, marked in red, signi cant results).

C Fraction of single isoform genes with (blue) or without (red) translated uORFs in either organ. The group with di erential and higher TE in liver contained signi cantly more translated uORF-containing transcripts than the genes not showing di erential TE in either organ (p=6.08e-04; Fisher's exact test); for kidney, there was a slight depletion of uORF-containing transcripts (non-signi cant). This analysis indicates that uORF usage may play a role in setting TE di erences across tissues. Note that in this analysis, we did not yet distinguish whether the uORF was translated in liver and/or kidney, but we treated the 1377 genes with a translated uORF in at least one organ as a single group.

D Organ-speci c uORF usage and its association with di erential TE. The group of genes with uORFs speci cally translated in liver was enriched for transcripts better translated in in kidney, and vice versa, consistent with a role of tissue-speci c uORF usage in setting TE di erences. However, due to the low number of di erential TE genes exhibiting uORF translation that was exclusive to one organ for this analysis, the enrichments and depletions did not reach statistical signi cance.

E Scatterplot of upstream ORF vs. CDS TE di erences across organs for genes containing translated uORFs in both organs and detected as di erential TE with

higher TE in kidney (yellow) or liver (green), or not showing di erential TE (grey). An anticorrelation between uORF usage and CDS TE was only observed for genes with di erential and higher TE in liver.

Figure S9. Relationship between transcript diversity and di erential TE.
A Cumulative distribution of the absolute kidney-to-liver TE ratio for genes whose transcript diversity is present or absent only in the indicated feature. The vertical dotted grey line marks the 1.5-fold di erence used to de ne di erential TE. In this Figure, all 7 groups are plotted (transcript diversity only/not in 5′ UTR, CDS, 3′ UTR; all genes); for better visibility, the analyses of individual features are also shown in separate panels, i.e. in Fig. 2G (5′ UTR), Fig. S9B (CDS) and Fig. S9C (3′ UTR). Collectively, these results showed that transcript diversity that originated only within the CDS (red), or that was excluded from the 5′ UTR (purple), or that was present only within the 3′ UTR (dark green), all showed smaller TE di erences across organs, thus pointing towards variability within the 5′ UTR as a contributor to tissue-speci c TE.
B As in (A) showing the genes with transcript diversity present (red) or absent (orange) only in the CDS. Note that when transcript diversity is only in CDS (i.e. UTRs are identical), there is a signi cant shift to more similar TEs in both organs. This is consistent with the speci c association of 5′ UTR diversity with di erential TE that is shown in Fig. 2G.
C As in (A) showing genes with transcript diversity present (dark green) or absent (light green) only in the 3′ UTR.

Figure S10. Analysis of phase di erences in RNA and RPF rhythms in kidney and across organs.
A Histogram of phase di erences (RPF – RNA, in hours) for all genes that were detected as rhythmic in the kidney RPF and RNA data (N=542; see Fig. 3A). Although the distribution mean was not signi cantly di erent from 0, more genes had their footprint abundance peak advanced (N=282) than delayed (N=260) with respect to their mRNA abundance peak.
B and C Histogram of the phase di erences (footprints to mRNA abundance, in hours) in kidney (B) and liver (C) for the 178 genes rhythmic in both organs (gene set shown in Fig. 3C). We observed a broader distribution of phase di erences in kidney and globally a phase advance of RPF with respect to RNA (-0.143 hours), as compared to overall stronger phase coherence of RPF and RNA in liver. See also Figures 3D-E.
D Histogram of the di erential (kidney – liver) phase di erence (RPF – RNA) for the 178 genes that were rhythmic throughout (Fig. 3C). Although statistically not reaching signi cance, the mean of -0.178 hours and the overall more genes for which the phase di erence had negative values (96 vs. 82 genes) were consistent with the nding that RPF rhythms peak earlier than RNA rhythms speci cally in kidney.

Figure S11. Higher expression of deadenylase complex subunits in kidney.
A and B Daily expression pro les of the CCR4-NOT complex components in kidney (A) and liver (B) at the RNA (orange) and RPF (blue) level.
C RPF expression of the CCR4-NOT subunits (averages over the day). Boxplots represent the interquartile range and whiskers extend to the minimum and maximum expression within 1.5 times the interquartile range. Note that di erences in protein biosynthesis are statistically signi cant for all subunits ($p<0.05$, two-sample t-test) apart from *Cnot8*.

Figure S12. Heatmaps of all detected RPF and RNA rhythms indicate false-negatives of the rhythmicity detection method.
A Same as Fig. 3A, but re-plotted here for ease of comparison with (B-D).
B-D Heatmap of RNA-seq (left) and RPF-seq (right) expression for genes detected as rhythmic only at the mRNA level (B, N=796), at both levels (C, N=542), and at the ribosome footprints only (D, N=435) in kidney. Gene expression levels are standardised by row (gene). Please note that even the panels that should represent "non-rhythmicity" (i.e. right panel in B and left panel in D) clearly showed underlying rhythmicity, albeit with more noise and/or lower amplitude. Many of these cases were therefore probably not truly "non-rhythmic" but rather false-negatives of the detection method (see Results section).

Figure S13. Core clock gene expression at RNA and RPF levels in both organs.
Left panels: Daily expression pro les of the 12 main core clock genes shown in Fig. 5A-C. **Right panels:** Hierarchical clustering of the organs' RNA and RPF pro les for each clock gene. Branch height represent the average Euclidean distance. Note that for 7 out of the 12 core clock genes, protein synthesis pro les were more conserved across organs than mRNA abundance and than RPF-RNA within organs.

Figure S14. Read distribution for uORF-containing core clock genes.
A Normalised read distribution for RPF (in blue) and RNA (in orange) along core clock transcripts containing uORFs in kidney (top) and liver (bottom) for the timepoint of maximal CDS translation. Red boxes indicate AUG-initiated uORFs as predicted in our analyses. For scaling issues and better visualisation, only a portion of the 3′ UTRs, corresponding to the same length as the full 5′ UTR, is depicted (exception *Nr1d1*, for which the 3′ UTR is so short that it is shown full length).
B Read distribution to the three translation frames showed a frame bias of footprint reads for most predicted uORFs that was in a similar range as the frame bias on the CDS. This frame preference is indicative of active translation on the uORFs.

uORF-mediated regulation of CLOCK protein abundance identified by translatome-wide mapping of targets of reinitation factor DENR*

Violeta Castelo-Szekely, Mara De Matos and David Gatfield

*Running Title. Manuscript in preparation.

uORF-mediated regulation of CLOCK protein abundance identified by translatome-wide mapping of targets of reinitation factor DENR*

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

Abstract

Translation initiation can be regulated through several mechanisms, including by upstream open reading frames (uORFs). DENR-MCTS1 are reinitiation factors that promote efficient translation on a subset of short and strong uORF-containing transcripts in *Drosophila*. We had previously identified uORF translation and DENR-mediated reinitiation as a novel regulatory mechanism of circadian clock function, but the DENR targets responsible for this phenotype are unknown. Here, we have used ribosome profiling in NIH3T3 cells to annotate uORFs acting through reinitiation transcriptome-wide, and have characterised the uORF specificities that confer DENR dependance through a comprehensive regression model. We identified 240 transcripts that required DENR for efficient protein production, including genes associated with increased proliferation and cancer. The presence of multiple uORFs, the start codon identity and sequence context, and the distance to the CDS start, were correlated with the magnitude of translational regulation, indicating a conserved mechanism of action. Finally, among circadian core clock genes, *Clock* was found as a DENR target. *Clock* contains two AUG-initiated uORFs, the second of which overlaps the main CDS and is followed by an additional start codon in frame with the annotated CDS one, that could give rise to a 5' alternative CLOCK protein. Systematic mutation of the uORF, annotated or alternative CDS start codons revealed that DENR would be required for reinitiation and biosynthesis of the alternative CLOCK isoform after translation of the overlapping uORF. Taken together, our study provided insights into uORF-mediated translational mechanisms that can regulate circadian clock function and gene expression at large.

*Running Title

[1]Center for Integrative Genomics, Génopode, University of Lausanne, 1015 Lausanne, Switzerland

Introduction

Initiation is the main rate-limiting step in translation, allowing rapid, reversible and transcript-specific control of gene expression [Jackson et al., 2010]. Translation initiation can be regulated by several mechanisms, including through upstream open reading frames (uORFs), short translated sequences defined by a start codon within the 5' untranslated region (5' UTR) of the transcript and an in-frame stop codon. Since eukaryotic translation involves scanning of the 5' UTR till a start codon is recognised, uORF translation requires the ribosome to resume scanning after termination, and reinitiate at the downstream coding sequence (CDS). Alternatively, the 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 can be regulated by several uORF and transcript features (reviewed in Barbosa et al. [2013]; Wethmar [2014]).

Density regulated protein (DENR) and Malignant T-cell amplified sequence 1 (MCTS1) form an heterodimer that binds the small ribosomal subunit and can promote eIF2-independent recruitment of Met-tRNA$_i$ to the ribosome on viral mRNAs, as well as ribosomal subunit recycling *in vitro* [Skabkin et al., 2010]. In *Drosophila*, DENR-MCTS1 was recently identified as a reinitiation factor required for efficient translation on short and strong uORF-containing transcripts [Schleich et al., 2014]. Furthermore, DENR targets in human cells have been predicted through an approach that used the DENR dependance measured for several synthetic uORF-containing reporters – of varying length or Kozak strength – to derive translational regulation genome-wide [Schleich et al., 2017]. However, an *in vivo* identification of DENR targets and a direct and transcriptome-wide quantification of the translational regulation exerted by DENR is 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 a reduced translation efficiency on

the main CDS due to the inefficiency of reinitiation and leaky scanning [Wethmar, 2014]. Previously, we observed that, also among transcripts comprising the core circadian clock circuitry in mouse liver, uORF translation was abundant. Moreover, loss-of-function of DENR led to a robust circadian period shortening in cells, uncovering a new layer of regulation of circadian function [Janich et al., 2015], but the DENR targets within the core clock that could underlie this phenotype remain unknown.

Here, we used ribosome profiling in murine NIH3T3 fibroblasts to quantify DENR-mediated translational regulation transcriptome-wide, and in particular, within core clock genes, and used a comprehensive regression analysis to characterise uORF features associated with translational regulation exerted by DENR.

Results

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

We hypothesized that if *Denr* is required for efficient protein production from transcripts that undergo re-initiation at the main coding sequence (CDS) 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 (figure 1a). We therefore performed shRNA-mediated knockdown of *Denr* in murine 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]. After lentiviral transduction with two different *Denr*-targeting shRNAs (shRNA1, shRNA2; in triplicate) or control RNAs (scramble shRNAs, in triplicate; Gfp shRNA in duplicate), cells were grown to confluency and pretreated with 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 figure S1a), that were enriched for protein coding sequences (Supplementary figure S1b) and that showed the characteristic triplet periodicity expected from high-quality footprint data (Supplementary figure S1c). Across replicates, CDS-mapping normalised reads were highly correlated ($R2 > 0.97$, Supplementary figure S1d), 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 (Figure 1b, 1c and Supplementary figure S1e).

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 translation as the ratio of 5' UTR to CDS-mapping footprint reads for *Denr*-depleted and control cells in a transcriptome-wide fashion. 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. 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]. For this gene set, we observed a significant increase in relative translation of 5' UTRs in *Denr*-deficient cells (Figure 1d), reflecting a shift from CDS to 5' UTRs that was specific for the ribosome footprints and absent from the RNA-seq data (Figure 1e). Nevertheless, on pairwise comparisons, less than 500 genes showed a translational shift that was statistically significant (FDR < 0.1, n $= 493$, Supplementary figure

S2a), supporting evidence that DENR-MCTS1 regulate protein synthesis rates on a subset of transcripts [Schleich et al., 2014].

The higher 5' UTR footprint density upon *Denr* depletion, which was more pronounced for the distal part of the UTR than for the proximal part (Figure 1f), suggested a redistribution of translating ribosomes, conceivably due to the presence of uORFs. The change on the translational landscape likely reflects alterations in the extent of leaky scanning and uORF usage, and is 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 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 (Figure 1d), we wished to narrow down to DENR targets that were potentially functionally relevant. We reasoned that changes in translation efficiency (TE) on the CDS 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 translation efficiencies (TEs), i.e. the ratio of CDS-mapping RPF-seq to RNA-seq reads, and then applied several available algorithms to detect significant changes in TE. The four methods we employed, Xtail [Xiao et al., 2016], RiboDiff [Zhong et al., 2017], Riborex [Li et al., 2017], and Babel [Olshen et al., 2013], are all conceived for the quantification of translational changes that deviate from the variation expected due to mRNA level differences. They differ in parameter estimation and modeling approaches, as well as in sensitivity, false discovery rates and computational speed. Used with options that maximize comparability of results (see Materials and Methods), we identified 240, 78, 41 and 30 genes that had a significant change in TE between *Denr*-deficient and control cells in Xtail, RiboDiff, Riborex, and Babel, respectively

(FDR < 0.1 for all four algorithms). Importantly, differential TE genes detected by Riborex, RiboDiff and Xtail completely overlapped, while only about half of the genes detected by Babel were also found by any of the other methods (Supplementary figure S3a). Moreover, most genes detected by Xtail, RiboDiff and Riborex showed decreased TE in *Denr*-depleted cells, as expected for the knockdown of a reinitiation factor, and they spanned the full range of expression levels (Figure 2a, and Supplementary figure S3b). By contrast, Babel identified similar numbers of up and down-regulated TE genes and was biased towards low abundance transcripts (Supplementary figure S3b). From these results, we concluded that Xtail was likely the most suitable method to identify DENR targets, and further estimated the false discovery rate of the algorithm by randomly shuffling the datasets, which eliminated nearly all differential TE genes (Supplementary figure S3c).

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 and periodicity (Supplementary figure S4a). Using this approach, we detected translated uORFs in nearly half of all expressed genes (n = 4755 genes, Figure 2b). 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 and non-overlapping uORFs (28.7%, n = 1365 genes) were frequent, too (Supplementary figure S4b). Notably, DENR targets were highly enriched for genes with translated uORFs (72.7%, p = 5.73e-4, Fisher test) and, in particular, with non-overlapping uORFs (Supplementary figure S4b).

Moreover, relative 5'UTR translation in *Denr*-depleted vs. control cells calculated over the full 5' UTR – as a proxy of uORF translation – was significantly higher for DENR targets than for the whole of uORF-containing genes (Supplementary figure S4c, and S4d). In summary, these observations were in line with an involvement of DENR in leaky scanning or reinitiation and efficient CDS translation of a subset of translated uORF-containing genes.

We next analysed whether the 240 differential TE genes showed any functional enrichment. GO term analysis revealed categories associated with translation and mRNA stability, as well as cellular functions related to protein phosphorylation and catabolism, kinase binding and cell cycle (Figure 2c and Supplementary figure S4e), in line with previous observations 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 (Figure 3a), as well as a shift of ribosome occupancy from the CDS to the 5' UTR in *Denr*-depleted cells (Figure 3b). This signature was characteristic of many differential TE genes (see Supplementary figure S5a and S5b), and correlated with the presence of translated uORFs (Supplementary figure S5c).

In order to validate the DENR dependance for several of the identified targets, we cloned the 5' UTR of the transcript of interest upstream of *Firefly* luciferase (FL) in a lentiviral vector that also expresses *Renilla* luciferase (RL) from the same promoter, for internal normalisation (Figure 3c, top). *Denr* depletion led to a 59 - 87% reduction in F/R luciferase signal, indicating that DENR was necessary for efficient reporter expression. Furthermore, mutation of the uORF start codon, increased luciferase signal and partially abolished DENR dependance (Figure 3c, bottom).

Taken together, we concluded that DENR regulates protein synthesis rates of a set of transcripts

containing uORFs and involved in translation, RNA stability and kinase activity, by promoting uORF by-passing and/or efficient reinitiation.

Linear regression analysis correlates DENR-dependence with uORF number, start codon identity, and position within the 5' UTR

There is evidence for a bewildering diversity of mechanisms through which uORFs can act. For example, they can affect main ORF translation (by affecting start site selection/leaky scanning/re-initiation/ribosome shunting), transcript abundance (by eliciting nonsense-mediated mRNA decay), and in some cases produce 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 features render certain uORF-containing transcripts DENR-dependent, in contrast to the majority of uORF-containing transcripts whose TE was not affected by DENR. Previous studies using synthetic uORF reporters in Drosophila S2 [Schleich et al., 2014] and human HeLa cells [Schleich et al., 2017] showed that very short uORFs (coding for 1-2 amino acids) with favourable Kozak contexts around the start codon were most strongly dependent on DENR for efficient expression (via reinitiation) of the downstream main coding sequence.

Here, we used a multiple regression analysis that modeled the CDS TE change observed upon *Denr* knockdown, considering nine features that could potentially regulate uORF TE and efficient translation reinitiation on the CDS. These features included the number of uORFs, the uORF start codon identity 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 (Figure 4a). The model was applied on all translated uORF-containing genes, and yielded several features with significant predictive value for regulation by DENR (Figure 3b-j, adjusted R2 = 0.057, F-statistic = 33.21 on 4731 DF, $p < 2.2e\text{-}16$). In

particular, the number of uORFs present on a transcript was significantly associated with TE regulation (Figure 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 (Figure 4c, b = 0.005, p = 0.44), as expected, given that re-initiation downstream of such uORF would already be within the body of the main CDS.

A second feature predictive of TE regulation by DENR was the uORF start codon. 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]. We computed the proportion of AUG uORFs present on a transcript to account for the presence of multiple uORFs with different start codons, allowing us to discriminate DENR preference for canonical vs. non-canonical start codons. The model showed that AUG uORF-containing transcripts showed the strongest TE reduction upon *Denr* knockdown (Figure 4d, b = -0.05, p = 1.15e-9), indicating that DENR could control the fidelity of start codon selection and, contrary to eIF1, bypass AUG uORFs on its targets to ensure efficient protein production. Moreover, DENR targets were enriched for AUG uORFs as compared to the ensemble of translated uORF-containing transcripts (p = 0.0044, Fisher test, Supplementary figure S6a), and genes harboring only AUG uORFs showed the strongest effect on CDS TE in *Denr*-deficient cells (Supplementary figure S6b). As shown before [Schleich et al., 2014, 2017], stronger Kozak contexts around uORF start codons were also associated with TE regulation (Figure 4e, b = -0.026, p = 0.013), suggesting that DENR allows leaky scanning of otherwise strong AUG uORFs. By contrast, the kozak context around the main ORF start codon bore no predictive value (Figure 4f, b = -0.0006, p = 0.38).

The location of a uORF with respect to the main ORF can dictate the efficiency of reinitiation, with longer spacing generally allowing the reacquisition of the necessary initiation factors to the scanning ribosome [Kozak, 1987]. Our model recapitulated this notion for DENR-regulated transcripts, where larger distances from the 3'-most uORF stop codon to the main ORF start were associated with a downregulation of TE (Figure 4g, b = -0.0101, p = 7.89e-7). This finding indicates that uORFs located further from the CDS are more strongly dependent of DENR for efficient reinitiation. This effect of uORF-CDS distance was not caused by longer 5' UTRs *per se*, as the distance from the 5' cap to the first uORF was not associated with DENR regulation (Figure 4h, b = -0.0018, p = 0.18). Moreover, the uORF-CDS distance was only weakly correlated with 5' UTR length (spearman rho = 0.16), and the correlation of 5'cap-uORF distance with 5' UTR length was even stronger (spearman rho = 0.27, Supplementary figure S6c, S6d).

No linear relationship was found between uORF length and the CDS TE (Figure 4i, b = 0.0006, p = 0.81), in agreement with previous results in human cells using synthetic uORFs of >9 nt in length [Schleich et al., 2017]. Of note, the length dependance reported in Schleich et al. [2017] was only found for 1 (or maximum 2) amino acid uORFs, which were excluded from our analyses (see Materials and Methods), thus potentially missing the regulatory role that such short uORFs could exert.

Finally, the uORF GC content showed a significant relationship with TE fold change (Figure 4j, b = 0.058, p = 0.005), suggesting that potential weaker secondary structures (low GC) favour DENR-mediated uORF translation, impairing efficient protein production from the main CDS.

We chose to use the CDS TE fold change as response – as opposed to a binary DENR-target vs. non-target response – in order to account for DENR regulation in a continuous manner that would not miss smaller magnitude cases of DENR control. Of note, however, such a logistic regression model also recapitulated our 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 distance from the last uORF to the CDS start (b = 0.005, p = 0.034) (Supplementary figure S6e, 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.

Together, our results showed that DENR facilitated efficient protein synthesis from transcripts containing AUG uORFs in strong kozak contexts, and that the distance of the most downstream uORF to the CDS start was critical for efficient reinitiation.

DENR-mediated reinitiation controls the levels of CLOCK protein biosynthesis*.

We recently showed that uORF translation is pervasive within core clock transcripts, and that depletion of *Denr* in NIH3T3 cells caused a robust circadian period shortening of 1-1.5 h, suggesting that circadian function can be regulated by uORFs and DENR-mediated reinitiation [Janich et al., 2015]. We therefore sought to uncover the specific core clock transcripts/uORFs acting through reinitiation that could explain the circadian phenotype.

Most core clock genes showed changes affecting mostly transcript levels (e.g. *Bmal1*), or small and unidirectional variations in both transcript abundance and absolute translation (e.g. *Nr1d1*, *Per1/2*, *Cry1/2*, *Ror*α) that therefore did not lead to changes in translational efficiency (Figures 5a, 5b, and Supplementary figures S7a, S7b). For *Clock*, however, the expression change was mostly translational (Figure 5a and Supplementary figure S7a), resulting in a $\sim$20% reduction in *Clock* protein biosynthesis rate (Figure 5b and Supplementary figure S7b). Of note, *Clock* did not pass the significance threshold in our analysis of differential CDS TE (Figure 2a and Supplementary File 1); however, Xtail algorithm evaluates differential TE statistical significance through two separate tests - one regarding differences between the RNA FC and the RPF FC distributions, and a second one evaluating differences in TE

*Results and conclusions from this section are not definite; experiments in progress.

distributions directly - and then derives the final p-value from the most conservative of the two tests [Xiao et al., 2016]. This additional test can therefore prevent false positives but miss differential TE genes close to the threshold. Importantly, for one of the tests, the TE fold change for *Clock* was significant ($p = 0.025$ in test 1; $p = 0.19$ for test 2; Supplementary File 1), and could thus represent a DENR target at the borderline of statistical significance.

In our analyses, we annotated two translated AUG-initiated uORFs, the second of which overlaps the main *Clock* CDS and could thus hamper translation of the annotated CDS. Downstream of this uORF 2 stop codon, there is an additional AUG in the coding frame of *Clock* cDNA, that could thus give rise to an alternative 5' truncated CLOCK protein, that would be 9 amino acid shorter, when uORF-2 is translated (Figure 5c and Supplementary figure S7c). We observed increased translation on both uORFs in *Denr*-deficient cells, leading to a reduced translation on the CDS (Figure 5d). In particular, uORF 2 showed $\sim$23% higher translation rate upon *Denr* depletion (Figure 5e), suggesting that DENR-mediated leaky scanning of, or reinitiation after, uORF 2 is crucial for efficient CLOCK production. We looked for further evidence of uORF 2 and alternative CDS translation using GWIPS-viz, an online genome browser for the visualisation of ribosome profiling data from published studies [Michel et al., 2014]. Initiating ribosomes can be detected on both uORF 2 and alternative CDS start codons in human and mouse data, and elongating ribosome profiling studies showed good footprint coverage on uORF 2 and the characteristic accumulation of ribosomes at the start and stop codons (Supplementary figure S8). Together, our and published data suggested abundant uORF 2 translation and CLOCK protein synthesis from the alternative AUG.

In order to investigate uORF usage, effect, and DENR dependance for CLOCK biosynthesis from the annotated and alternative CDS, we cloned *Clock* 5' UTR and first 14 codons (containing both the annotated and alternative ATGs) upstream of *Firefly* luciferase in our dual luciferase vector (Figure

5f, top). We then systematically mutated the uORFs, annotated or alternative CDS start codons, or combinations thereof (Supplementary figure S7d), and carried out dual luciferase assays in control and *Denr*-depleted cells. *Denr* depletion resulted in a 20% reduction in luciferase signal for the *Clock Wt* construct (Figure 5f, bottom), in line with the decrease in protein production quantified *in vivo* (Figure 5d). Interestingly, mutation of the annotated CDS initiation codon did not affect CLOCK synthesis or DENR dependance when compared to *Clock Wt*; however mutation of the alternative ATG decreased luciferase signal by $\sim$95% and abolished DENR dependance (Figure 5f, bottom), suggesting that most CLOCK biosynthesis in NIH3T3 cells actually comes from the alternative ATG, and that DENR is required for reinitiation after uORF 2. Mutation of uORF 1 led to a 14% increase in luciferase signal and partial relief of DENR dependance. By contrast, uORF 2 mutation led to a 71% increase in protein synthesis and a reduced DENR dependance by 50%, indicating that uORF 2 translation strongly downregulated CLOCK biosynthesis and confers DENR dependance for efficient translation. Moreover, mutation of both uORF 1 + 2 start codons resulted in a 95% increase in protein synthesis, suggesting additive effects. In line with minor synthesis of the long CLOCK isoform, mutation of the annotated CDS start codon in combination with uORF mutations led to similar results than mutation of the uORFs alone. Furthermore, mutation of the alternative ATG in combination with uORF mutations led to strong decrease in luciferase signals, consistent with the predominant synthesis of the short CLOCK isoform. In particular, the case where both the alternative CDS and uORF 2 (or uORF 1+2) start codons were mutated – an scenario that would lead to CLOCK synthesis from the annotated ATG without inhibitory effects – resulted in a $>$80% reduction in CLOCK protein production (Figure 5f, bottom).

From these results we concluded that *Clock* requires DENR for efficient protein synthesis from the 9-codon shorter, alternative CDS, largely due to the strong translation of uORF 2. Moreover CLOCK start codon is likely wrongly annotated, at least for the case of NIH3T3 cells.

Discussion

DENR has been recently identified as a reinitiation factor in *Drosophila*, required for efficient translation of a subset of short and strong uORF-containing genes [Schleich et al., 2014]. Using a series of synthetic uORF-containing reporters, the effect of varying uORF lengths and Kozak context strength on translational regulation were measured in human HeLa cells, and used to derive genome-wide DENR targets and their predicted translational decrease upon Denr knockdown [Schleich et al., 2017]. Here, we have used ribosome profiling in *Denr*-deficient NIH3T3 cells to quantify CDS TE changes *in vivo*, identifying 240 transcripts that needed DENR for efficient protein production, and have characterised the uORF features that are associated with DENR dependance using a quantitative model approach.

DENR and its heterodimerization partner MCTS1 are oncogenes [Reinert et al., 2006], upregulated, for example, in prostate cancer CERAMI. Not surprisingly, many of the identified targets are therefore genes involved in proliferation and upregulated in several cancer types. Kelch Domain Containing 8A (KLHDC8A) protein is overexpressed in human gliomas that become resistant to epidermal growth factor receptor (EGFR) silencing, commonly used as therapeutic approach, allowing aggressiveness maintenance in these tumors through an unknown EGFR-independent pathway [Mukasa et al., 2010]. It is thus possible that upon loss of EGFR-silencing sensitivity, KLHDC8A levels increase through DENR-mediated translational control. Similarly, Vascular endothelial growth factor D (VEGF-D), or Mitogen-Activated Protein Kinase Kinase 5 (MAP2K5) are upregulated in different cancers [Thelen et al., 2008; Diao et al., 2016]. DENR could therefore be an attractive drug target candidate: decreasing DENR levels could, in turn, reduce the oncogenic levels of the proteins it regulates in proliferating, but not in quiescent, cells.

A comparison of DENR targets found in this and previous studies returned little overlap. Notably, *Dock1*, the mouse orthologue of Drosophila's *mbc* shown in Schleich et al. [2014] to be a strong DENR target, also showed DENR dependance in our reporter assays, although it was not detected in our

differential TE analysis, indicating that our list might not be fully exhaustive. Of the eight targets validated in HeLa cells in Schleich et al. [2017], all but one (*Tmem60*) were not expressed in NIH3T3 cells. *Tmem60* was predicted to result in a 25% translational downregulation upon *Denr* knockdown, whereas in our system a $\sim$8% decrease in CDS TE was quantified. Similarly, for the ensemble of 104 predicted targets by Schleich et al. [2017], $>$ 90% were not expressed or lacked a mouse orthologue in NIH3T3 cells, and others (e.g. *Lca5*, *Mllt11* or *Dsel*) were downregulated to a lesser extent. It is therefore likely that DENR targets are cell-type and organism-specific, but conceivably as well, the differences in the magnitude of translational regulation observed across studies are the result of different experimental and analytical approaches.

Several transcript and/or uORF features are known to regulate the efficiency of uORF skipping and translation reinitiation. The start codon and context can dictate the strength of its recognition; sequences upstream, within or downstream a uORF can be recognised by *trans*-acting factors affecting reinitiation [Mohammad et al., 2017]; and uORF length and structural complexity, as well as the intercistronic distance to the CDS, can influence reinitiation efficiency [Kozak, 1987, 2001]. What rules govern DENR dependance on certain uORFs but not on others thus remain an intriguing question. Using all uORF-containing transcripts annotated in our analyses, we have identified four prominent features that added predictive value to a comprehensive model of CDS TE regulation exerted by DENR. The number of uORFs, the start codon identity and sequence context, and the distance to the main ORF, were associated to the magnitude of translational regulation. The presence of several uORFs is a well-known modulator of CDS TE, by requiring successive reinitiation events [Calvo et al., 2009], and DENR-regulated transcripts would not be an exception. Similarly, uORFs located further from the CDS would require DENR for efficient reinitiation, consistent with the idea that longer distances allow loss of termination factors and reacquisition of the initiation machinery on time for CDS translation [Kozak, 1987]. Finally,

AUG-starting uORFs and those with strong Kozak contexts showed the strongest DENR dependance, as had been shown in fly [Schleich et al., 2014] and human cells [Schleich et al., 2017], suggesting a conserved mechanism of action.

On the other hand, the uORF length was not associated with DENR-mediated translational regulation. Of note, already between *Drosophila* and human cells, DENR dependance on uORF length was quantitatively different: in S2 cells, uORFs coding for up to 4 amino acids could be regulated by DENR [Schleich et al., 2014], whereas only 1 amino acid-coding uORFs were strongly DENR dependent in HeLa cells [Schleich et al., 2017]. Therefore, the uORF length specificities could differ across species, or reflect essential differences in methodological approaches: deriving genome-wide dependance from synthetic uORF-containing reporters [Schleich et al., 2017] could be biased to the results measured from a selected number of constructs. It is therefore not unlikely that *in vivo* targets contain uORFs of varying lengths, given that it is the time required for translation rather than the linear length *per se* what determines scanning resumption [Poyry et al., 2004; Calvo et al., 2009; Barbosa et al., 2013].

uORF-mediated translational regulation can occur through several different mechanisms: leaky scanning, in which uORFs are bypassed; reinitiation, in which the ribosome resumes scanning after uORF translation termination; and other mechanisms that involve peptide or ribosome-induced ribosome stalling and NMD [Wethmar, 2014]. Structural studies have recently shown that DENR-MCTS1 can bind the small ribosomal subunit close to the P-site and the mRNA entry channel [Lomakin et al., 2017; Weisser et al., 2017]. However iClip studies did not show DENR crosslinking to mRNAs (data not shown), suggesting that DENR travels on the ribosome or transiently interacts with it, although recognition and binding to specific motifs could occur through a yet unknown factor. DENR-MCTS1 presence on the ribosome could sterically clash with that of canonical initiation factors, such as certain eIF3 subunits [Lomakin et al., 2017; Weisser et al., 2017]. Since eIF3 is involved in translation termination

and ribosome recycling [Beznoskova et al., 2013], the recruitment of DENR-MCTS1 to the ribosome could displace eIF3 to promote reinitiation on a downstream ORF *via* DENR's SUI domain involved in start codon recognition [Yoon and Donahue, 1992]. Alternatively, DENR-MCTS1 could compete with and replace eIF1 for cap-dependent scanning, given that simultaneous presence of DENR-MCTS1 and eIF2-tRNA-Met$_i$ is not incompatible [Lomakin et al., 2017]. In this scenario, overexpression of DENR in proliferating cells would shift the balance to more DENR-MCTS1-containing ribosomes, leading to increased CDS translation through skipping of strong uORFs that would normally be translated in a eIF1-containing ribosome. Future studies that discern between leaky scanning and translation reinitiation would gain insights into the precise mode of action of DENR-MCTS1.

[For *Clock*-related discussion see Thesis.]

Methods

Cell culture. NIH3T3 and HEK293FT cells were cultured under standard conditions (DMEM; 10% FCS, 1% penicillin/streptomycin, all from Invitrogen; 37C; 5% CO2).

Lentiviral particles production in HEK293FT cells from pLkO.1puro-shRNA using envelope pMD2.G and packaging psPAX2 plasmids, and viral transduction of NIH3T3 cells were performed following published protocols [Salmon and Trono, 2007], with puromycin selection at 5 μg/ml for 4 days.

For dual luciferase assays, 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).

Cloning and plasmids. For the generation of lentiviral shRNA expression vectors, two different sequences targeting Denr were cloned into pLKO.1puro backbone vector (Addgene no. 10878, [Moffat et al., 2006]): shRNA1: GTACCACAGAAGGTCACGATA, corresponding to clone TRCN0000308443 of the TRC shRNA Library from the Broad Institute; and shRNA2: GTGCCAAGTTAGATGCGGATT, corresponding to clone TRCN0000098826. The empty pLKO.1puro vector, containing Gfp, and Scramble

(Addgene no. 1864) shRNAs served as control.

For the generation of dual luciferase (Firefly/Renilla) reporter plasmids, fragments containing the 5'UTR and first 5-14 codons of the selected targets were amplified by PCR and cloned into the BamHI site of the prLV1 dual luciferase reporter plasmid [Du et al., 2014]. The following primers were used for PCR: Dock1, forward: aaaggatccCCATCCTGTCGCGGCGGGGAAGGAATG, reverse: tttggatccactagtGACCCAGCGCGTCATGGTGGCGCTGG; 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.

The target 5'UTR mutants were generated by site-directed mutagenesis of the corresponding plasmid, using the following primers: Dock1-Mut, forward; GTCGCGGCGGGGAAGGAATGGgtaccGGCGGCCTAGGCGCAGAGTTTC, reverse: GAAACTCTGCGCCTAGGCCGCCggtacCCATTCCTTCCCCGCCGCGAC; Etaa1-Mut, forward: cgaaggatcCGACTTGCAAAGgtaccGCTGCACGGTGGCGCGCGGGCTC, reverse: GAGCCCGCG CGCCACCGTGCAGCggtacCTTTGCAAGTCGgatccttcg; Map2k5-Mut, forward: CTCTTACTCACAAGGACTACAGgctagcTTGTGGGTTCTGTTTTGTCCAC, reverse: GTGGACAAAACAGAACCCACAAgctagcCTGTAGTCCTTGTGAGTAAGAG; Klhdc8a-Mut, forward: GGCAGGGA GCCCGTGCAGCGCGgtaccGAGGCTGAGAGAGGGGACGCGCC, reverse: GGCGCGTCCCCTCTCTCAGCCTCggtacCGCGCTGCACGGGCTCCCTGCC; Clock-uORF1-Mut, forward: CCCAAAATCACCAGCAAGAGTTgctagcGTCAGTCACACAGAAGACGGCC, reverse: GGCCGTCTTCTGTGTGACTGACgctagcAACTCTTGCTGGTGATTTTGGG; Clock-uORF2-Mut, forward: AAAGTGAAAGAGGAGAAGTACAggt acCTACCACAAGACGAAAACATAAT, reverse: ATTATGTTTTCGTCTTGTGGTAGgtaccTGTACTTCTCCTCTTTCACTTT; Clock-AnnotCDS-Mut, forward: CAAGACGAAAACATAATGTGTTtcGGTGTTTACCGTAAGCTGTAG, reverse: CTACAGCTTACGGTAAACACCgaAACACATTATGTTTTCGTCTTG; Clock-AlterCDS-Mut, forward: GTTTACCGTAAGCTGTAGTAAAcaGAGCTCAATTGTTactagtg, reverse: cactagtAACAATTGAGCTCtgTTTACTACAGCTTACGGTAAAC.

Ribosome profiling. One 15-cm dish of confluent NIH3T3 cells was used per replicate. Ribosome profiling and parallel RNA-seq was performed in triplicate for Denr shRNA1, Denr shRNA2 and scr

shRNA, and in duplicate for Gpf shRNA, following the ART-Seq protocol by Epicentre-Illumina (Cat.no. RPHMR12126) with minor modifications as described below. Cells were pre-incubated with cycloheximide-supplemented media (100 μg/ml) for 2 min at 37C to arrest translation elongation. For the generation of ribosome-protected fragments, cell lysates were treated with 5 units ART-Seq nuclease 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 in Janich et al. [2015]. 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 TPM 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 TPM to RNA TPM 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 Kd and Wt 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 consistency among them, model raw read counts as following a negative binomial distribution, and use global normalization methods for sequencing depth that would unaccount 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 a errors-in-variables regression to test for significant changes in translation within and between conditions.

Differential TE was set at FDR < 0.1 for all four methods and performance for false positive discovery rate was further assessed by randomly shuffling Kd and Wt datasets which should not yield differential TE genes.

uORF annotation and translation analyses. For uORF detection, a set of transcripts was used for which reads can be unambiguously assigned to the 5'UTR, i.e. transcripts that are the only protein-coding isoform expressed (n=4637). In order to include as many transcripts as possible for the annotation, genes with multiple protein-coding isoforms were also considered if 1) an isoform represented $> 65\%$ of all expressed (as derived from our transcript expression estimation, see above, n=3211) or 2) all expressed isoforms had the same CDS start, in which case the transcript with the highest expression was selected (n=1826). For the selected transcripts, uORFs were annotated and considered as translated with the following criteria: 1) started with AUG, CUG, GUG or UUG, 2) had an in-frame stop codon within the 5'UTR or could overlap the CDS, 3) were at least 9 nt long, 4) had a coverage $> 25\%$, 5) showed a frame preference and 6) the preferred frame was the first one relative to the uORF 5'. Importantly, we considered only uORFs coding for at least 2 aminoacids (i.e. 9 nt, including the stop codon), as for shorter uORFs frame bias assessment could be more proned 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].

References

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

Beznoskova, P., Cuchalova, L., Wagner, S., Shoemaker, C. J., Gunisova, S., von der Haar, T., and Valasek, L. S. (2013). Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. *PLoS genetics*, 9(11):e1003962.

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.

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

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

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.

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.

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.

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

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.

Lomakin, I. B., Stolboushkina, E. A., Vaidya, A. T., Zhao, C., Garber, M. B., Dmitriev, S. E., and Steitz, T. A. (2017). Crystal Structure of the Human Ribosome in Complex with DENR-MCT-1. *Cell Reports*, 20(3):521–528.

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.

Mohammad, M. P., MunzarovaÂšPondelickova, V., Zeman, J., Gunisova, S., and Valasek, L. S. (2017). In vivo evidence that eIF3 stays bound to ribosomes elongating and terminating on short upstream ORFs to promote reinitiation. *Nucleic Acids Research*, 45(5):2658–2674.

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.

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

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

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

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5):511–515.

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

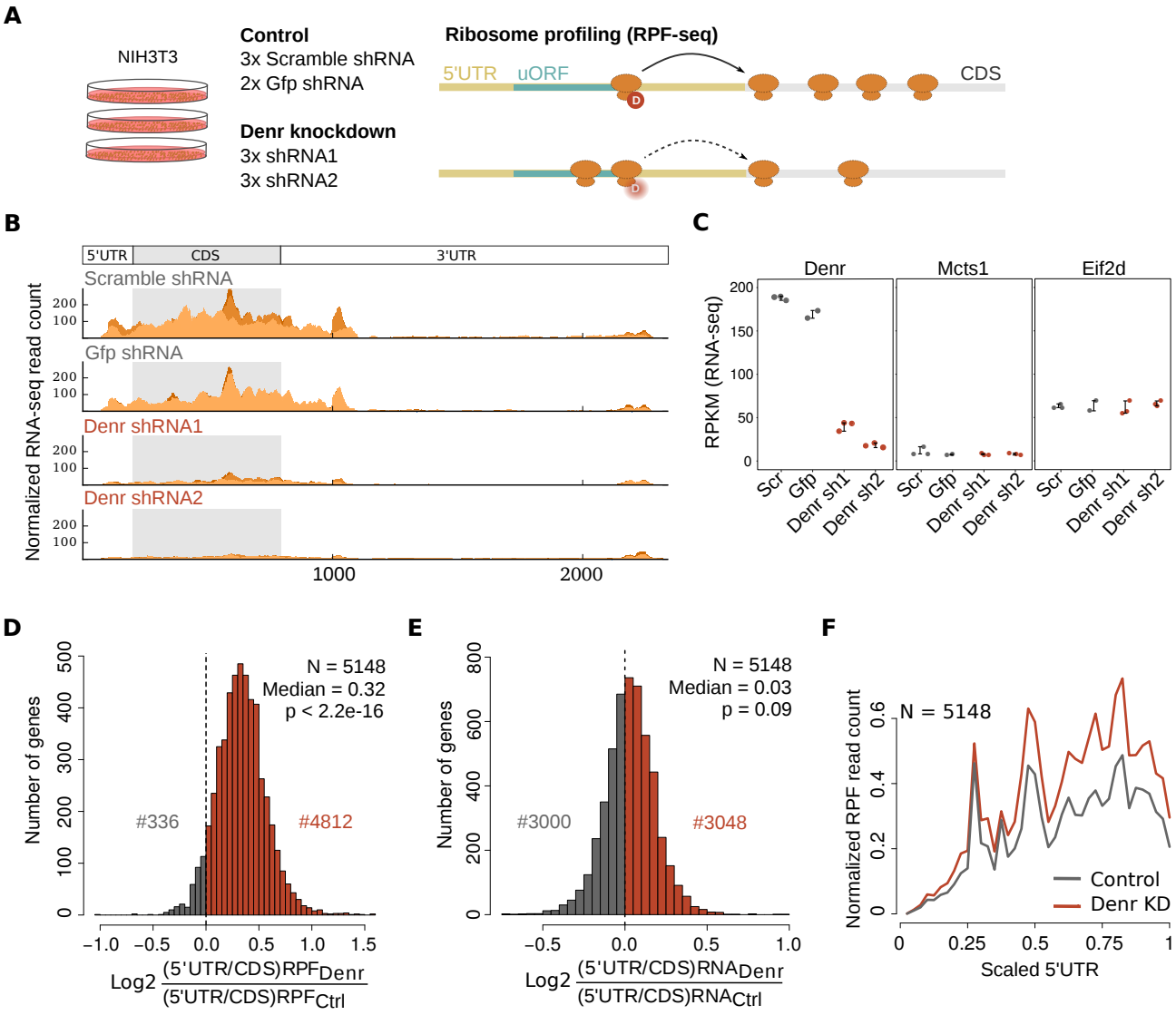
Weisser, M., SchÃďfer, T., Leibundgut, M., BÃűhringer, D., Aylett, C. H. S., and Ban, N. (2017). Structural and Functional Insights into Human Re-initiation Complexes. *Molecular Cell*, 67(3):447–456.e7.

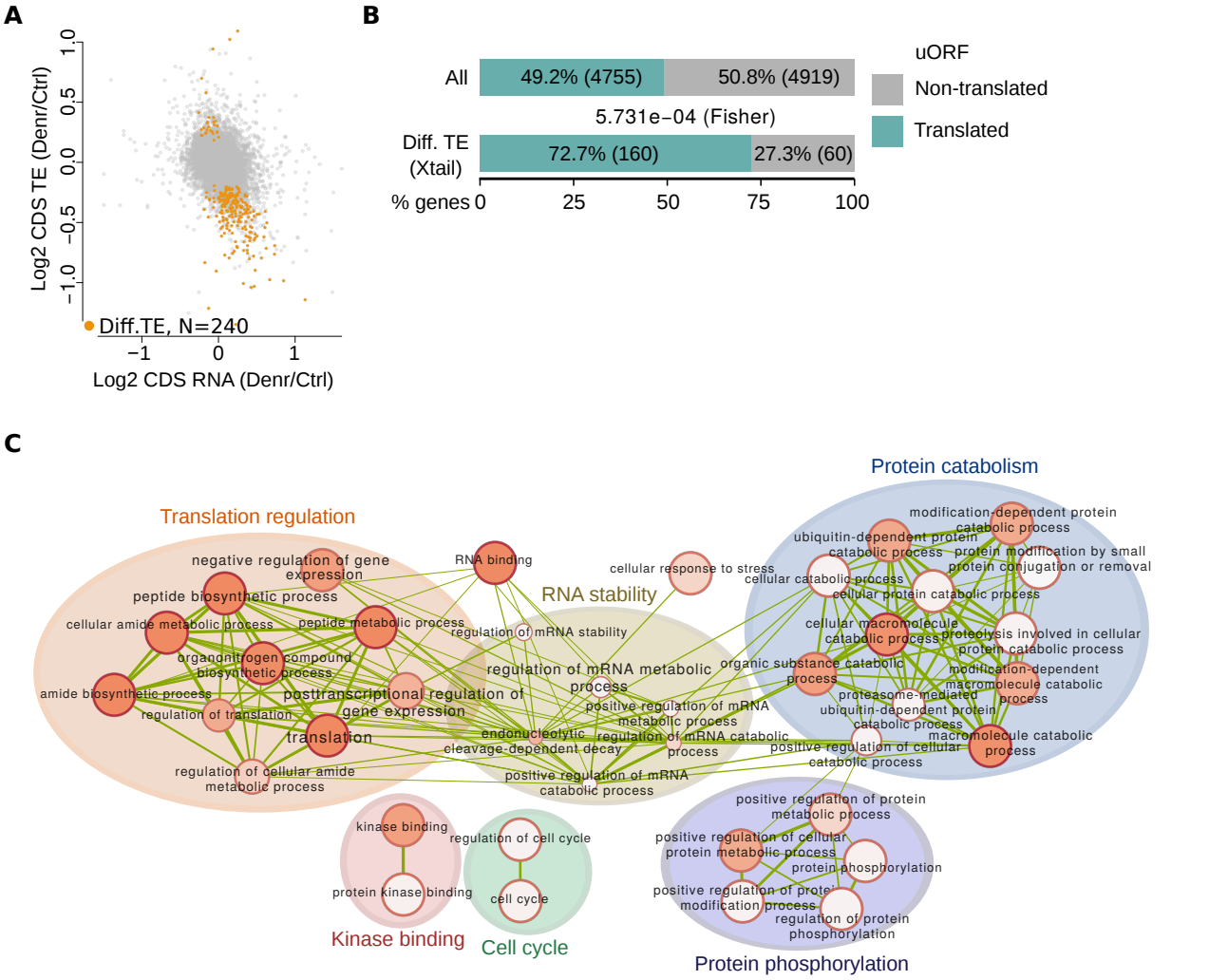
Wethmar, K. (2014). The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdisciplinary Reviews: RNA*, 5(6):765–768.

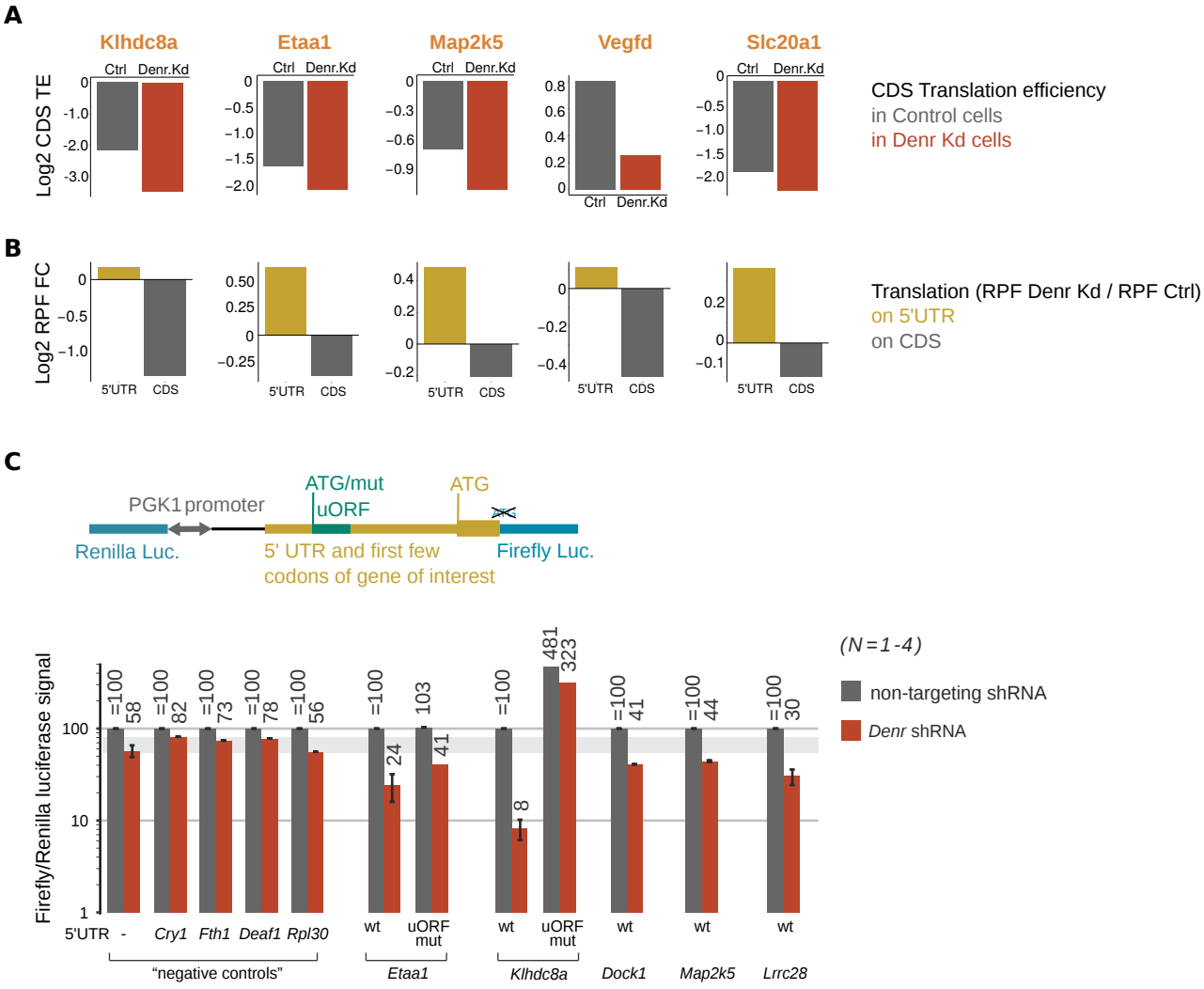
Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nature Communications*, 7:11194.

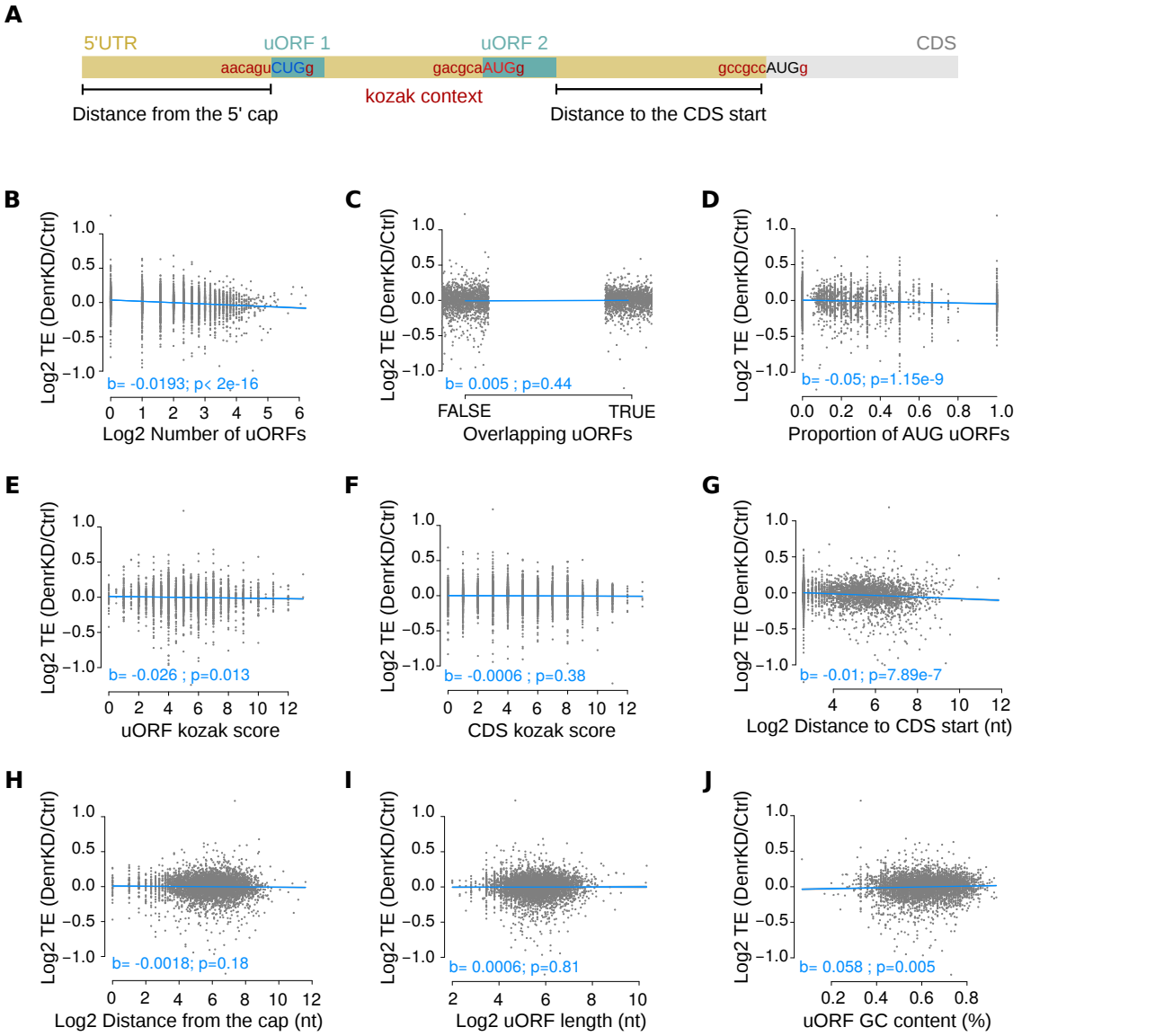
Yoon, H. J. and Donahue, T. F. (1992). The suil suppressor locus in Saccharomyces cerevisiae encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. *Molecular and Cellular Biology*, 12(1):248–260.

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.


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
Number of genes
N = 5148
Median = 0.32
p < 2.2e-16
#336
#4812
Log2 (5'UTR/CDS)RPFDenr / (5'UTR/CDS)RPFCtrl
E
Number of genes
N = 5148
Median = 0.03
p = 0.09
#3000
#3048
Log2 (5'UTR/CDS)RNADenr / (5'UTR/CDS)RNACtrl
F
N = 5148
Normalized RPF read count
Control
Denr KD
Scaled 5'UTR


A
Log2 CDS TE (Denr/Ctrl)
1.0
0.5
0.0
−0.5
−1.0
Diff.TE, N=240
−1
0
1
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 0
25
50
75
100
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
regulation of mRNA stability
RNA stability
cellular response to stress
regulation of mRNA metabolic process
positive regulation of mRNA metabolic process
endonucleolytic cleavage-dependent decay
regulation of mRNA catabolic process
positive regulation of mRNA catabolic process
posttranscriptional regulation of gene expression
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
positive regulation of cellular catabolic process
macromolecule catabolic process
kinase binding
protein kinase binding
Kinase binding
regulation of cell cycle
cell cycle
Cell cycle
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
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.
Firefly/Renilla luciferase signal
(N = 1 - 4)
non-targeting shRNA
Denr shRNA
5'UTR
Cry1
Fth1
Deaf1
Rpl30
wt
uORF mut
"negative controls"
Etaa1
Klhdc8a
Dock1
Map2k5
Lrrc28


A
5'UTR
uORF 1
uORF 2
CDS
aacaguCUGg
gacgcaAUGg
gccgccAUGg
kozak context
Distance from the 5' cap
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= -0.026 ; 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 (%)

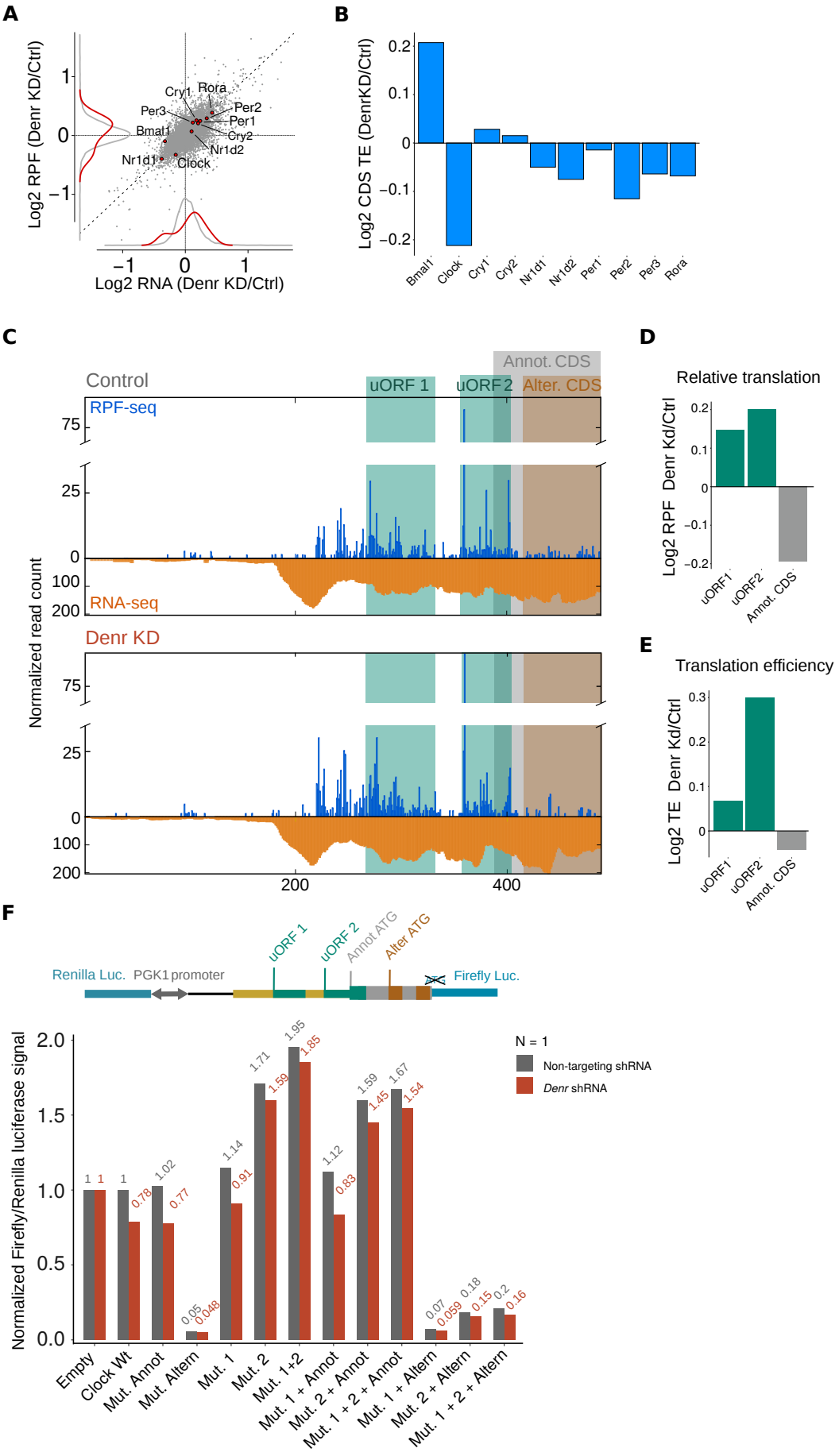

A
Log2 RPF (Denr KD/Ctrl)
Log2 RNA (Denr KD/Ctrl)
Cry1
Rora
Per3
Per2
Per1
Bmal1
Cry2
Nr1d2
Nr1d1
Clock
B
Log2 CDS TE (DenrKD/Ctrl)
Bmal1
Clock
Cry1
Cry2
Nr1d1
Nr1d2
Per1
Per2
Per3
Rora
C
Control
RPF-seq
RNA-seq
Denr KD
uORF 1
uORF 2
Annot. CDS
Alter. CDS
Normalized read count
D
Relative translation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

Figure legends

Figure 1. Translation is shifted towards 5'UTRs upon *Denr* knockdown. **A)** Experimental setup: ribosome profiling (and parallel RNA-seq) was performed on shRNA-mediated *Denr* knockdown NIH3T3 mouse fibroblasts, using scramble and *Gpf* shRNA as controls (wild-type cells, Wt). Two different shRNAs were used for knockdown experiments and ribosome profiling was carried out in triplicate for *Denr* shRNA1, *Denr* shRNA2 and 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 a reduced translation efficiency on the main coding sequence of DENR targets. **B)** *Denr* knockdown efficiency. Distribution of normalized RNA-seq read counts along the *Denr* transcript. Different color shadings (light/dark orange) correspond to the different condition replicates. **C)** RPKM values for CDS-mapping RNA-seq read counts for *Denr* (left), *Mcts1* (middle), and *Eif2d* (right). Each dot correspond to a sample and replicate and error bars indicate mean$\pm$sd. **D)** Histogram of 5' UTR-to-CDS translation (from RPF-seq) between *Denr* Kd and control cells. Genes with a minimum of 20 RPF-seq reads in the 5'UTR in 3/4 of samples were used (n = 5148), to exclude noisy genes. Relative 5' UTR translation was significantly higher in *Denr* Kd cells (median = 0.32, p < 2.2e-16, Wilcoxon signed rank test), suggesting a ribosome footprint redistribution to 5' UTRs upon *Denr* knockdown. **E)** Histogram of 5' UTR-to-CDS transcript levels (from RNA-seq) ratio 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 count on the 5' UTRs quantified in (C, D). 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, likely due to the presence of translated uORFs.

Figure 2. DENR targets show decreased CDS TE, are enriched for uORF-containing genes and are involved in translation and mRNA stability. **A)** Scatterplot of *Denr* Kd-to-Wt 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 a 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 an 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 considered. Node size is proportional to the number of genes associated with the GO category, shading corresponds with the statistical significance, and edge width is proportional to the number of

genes shared between functional categories.

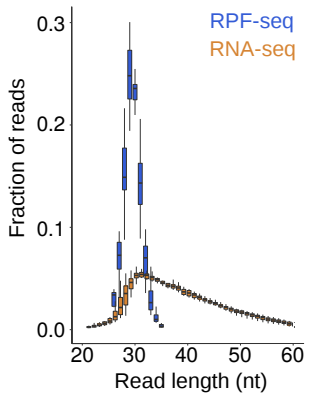
Figure 3. Mutation of uORF start codons on selected targets relieves DENR dependance. A) Representative examples of genes with differential TE. Panels show average CDS translation efficiency in Wt (gray) and *Denr* Kd (red) cells. **B)** *Denr* Kd-to-Wt translation on the 5'UTR (brown) and on the CDS (grey) for the same examples as in (A). DENR targets showed a decreased CDS TE (A) and an increased 5'UTR translation concomitant with a reduced CDS translation in *Denr*-deficient cells. **C)** *(Top)* Schematic representation of the dual luciferase reporter construct, where the 5' UTR of the gene of interest – with the Wt or mutated uORF – and first few codons of the CDS was cloned upstream of FL. *(Bottom)* Results of dual luciferase assay, where influence of the 5' UTR (Wt/mutant) on the expression of FL was measured in control and *Denr* knockdown cells, and internally normalised to RL, expressed from the same bidirectional promoter. Note that, for all targets tested, depletion of *Denr* reduced the F/R luciferase signal, indicating DENR dependance for efficient FL expression; and mutation of the uORF start codon relieved the inhibition and partially abolished DENR dependance. For each 5' UTR assayed, F/R signals were normalised to the scr-transduced non-mutated construct. Gray area correspond to the background of the assay, reflecting *Denr* effects that were not due to the presence/absence of the 5' UTR of interest. Empty vector ("-") contained only the vector-encoded 5' UTR. Three additional constructs – not detected as DENR targets – were used as negative controls.

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 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. Wt) on each uORF feature. Indicated in blue are the coefficient of the slope for the partial regression line and its significance level (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.

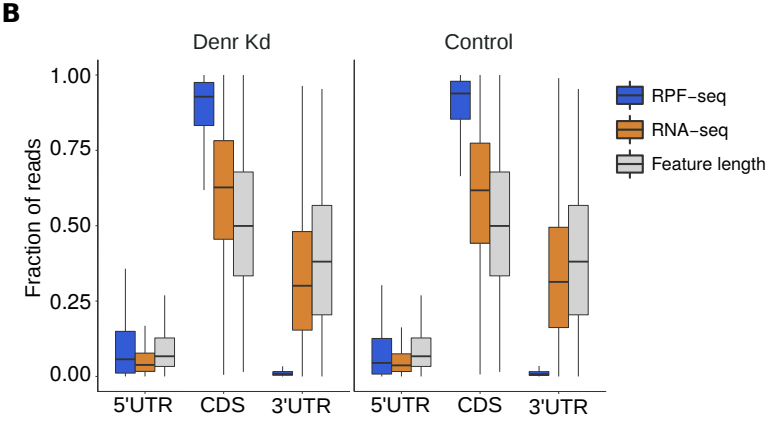
Figure 5. DENR-mediated reinitiation gives rise to CLOCK protein heterogeneity. A) Scatterplot of *Denr* Kd-to-control transcript levels (RNA-seq) vs. footprint levels (RPF-seq), with circadian core clock genes highlighted. *Clock* showed a decreased translation without changes in mRNA abundance. **B)** 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 area indicates the AUG-initiated uORFs; gray area corresponds to the annotated CDS; and brown are 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 empty vector/scr-shRNA, and all *Clock* mutant constructs were then renormalised to *Clock* Wt/scr-shRNA. Signal from all *Denr* shRNA-transfected *Clock* constructs were also normalised to empty vector/*Denr*-shRNA, to remove unspecific DENR effects.

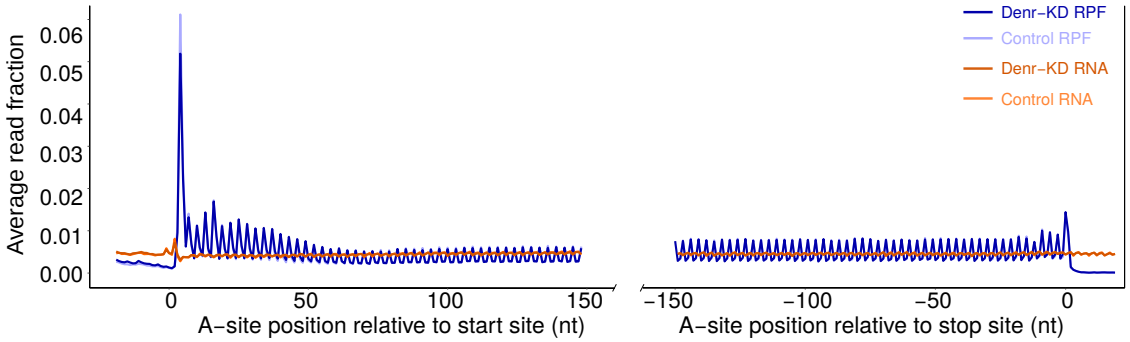
A



B



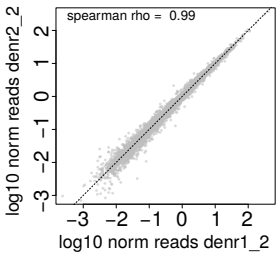
C



D

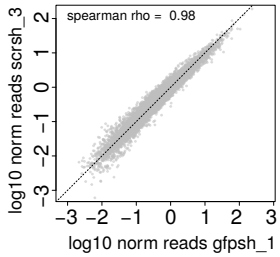
Correlation between samples RPF-seq

	Denr1_1	Denr1_2	Denr1_3	Denr2_1	Denr2_2	Denr2_3	Gfpsh_1	Gfpsh_2	Scrsh_1	Scrsh_2	Scrsh_3
Denr1_1	1	1	0.99	1	0.99	0.99	0.99	0.99	0.99	0.98	0.97
Denr1_2		1	0.99	1	1	0.99	0.99	1	0.99	0.99	0.98
Denr1_3			1	0.99	0.99	1	0.98	0.99	0.99	0.99	0.99
Denr2_1				1	1	0.99	0.99	1	0.99	0.99	0.98
Denr2_2					1	1	0.99	1	0.99	0.99	0.99
Denr2_3						1	0.99	0.99	0.99	0.99	0.99
Gfpsh_1							1	1	0.99	0.99	0.98
Gfpsh_2								1	0.99	0.99	0.98
Scrsh_1									1	0.99	0.98
Scrsh_2										1	1
Scrsh_3											1

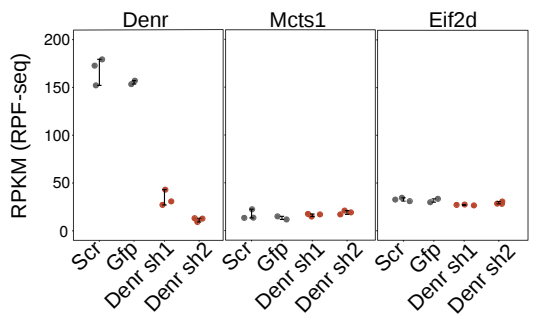


Correlation between samples RNA-seq

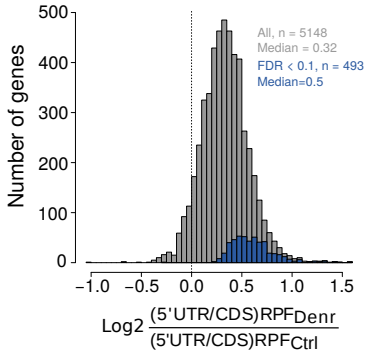
	Denr1_1	Denr1_2	Denr1_3	Denr2_1	Denr2_2	Denr2_3	Gfpsh_1	Gfpsh_2	Scrsh_1	Scrsh_2	Scrsh_3
Denr1_1	1	0.99	0.97	0.99	0.98	0.98	0.99	0.98	0.98	0.97	0.95
Denr1_2		1	0.99	0.99	1	0.99	0.99	1	0.99	0.99	0.97
Denr1_3			1	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99
Denr2_1				1	0.99	0.99	1	0.99	0.99	0.99	0.97
Denr2_2					1	1	0.99	1	0.99	1	0.98
Denr2_3						1	0.99	0.99	0.99	0.99	0.98
Gfpsh_1							1	1	0.99	0.99	0.97
Gfpsh_2								1	0.99	0.99	0.98
Scrsh_1									1	1	0.98
Scrsh_2										1	0.98
Scrsh_3											1

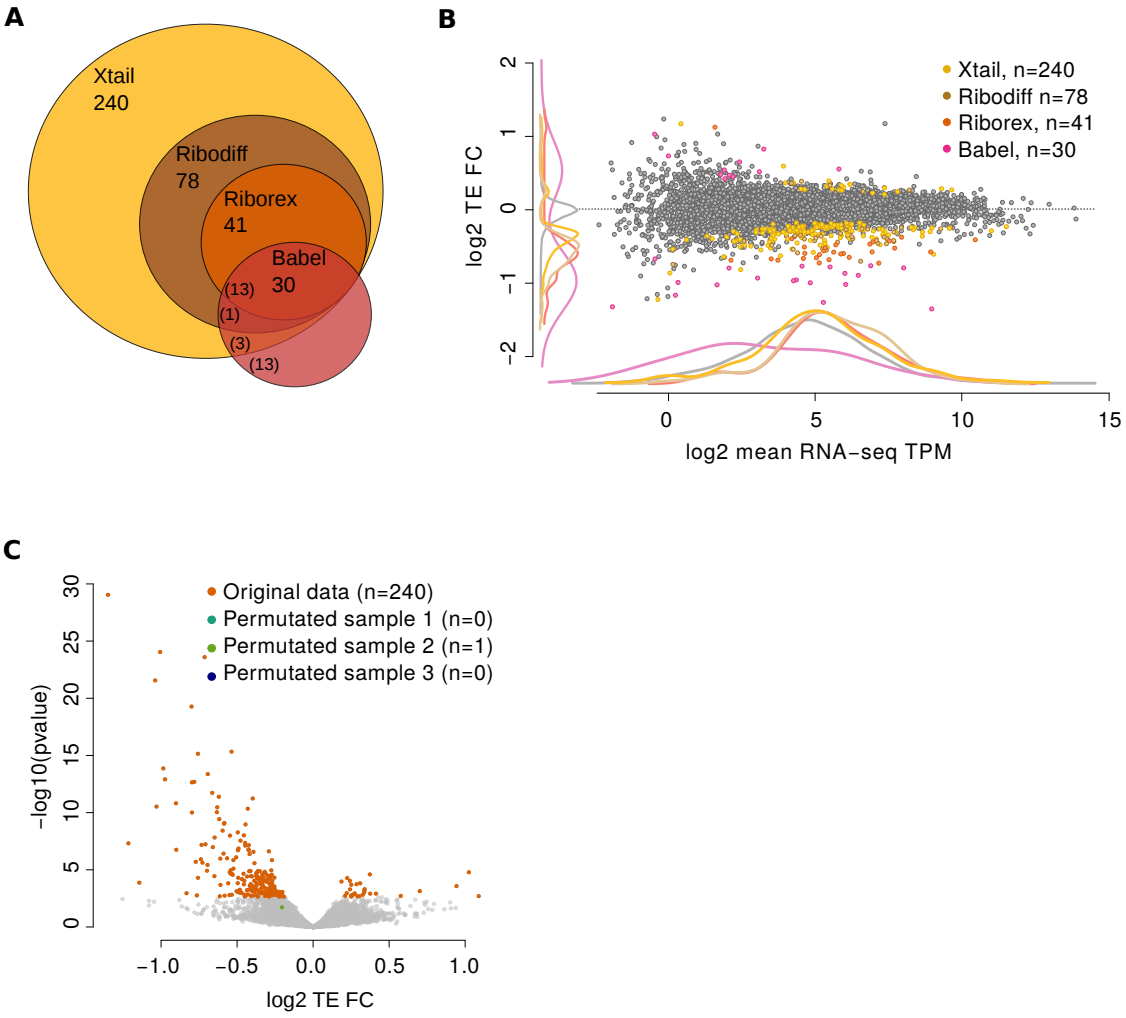


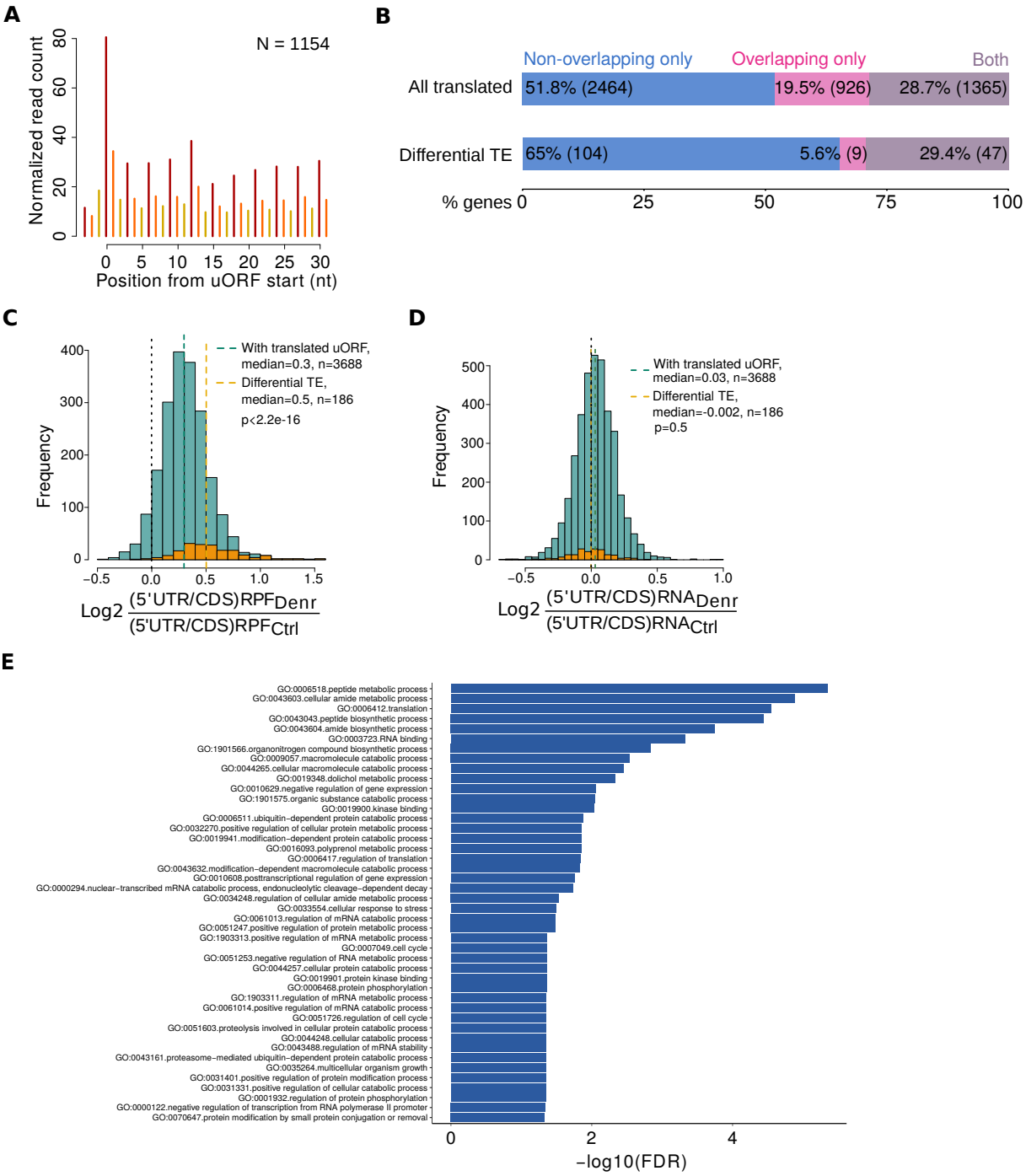
E

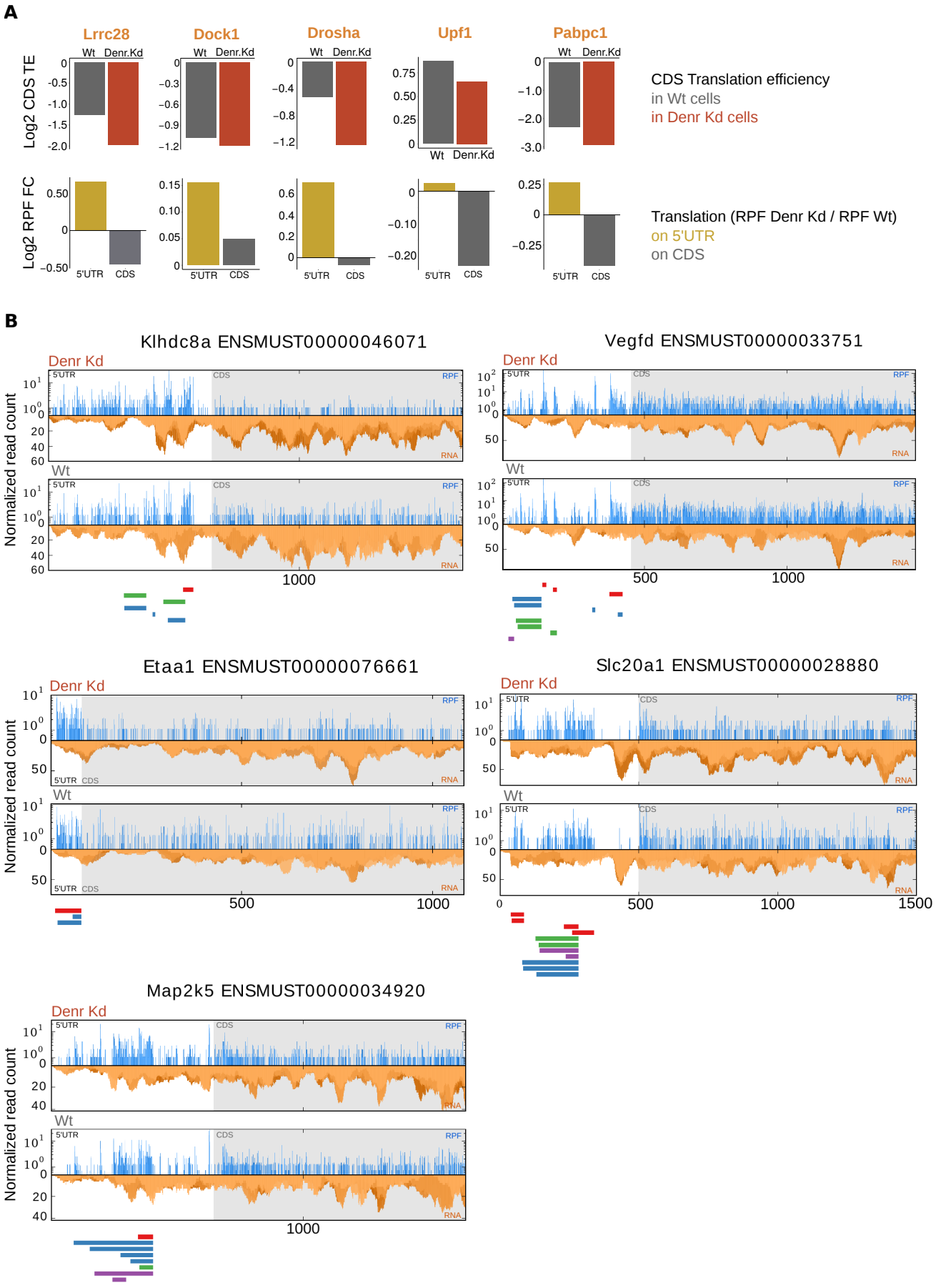


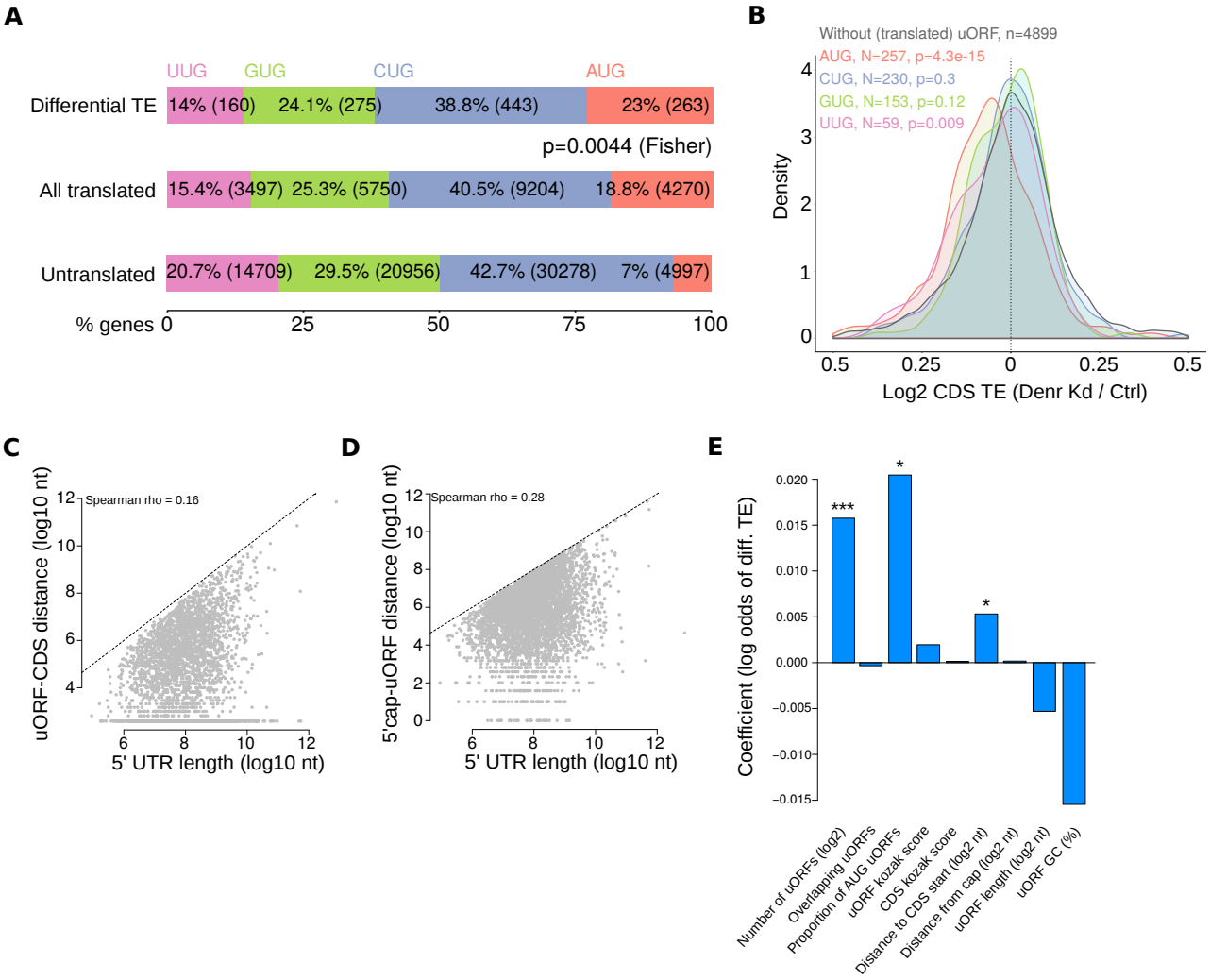
A

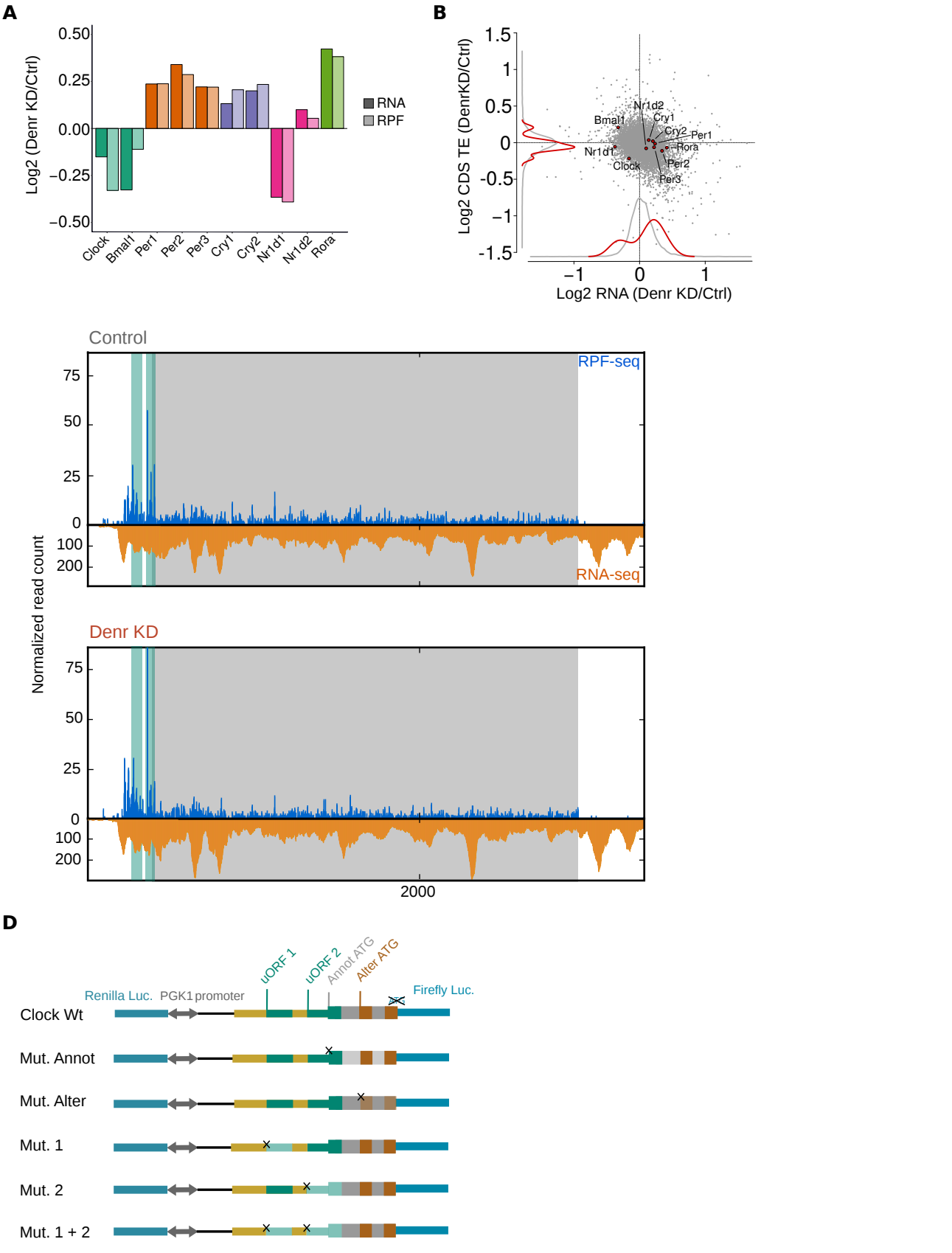

500
400
300
200
100
0
Number of genes
All, n = 5148
Median = 0.32
FDR < 0.1, n = 493
Median=0.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Log2 (5'UTR/CDS)RPFDenr / (5'UTR/CDS)RPFCtrl

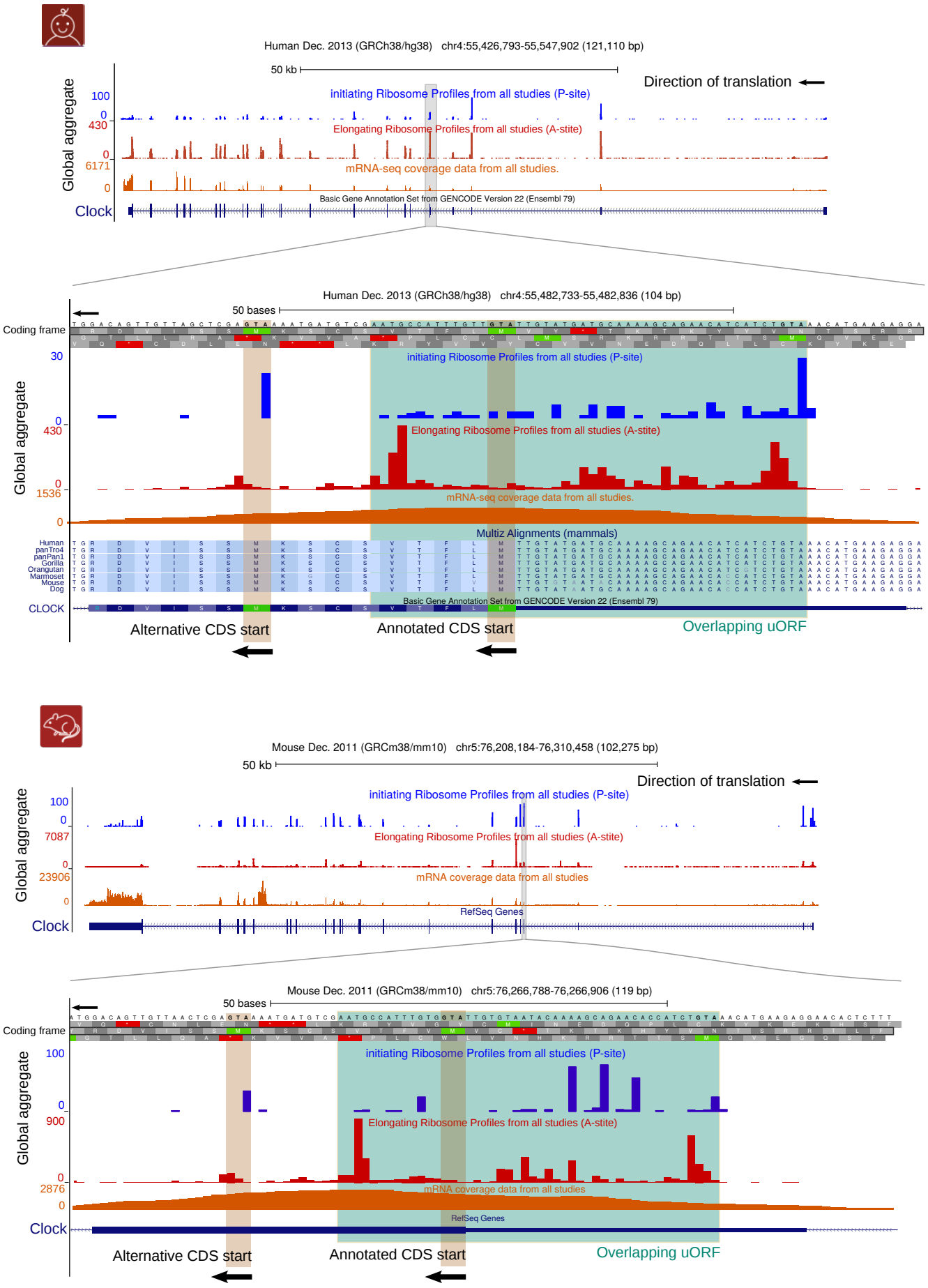

A
Xtail
240
Ribodiff
78
Riborex
41
Babel
30
(13)
(1)
(3)
(13)
B
Xtail, n=240
Ribodiff n=78
Riborex, n=41
Babel, n=30
log2 TE FC
log2 mean RNA−seq TPM
C
Original data (n=240)
Permutated sample 1 (n=0)
Permutated sample 2 (n=1)
Permutated sample 3 (n=0)
−log10(pvalue)
log2 TE FC


A
N = 1154
Normalized read count
Position from uORF start (nt)
B
Non-overlapping only
Overlapping only
Both
All translated
51.8% (2464)
19.5% (926)
28.7% (1365)
Differential TE
65% (104)
5.6% (9)
29.4% (47)
% genes
C
With translated uORF, median=0.3, n=3688
Differential TE, median=0.5, n=186
p<2.2e-16
Frequency
Log2 (5'UTR/CDS)RPFDenr / (5'UTR/CDS)RPFCtrl
D
With translated uORF, median=0.03, n=3688
Differential TE, median=-0.002, n=186
p=0.5
Frequency
Log2 (5'UTR/CDS)RNADenr / (5'UTR/CDS)RNACtrl
E
GO:0006518.peptide metabolic process
GO:0043603.cellular amide metabolic process
GO:0006412.translation
GO:0043043.peptide biosynthetic process
GO:0043604.amide biosynthetic process
GO:0003723.RNA binding
GO:1901566.organonitrogen compound biosynthetic process
GO:0009057.macromolecule catabolic process
GO:0044265.cellular macromolecule catabolic process
GO:0019348.dolichol metabolic process
GO:0010629.negative regulation of gene expression
GO:1901575.organic substance catabolic process
GO:0019900.kinase binding
GO:0006511.ubiquitin-dependent protein catabolic process
GO:0032270.positive regulation of cellular protein metabolic process
GO:0019941.modification-dependent protein catabolic process
GO:0016093.polyprenol metabolic process
GO:0006417.regulation of translation
GO:0043632.modification-dependent macromolecule catabolic process
GO:0010608.posttranscriptional regulation of gene expression
GO:0000294.nuclear-transcribed mRNA catabolic process, endonucleolytic cleavage-dependent decay
GO:0034248.regulation of cellular amide metabolic process
GO:0033554.cellular response to stress
GO:0061013.regulation of mRNA catabolic process
GO:0051247.positive regulation of protein metabolic process
GO:1903313.positive regulation of mRNA metabolic process
GO:0007049.cell cycle
GO:0051253.negative regulation of RNA metabolic process
GO:0044257.cellular protein catabolic process
GO:0019901.protein kinase binding
GO:0006468.protein phosphorylation
GO:1903311.regulation of mRNA metabolic process
GO:0061014.positive regulation of mRNA catabolic process
GO:0051726.regulation of cell cycle
GO:0051603.proteolysis involved in cellular protein catabolic process
GO:0044248.cellular catabolic process
GO:0043488.regulation of mRNA stability
GO:0043161.proteasome-mediated ubiquitin-dependent protein catabolic process
GO:0035264.multicellular organism growth
GO:0031401.positive regulation of protein modification process
GO:0031331.positive regulation of cellular catabolic process
GO:0001932.regulation of protein phosphorylation
GO:0000122.negative regulation of transcription from RNA polymerase II promoter
GO:0070647.protein modification by small protein conjugation or removal
−log10(FDR)


A
Lrrc28
Dock1
Drosha
Upf1
Pabpc1
Wt
Denr.Kd
Log2 CDS TE
Log2 RPF FC
5'UTR
CDS
CDS Translation efficiency
in Wt cells
in Denr Kd cells
Translation (RPF Denr Kd / RPF Wt)
on 5'UTR
on CDS
B
Klhdc8a ENSMUST00000046071
Vegfd ENSMUST00000033751
Etaa1 ENSMUST00000076661
Slc20a1 ENSMUST00000028880
Map2k5 ENSMUST00000034920
Denr Kd
Wt
RPF
RNA
Normalized read count


A
UUG
GUG
CUG
AUG
Differential TE
14% (160)
24.1% (275)
38.8% (443)
23% (263)
p=0.0044 (Fisher)
All translated
15.4% (3497)
25.3% (5750)
40.5% (9204)
18.8% (4270)
Untranslated
20.7% (14709)
29.5% (20956)
42.7% (30278)
7% (4997)
% genes 0
25
50
75
100
B
Without (translated) uORF, n=4899
AUG, N=257, p=4.3e-15
CUG, N=230, p=0.3
GUG, N=153, p=0.12
UUG, N=59, p=0.009
Density
Log2 CDS TE (Denr Kd / Ctrl)
C
Spearman rho = 0.16
uORF-CDS distance (log10 nt)
5' UTR length (log10 nt)
D
Spearman rho = 0.28
5'cap-uORF distance (log10 nt)
5' UTR length (log10 nt)
E
Coefficient (log odds of diff. TE)
Number of uORFs (log2)
Overlapping uORFs
Proportion of AUG uORFs
uORF kozak score
CDS kozak score
Distance to CDS start (log2 nt)
Distance from cap (log2 nt)
uORF length (log2 nt)
uORF GC (%)


A
Log2 (Denr KD/Ctrl)
0.50
0.25
0.00
−0.25
−0.50
RNA
RPF
Clock
Bmal1
Per1
Per2
Per3
Cry1
Cry2
Nr1d1
Nr1d2
Rora
B
Log2 CDS TE (DenrKD/Ctrl)
Log2 RNA (Denr KD/Ctrl)
Nr1d2
Cry1
Bmal1
Cry2
Per1
Rora
Nr1d1
Clock
Per2
Per3
Control
RPF-seq
RNA-seq
Denr KD
Normalized read count
2000
D
uORF 1
uORF 2
Annot ATG
Alter ATG
Renilla Luc.
PGK1 promoter
Firefly Luc.
Clock Wt
Mut. Annot
Mut. Alter
Mut. 1
Mut. 2
Mut. 1 + 2


Human Dec. 2013 (GRCh38/hg38) chr4:55,426,793-55,547,902 (121,110 bp)
50 kb
Direction of translation
Global aggregate
initiating Ribosome Profiles from all studies (P-site)
Elongating Ribosome Profiles from all studies (A-site)
mRNA-seq coverage data from all studies.
Basic Gene Annotation Set from GENCODE Version 22 (Ensembl 79)
Clock
Human Dec. 2013 (GRCh38/hg38) chr4:55,482,733-55,482,836 (104 bp)
50 bases
Coding frame
Multiz Alignments (mammals)
Human
panTro4
panPan1
Gorilla
Orangutan
Marmoset
Mouse
Dog
CLOCK
Alternative CDS start
Annotated CDS start
Overlapping uORF
Mouse Dec. 2011 (GRCm38/mm10) chr5:76,208,184-76,310,458 (102,275 bp)
mRNA coverage data from all studies
RefSeq Genes
Mouse Dec. 2011 (GRCm38/mm10) chr5:76,266,788-76,266,906 (119 bp)

Supplementary figure legends

Supplementary Figure 1. Ribosome profiling in *Denr* knockdown NIH3T3 mouse fibroblasts. **A)** Read length distribution for ribosome footprint (PPF-seq) and total mRNA (RNA-seq). Footprint length peak was 29-30 nt, whereas RNA-seq reads showed a broader distribution, as expected. **B)** Read distribution to transcript features. RPF-seq and RNA-seq compared to a distribution expected from the relative feature size (gray). In both Denr-Kd and control cells, RPF-seq but not RNA-seq reads were depleted from 3' UTRs and enriched on the CDS. A significant amount of reads also mapped to 5' UTRs. **C)** Metagene analysis of ribosome's A-site position relative to the start and stop positions. Read density at each position was average across genes with a single protein-coding isoform, with a CDS RPF RPKM > 5 and longer than 400 nt ($n = 3141$). This analysis revealed the trinucleotide periodicity of RPF-seq, but not RNA-seq, reads. **D)** Tables of Spearman correlation coefficients between samples for normalized CDS reads. On the right, one representative example of normalised CDS reads correlation between replicate samples. **E)** RPKM values for CDS-mapping RPF-seq read counts for *Denr* (left), *Mcts1* (middle), and *Eif2d* (right). Each dot correspond to a sample and replicate and error bars indicate mean$\pm$sd.

Supplementary Figure 2. Analyies of significant relative 5' UTR translation. **A)** Histogram of 5' UTR-to-CDS translation (from RPF-seq) between *Denr* Kd and control cells, for all genes as in Figure 1D (gray). In blue, same distribution for genes with statistically significant change between conditions, as assessed from pairwise comparisons (Wilcox test, FDR < 0.1). Although a global ribosome footprint redistribution to 5' UTRs upon *Denr* knockdown is observed from the analysed genes, the difference was significant for 493 genes, indicating that DENR affects the translational landscape of a subset of genes.

Supplementary Figure 3. Comparison of algorithms for differential translation efficiency analysis. **A)** Venn diagram of differential TE genes detected by Xtail (n=240), RiboDiff (n=78), Riborex (n=41) or Babel (n=30) with an FDR < 0.1. The analysis showed an accordance among the first three algorithms, with Xtail detecting the largest number of significant cases. **B)** Scatterplot with marginal densities of average mRNA abundance vs. TE fold change (*Denr* Kd/Wt), with differential TE genes highlighted for each algorithm. For genes detected by several algorithms, the data-point is coloured according to the method detecting less genes; however density lines correspond to the distribution of all detected cases. The majority of Xtail, RiboDiff and Riborex detected genes showed a decreased TE in *Denr* deficient cells as expected, and spanned all levels of expression, whereas Babel detected similar number of up- and downregulated TE genes and showed a bias for lowly expressed genes. **C)** Volcano plot of differential TE genes detected by Xtail on the original data (n=240) and on three condition-shuffled datasets. Sample permutation yielded nearly no significant genes, indicating no/low false positive detection by Xtail.

Supplementary figure 4. Analysis of differential TE genes. **A)** Proportion of translated uORF-containing genes with non-overlapping uORFs only, overlapping uORFs only, or both, transcriptome-wide and among DENR targets. The latter showed a depletion of overlapping uORFs and enrichment for non-overlapping uORFs only-containing genes. **B)** Triplet periodicity of ribosome footprints within translated uORFs. Genes containing a single uORF were included in the analysis (n = 1154 genes/uORFs) to avoid uORFs overlapping to each other. **C)** Histogram of 5'UTR-to-CDS translation between *Denr* Kd and control cells, for all uORF-containing genes (gray, n=3688) and for differential TE genes (orange, n=186). Genes with a minimum of 20 RPF-seq reads in the 5'UTR in 3/4 of samples were included for the analysis. **D)** Same as in (C) for RNA-seq reads. Differential TE genes showed an increased relative 5'UTR translation, but equal transcript levels, in *Denr*-deficient cells, that was higher than for the whole set of uORF-containing genes, suggestive of impaired translation reinitiation among targets. **E)** Gene Ontology (GO) analysis for Biological process, Metabolic function and Reactome functional categories of genes with differential TE. Bars show -log10 for categories with an FDR $<$ 0.05.

Supplementary figure 5. Examples of differential TE genes. **A)** Representative examples of genes with differential TE. Top panels show average CDS translation efficiency in Wt (gray) and *Denr* Kd (red) cells, and bottom panels show *Denr* Kd-to-Wt translation on the 5'UTR (brown) and on the CDS (gray). DENR targets showed a decreased CDS TE and an increased 5'UTR translation concomitant with a reduced CDS translation in *Denr*-deficient cells. **B)** Normalized RPF (blue) and RNA (orange) read count distribution on several DENR targets, in knockdown (top) and control (bottom) cells. Different blue and orange shadings correspond to each replicate per condition. The first 1000 nt of the CDS are shown. Coloured boxes below profiles indicate translated uORFs (red: AUG, blue: CUG, green: GUG, pink: UUG).

Supplementary figure 6. Regression analyses of TE regulation by DENR. **A)** Proportion of AUG, CUG, GUG, and UUG uORFs detected as translated, untranslated and among differential TE genes. DENR targets are enriched for translated AUG uORFs (p = 0.0044, Fisher test), indicating start codon specificity among DENR-regulated transcripts. **B)** Density plot of the CDS TE ratio (*Denr* Kd/Wt) for genes containing only AUG (n = 257), only CUG (n = 230), only GUG (n = 153), or only UUG (n = 59) uORFs. P-values indicate wilcoxon test results for the difference between each distribution and that of genes without uORFs (or with untranslated uORFs, n = 4899). Although the number of genes per category is low, this analysis showed that AUG-containing genes have the strongest CDS TE regulatory effect upon *Denr* knockdown. **C)** Scatterplot of 5' UTR length vs. the distance from the last uORF stop codon to the CDS start for all uORF-containing genes used for the regression model. No strong correlation was quantified between the two variables, ruling out that the correlation of the intercistronic distance with the magnitude of DENR regulation was due to longer 5'

UTRs. **D)** Scatterplot of 5' UTR length vs. the distance from the 5' cap to the first uORF start codon for all uORF-containing genes used for the regression model. No strong correlation was quantified between the two variables, but correlation was higher than in (C). **E)** Coefficients for the logistic regression model, with "detected as differential TE, TRUE/FALSE" as binary response. Coefficients thus indicate the log odds change for being detected as differential TE (by Xtail) per unit change in the predictor. Same genes as for figure 3 were used. Asterisks indicate coefficient estimates significance as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

Supplementary figure 7. *Clock* as DENR target. **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)** Scatterplot of *Denr* Kd-to-control transcript levels (RNA-seq) vs. CDS translation efficiency, with circadian core clock genes highlighted. **C)** Distribution of normalised RPF-seq (blue) and RNA-seq (orange) reads along the *Clock* transcript, in control (top) and *Denr*-Kd (bottom) cells. Green shaded area indicates the AUG-initiated uORFs; gray area corresponds to the annotated CDS. Only the first 500 nt of the 3' UTR are shown for better visualisation. **D)** Schematic representation of the *Clock Wt* and mutants tested for effect on CLOCK synthesis and DENR dependance. *Clock Wt* contains the whole 5' UTR and first 14 aminoacids of the CDS. All other mutants tested were combinations of these.

Supplementary figure 8. *Clock* ribosome profiling data from published studies. All available data from initiating and elongating ribosome profiling studies, gathered from GWIPS-viz [Michel et al., 2014]. An overview of the whole Clock transcript and a zoom over the CDS start is shown, where reads mapping on the overlapping uORF and annotated and alternative CDS starts are visualised. Both human (top) and mouse (bottom) data are displayed. Initiating ribosomes can be seen at the start of the uORF and alternative CDS start, as well as good uORF coverage from elongating ribosomes.