

Transcriptome-wide sites of collided ribosomes reveal principles of translational pausing

Alaaddin Bulak Arpat^{a,b}, Angélica Liechti^a, Mara De Matos^a, René Dreos^a,
Peggy Janich^{a,c}, David Gatfield^{a,*}

^aCenter for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland

^bVital-IT, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

^cpresent address: Krebsliga Schweiz, 3001 Bern, Switzerland

Summary

Translation initiation is considered overall rate-limiting for protein biosynthesis, whereas the impact of non-uniform ribosomal elongation rates is largely unknown. Using a modified ribosome profiling protocol based on footprints from two closely packed ribosomes (disomes), we have mapped ribosomal collisions transcriptome-wide in mouse liver. We uncover that the stacking of an elongating onto a paused ribosome occurs frequently and scales with translation rate, trapping ~10% of translating ribosomes in the disome state. A distinct class of pause sites, independent of translation rate, is indicative of deterministic pausing signals. Pause sites are associated with specific amino acids, peptide motifs, and with structural features of the nascent polypeptide, suggestive of programmed pausing as a widespread mechanism associated with protein folding. Evolutionary conservation at disome sites and experiments indicate functional relevance of translational pausing. Collectively, our disome profiling approach allows novel, unexpected insights into gene regulation occurring at the step of translation elongation.

Keywords: disome, ribosome profiling, translation, ribosome stalling, ribosome collisions, footprint, elongation, protein structure

*Lead Contact:

Email address: david.gatfield@unil.ch (David Gatfield)

Introduction

The translation of messenger RNA (mRNA) to protein is a central step in gene expression. Our knowledge of this process has exploded over the past decade due to the emergence of ribosome profiling (Ingolia et al., 2009), a technique based on the high-throughput sequencing of the ~ 30 nt mRNA footprints that are buried inside the translating ribosome and thus protected from the nuclease treatment that is used to digest the mRNA regions that are not occupied by ribosomes. A plethora of studies have built on the quantitative, transcriptome-wide and nucleotide-resolved information that ribosome profiling provides to gain insight into a variety of aspects of protein biosynthesis (see Ingolia et al. (2019) for a recent review). This includes, among others, the annotation of translated mRNA regions, the study of differential translation across various biological and experimental paradigms, the characterization of intermediate states of the translating ribosome, the subcellular compartmentalization of protein biosynthesis, or functional differences in translational capacity within a heterogeneous cellular ribosome population.

Most available ribosome profiling data is consistent with the longstanding notion that of the four distinct phases defining translation, i.e. initiation, elongation, termination and ribosome recycling, it is the first step – the commitment of the ribosome to initiate – that is rate-limiting for the overall process in eukaryotes (Hinnebusch, 2014). It is thus assumed that the quantity of elongating ribosome footprints (i.e., the species mainly captured by conventional ribosome profiling methodology) is proportional to initiation rate and to overall protein biosynthesis. Elongating ribosome footprints typically distribute in a distinctly non-uniform fashion across a given protein coding sequence (CDS), which has been attributed to variations in ribosome decoding speed and dwell times (Ingolia et al., 2011). Integrating footprint reads across the entire CDS is thought to correct for local variation in footprint density and to allow for accurate estimates of relative translation efficiencies per gene (TEs, calculated as CDS-mapping footprint reads normalized to RNA abundance) in a transcriptome-wide fashion.

Nevertheless, a possible influence of local footprint variation on overall translation speed of an mRNA has been suggested early on (Dana and Tuller, 2012) and, in general, how to interpret apparent local differences in footprint densities is not fully resolved. Of note, it remains an intrinsic limit of the technique
35 that it delivers static snap-shots of ribosome occupancy rather than dynamic data of the translation process. Therefore, and somewhat paradoxically, in the two extreme, hypothetical scenarios of one transcript whose elongating ribosomes are translationally paused (resulting in low or no protein biosynthesis), and of another transcript that shows strong, productive flux of elongating ribo-
40 somes (resulting in high protein biosynthesis), the actual footprint snap-shots that would be seen in ribosome profiling may actually be indistinguishable. To discern such cases, a dedicated genome-wide method for the direct detection of ribosomal pausing would be crucial; in yeast, specific footprint size classes associated with stalled ribosomes have been described (Guydosh and Green,
45 2014).

Historically, early evidence for paused elongation – leading to the subsequent stacking of upstream elongating ribosomes onto the paused one – has come from *in vitro* translation reactions (Wolin and Walter, 1988). For a limited number of prominent cases, pausing has since then been shown to be function-
50 ally important for proper protein localization to membranes (Yanagitani et al., 2011; Mariappan et al., 2010), to serve as a mechanism for start codon selection (Ivanov et al., 2018), and to regulate the extent of productive full-length protein biosynthesis (Yordanova et al., 2018). It is tempting to extrapolate from such individual examples to general roles for elongation pausing that cells could
55 employ to control protein biosynthesis post-initiation. At the other end of the spectrum, hard elongation stalls caused by various obstacles to processive translation (including defective mRNAs or specific amino acid motifs in the nascent peptide) require resolution by the ribosome-associated quality control pathway (RQC), and the mechanisms through which such terminally stalled ribosomes
60 are sensed and handled is a highly active field of current research (reviewed in Joazeiro (2019)).

An early ribosome profiling study in mouse embryonic stem cells (mESCs) already addressed the question of how to extract potential pause sites from footprint data (Ingolia et al., 2011), which resulted in the identification of thousands
65 of alleged pauses in the body of genes as well as an enrichment at termination codons. In combination with quantitative modelling approaches, subsequent studies have identified parameters that can impinge on local translation speed and pausing (reviewed in Schuller and Green (2018)). Among these are, notably, specific amino acids (Charneski and Hurst, 2013), codon pairs (Gamble
70 et al., 2016), tRNA availability (Darnell et al., 2018; Guydosh and Green, 2014), RNA secondary structures (Zhang et al., 2017; Pop et al., 2014), or the folding (Doring et al., 2017) and exit tunnel interactions (Dao Duc and Song, 2018; Charneski and Hurst, 2013) of the nascent peptide. However, to what extent translational pausing occurs *in vivo* in a mammalian system, which character-
75 istics these pause sites have, and whether they are functionally relevant is still poorly understood.

Here, we have applied a modified ribosome profiling strategy to a mammalian organ, mouse liver, in order to directly reveal in a transcriptome-wide fashion the sites where two ribosomes collide. The characteristics associated with these
80 ~ 60 nt "disome footprints" are consistent with the expectations for collision events of an upstream elongating ribosome onto a downstream, paused ribosome. Through the use of synthetic footprint spike-ins, we estimated the quantitative relationship between disome and monosome footprints. Deep analysis of the disome sites allowed identifying features predictive of ribosome pausing,
85 including sequence features of the mRNA and structural features of the nascent polypeptide. Finally, we address the question of the functional relevance of pausing events via the analysis of evolutionary conservation and through an experimental approach.

Results

90 *Disome footprint sequencing allows transcriptome-wide mapping of ribosomal collisions*

A critical step in ribosome profiling is the quantitative, nuclease-mediated conversion of polysomes to individual monosomes, from which protected mRNA footprints of ~ 30 nt can be purified and converted to sequenceable libraries. During the setup of this technique from mouse liver polysomal extracts for a previous study (Janich et al., 2015), we used northern blots to monitor the efficiency of RNase I-mediated footprint generation. Radioactively labelled short oligonucleotide probes antisense to the protein coding sequences (CDS) of the highly abundant *Albumin (Alb)* and *Major urinary protein 7 (Mup7)* mRNAs indeed 100 showed the expected ~ 30 nt monosome footprints (Figure 1A, B). Moreover, several of the probes also detected distinct higher-order bands whose estimated sizes were compatible with those expected for multiples of monosome footprints (i.e. ~ 60 nt, ~ 90 nt etc.). These additional bands were particularly prominent for the probes designed to anneal to the transcripts just downstream of where the signal peptide (SP) was encoded (see monosome footprint and two higher-order bands for probe Alb_{71–101} in Figure 1A; see monosome footprint and >5 higher-order bands for probe Mup7_{58–81} in Figure 1B). We initially interpreted the presence of these bands as an indication of suboptimal conditions during nuclease treatment, leading to an incomplete collapse of polysomes to monosomes. We hence tested other nuclease digestion conditions. However, neither 110 changes in temperature or detergent concentrations during extract preparation and nuclease treatment (Supplemental Figure S1A), nor higher RNase I activity (Supplemental Figure S1B), nor a different nuclease altogether, micrococcal nuclease (Supplemental Figure S1C), were able to collapse the higher-order bands 115 quantitatively to the size of the monosome footprint. We thus speculated that the higher-order footprints reflected a distinct, relatively stable state of translating ribosomes, possibly resulting from two (disome), three (trisome) or, in the case of the bands seen for Mup7_{58–81}, even higher numbers of ribosomes whose

dense stacking rendered the mRNA inaccessible to nucleases. This scenario was
120 reminiscent of the ribosomal pausing and stacking described in the 1980s for *in*
vitro translated preprolactin mRNA (Wolin and Walter, 1988). Here, a major
translation stall site at codon 75 (a GGC glycine codon), which led to the queu-
ing up of subsequent incoming ribosomes, was related to the recruitment of the
signal recognition particle (SRP) to the signal peptide.

125 We wished to determine whether our higher-order footprints reflected a sim-
ilar phenomenon and whether they would allow us to detect ribosomal pause
and collision sites transcriptome-wide and *in vivo*. We thus selected a subset
of samples from our previously collected mouse liver time series (Janich et al.,
2015), corresponding to three timepoints at the beginning of the daily light
130 (*Zeitgeber* Times ZT0 and ZT2) and dark phases (ZT12), and subjected them
to ribosome profiling for both the ~30 nt monosome footprints and the ~60 nt
alleged disome footprints; we also determined RNA abundances from the same
samples by RNA-seq (Figure 1C). Libraries were sequenced sufficiently deeply
to obtain $>10^8$ cDNA-mapping reads per footprint species (Figure 1D, Supple-
135 mental Figure S2A and Supplemental Table S1). Monosome footprints showed
the expected length and mapping features, i.e. the majority was 29-30 nt in size
(Figure 1E) and they were enriched on CDS and depleted from untranslated
regions (UTRs) (Supplemental Figure S2B). This depletion was considerably
stronger for 3' UTRs than for 5' UTRs, which was expected given that 5' UTRs
140 are known to harbour significant translational activity on upstream open reading
frames (uORFs). Disome footprint lengths showed a broader distribution with
two distinct populations at 59-60 nt and 62-63 nt (Figure 1E) that resembled the
bimodal pattern that has been seen in yeast as well (Guydosh and Green, 2014).
The mapping to transcript regions was similar to that of monosome footprints,
145 albeit with a more pronounced depletion from 5' UTRs (Supplemental Figure
S2B). As the median uORF length in mice is <40 nt (Johnstone et al., 2016),
it is likely that many uORFs are simply too short to accommodate two trans-
lating ribosomes simultaneously. Reduced levels in 5' UTR disome footprints
were thus fully compatible with the hypothesis that they reflected ribosomal

150 collisions.

We next analyzed frame preference and distribution along the CDS for the two footprint species. To this end, we mapped the predicted position of the ribosomal aminacyl-tRNA acceptor site (A-site) codon of each monosome footprint (i.e. nucleotides 15-17 for 29-30 nt footprints, see Janich et al. (2015)) onto the meta-transcriptome (N=5456 genes). Transcriptome-wide, we observed the characteristic 3-nt periodicity of ribosome footprints across coding sequences, starting at the +1 codon relative to the initiation site (note that initiating ribosomes carry the first tRNA already in their P-site, and the A-site is thus placed over the +1 codon) and ending at the termination codon (Figure 1F). Moreover, the profile showed previously reported features, including elevated ribosome density at the start and at the +5 codon (which has been interpreted to reflect a pause occurring between initiation and elongation commitment, see Han et al. (2014)), and an underrepresentation of stop codon footprints (which is a known consequence of the elongation inhibitor cycloheximide that was added during polysomal extract preparation, see Ingolia et al. (2011)).

The equivalent analysis for the disome footprints showed transcriptome-wide 3-nt periodicity as well, albeit with two noteworthy peculiarities. First, at the 3' end of the CDS, disome coverage ended 10 codons upstream of monosome coverage, precisely at the position expected when the leading ribosome of two stacked ribosomes would occupy the termination codon. Second, at the 5' end, disome occurrence was low at and immediately after the initiation site, highest on the +5 codon, and then overall rather uniform (though showing some 5'-to-3' decrease) over the remainder of the CDS. Aligning the disome footprints relative to the predicted A-site of the leading, rather than the stacked ribosome, gave fully compatible results (Supplemental Figure S3A). Taken together, these findings were consistent with the hypothesis that the ~60 nt higher-order bands indeed represented footprints originating from translated mRNA that was protected by two adjacent, and possibly collided, ribosomes. Moreover, it would appear that transcriptome-wide these alleged collision events could occur at most CDS positions, although the likelihood of stacking onto a downstream ri-

bosome would seem reduced for the first few codons, possibly due to sterical constraints to initiation (Guydosh and Green, 2014).

Disome occurrence is locally favored by signal peptides, and globally by high translation efficiency

185 The SRP-dependent ribosome pausing and stacking described for prepro-
lactin mRNA (Wolin and Walter, 1988) suggested that transcripts encoding
signal peptide-containing proteins (termed SP transcripts in the following) could
serve as a positive control to validate that our approach was indeed capturing
similar events. We thus assessed footprint densities along the CDS for SP tran-
190 scripts (N=713) vs. non-SP transcripts (N=4743). SP transcripts showed a
striking build-up of disome footprints towards the 5' end of the CDS that ex-
tended virtually to the position (~ 75 amino acids) Wolin & Walter had described
for preprolactin mRNA, whereas downstream of this region disome read den-
sities were reduced (Figure 2A-B and Supplemental Figure S3B). Importantly,
195 these features were specific to disomes, as they were absent from monosome
footprint data (Figure 2C-D). We concluded that the profiling of disome foot-
prints was able to capture the same translational pausing and stacking events
that had been previously described (Wolin and Walter, 1988).

We next analyzed monosome and disome footprints per gene. Of note, many
200 transcripts that were detectable at the monosome footprint and RNA-seq levels
also showed sufficient coverage in the disome data, allowing for robust quanti-
fication of both footprint species across a large portion of the expressed genome
(N=8626 genes). We first computed for each gene individually the ratio of
CDS-mapping footprint to RNA-seq reads. For monosome footprints, this ra-
205 tio is frequently referred to as "ribosome density" and considered to reflect a
transcript's relative translational efficiency (TE). In an analogous fashion, for
the disome footprints this density would correspond to a measure for the extent
of ribosomal pausing and stacking. When comparing disome and monosome
footprint densities per gene, we made two main observations. First, disome
210 densities were positively correlated with translation efficiencies (Figure 2E). SP

transcripts showed this correlation as well; however, they were globally shifted to lower disome footprint levels, indicating that the high disome occurrence up to codon ~ 75 was outweighed by the reduction seen over the remainder of the CDS (Figure 2B). Second, the steepness of the fit in the double-log plot in Figure 2E was significantly greater than 1, indicating an exponential relationship between disome and monosome densities. Conceivably, increased ribosomal flux on mRNAs was thus associated with an even higher relative increase in ribosomal collisions. We next assessed whether this relationship between disome and monosome footprint levels was not only observable across different transcripts, but also for a given transcript at different translation efficiencies. To this end, we analyzed mRNAs encoding for ribosomal proteins (RPs), which show prominent, feeding-dependent daily rhythms in TE (Sinturel et al., 2017). Using the two timepoints from our datasets that corresponded to states of low (ZT2) and high (ZT12) RP mRNA translation (Janich et al., 2015), our analyses revealed that the increase in disome density on RP transcripts was significantly greater than the ~ 2 -fold increase in monosome TE between the two timepoints (Supplemental Figure S3C). This outcome is consistent with the exponential relationship between disome and monosome footprints detected transcriptome-wide.

These findings suggested that – at least in part – disome footprints were a consequence of ribosomal crowding that could in principle occur during any translation event, but that was favoured under high ribosomal traffic. In such a case, one might hypothesize that the positions on the CDS where ribosomes collided could have a sizable stochastic component; in addition, local differences in ribosomal dwell times – which are associated with amino acid/codon usage and the size of the amino acid-loaded tRNA pool (Gobet et al., 2019) – would be expected to bias the collision sites as well. In contrast, however, the observation that signal peptides appeared to act as general triggers for ribosome stalling and queuing, as well as differences in disome levels across transcripts that were not simply attributable to TE differences (Supplemental Figure S3D), suggested that beyond the alleged "stochastic" sites, more specific, "deterministic" signals

and stall sites may exist, too.

In order to gain a sense of whether these different scenarios and distinct classes of disome sites (i.e. stochastic vs. deterministic) truly existed, we first inspected various transcript examples. To begin with, we noted that individual SP transcripts indeed exhibited the expected disome patterns. As shown for the case of *Adhesion G protein-coupled receptor G3 (Adgrg3)* – whose annotated signal peptide spans amino acids (aa) 1-18 – disome footprint coverage was elevated in the region up to codon ~ 75 and was lower and appeared more randomly distributed over the remainder of the transcript (Figure 2F). Similarly, *Transferrin receptor (Tfrc)*, which contains an SRP-dependent signal anchor (SA) sequence at aa 68-88 (Zerial et al., 1986), showed elevated disome levels extending until codon ~ 145 (Figure 2G), indicating a direct relationship between the positions of disome buildup and of the signal sequence/SRP interaction. We next examined individual transcripts lacking signal sequences for the presence of the alleged stochastic and deterministic sites. For example, the transcripts coding for two subunits of the 19S regulatory particle of the proteasome, *Psm α 4* (Figure 2H) and *Psm α 5* (Figure 2I), showed distinct patterns of disome distribution that were consistent with our expectations for stochastic and deterministic sites, respectively. *Psm α 4* thus showed disome coverage at numerous positions along the CDS. By contrast, a specific, dominant site was apparent for *Psm α 5*. Many other transcripts showed such patterns with distinct dominant sites as well, e.g. *Aldehyde dehydrogenase 1a1 (Aldh1a1)* (Figure 2L), *Pyruvate kinase liver and red blood cell (Pklr)* (Figure 2N) and *Eukaryotic translation initiation factor 5A (Eif5a)* (Figure 2P). Dispersed disome patterns similar to *Psm α 4*, as well as mixed cases combining broad coverage with specific dominant sites were frequent, too, e.g. *Aldolase A (AldoA)* (Figure 2K), *Acyl-Coenzyme A oxidase 3 (Acox3)* (Figure 2M) and *Eukaryotic translation initiation factor 2A (Eif2a)* (Figure 2O). Furthermore, we made the empirical observation that in some cases there was not (e.g. *Aldh1a1*) and in others there was (e.g. *Pklr*) a correspondence between the sites of strong disome and monosome accumulation. Indeed, both scenarios – correlation and anti-correlation – between strong disome and

monosome sites appear plausible: On the one hand, extended ribosomal dwell times should lead to the capture of more monosome footprints from slow codons – and since these positions would also represent sites of likely ribosomal collisions, they would be enriched in the disome data as well. On the other hand, however, for sites where collisions are very frequent – to the extent that stacked ribosomes become the rule – one may expect to see an effective loss of these positions in the monosome footprint data.

An important consequence of elongating ribosomes getting trapped in disomes is that conventional (i.e., monosome footprint-based) ribosome profiling datasets will inevitably underestimate the number of translating ribosomes per transcript, in particular for high TE mRNAs. We wished to quantify this effect. This was, however, not possible from our existing monosome and disome footprint datasets because they originated from independent libraries (Figure 1C). Consequently, we had no means of normalizing the disome and monosome data between each other and of aligning them in a way that would have allowed us to determine the quantitative relationship between the two footprint species in the original sample (i.e., global quantitative differences between monosome and disome reads could not be discriminated from technical differences such as unequal sequencing depths). We therefore sequenced new libraries from liver samples to which, early in the protocol, we had added defined quantities of synthetic 30mer and 60mer RNA spike-ins (Supplemental Figure S4A-B), which subsequently allowed for a quantitative realignment of monosome and disome footprint data. This approach revealed that for transcripts with high translation efficiency, typically ~10% of translating ribosomes were trapped in disomes (Figure 2Q). This percentage decreased with decreasing transcript TE and was generally reduced for SP-transcripts, as expected.

In summary, we concluded that disome formation was a common phenomenon and observable across most of the transcriptome. The association with signal peptides and with high translational flux supported the notion that disome footprints were indeed indicative of ribosomal collisions between a downstream, slow decoding event and an upstream ribosome that stacked onto the stalled

ribosome.

305 *Disome sites are associated with specific amino acids and codons*

We next investigated whether disome sites were associated with mRNA sequence features, in particular with specific codons or amino acids. To this end, we adapted a method developed for the analysis of monosome-based footprint data, termed Ribo-seq Unit Step Transformation (RUST), which calculates observed/expected ratios for a given feature at each codon position within a window that encompasses the footprint and surrounding upstream and downstream regions (O'Connor et al., 2016). RUST-based enrichment analyses in O'Connor et al. showed that ribosome footprints had the highest information content (relative entropy, expressed as Kullback–Leibler divergence) on the codons placed within the ribosome decoding center. Moreover, the sequence composition at the 5' and 3' termini of the mRNA fragments showed pronounced non-randomness as well, which was, however, not specific to footprints but found in RNA-seq data as well. It was thus concluded that the main contributing factors to footprint frequency at a given mRNA location were, first, the identity of the codons in the A- and P-sites and, second, the sequence-specificity of the enzymes used for library construction (O'Connor et al., 2016).

To be able to apply the RUST pipeline for the disome footprints, we first needed to investigate the origins of their bimodal length distribution (Figure 1E) and determine which footprint nucleotides likely corresponded to the ribosomal E-, P- and A-sites. To this end, enrichment analyses for codons, conducted individually for the different size classes of disome footprints (Supplemental Figure S5A), resulted in profiles that resembled the previously described RUST-based enrichment profiles for monosome data (O'Connor et al., 2016). Increased information content was thus apparent at the footprint 5' and 3' boundaries, indicative of the aforementioned biases from library construction. Moreover, codon selectivity was consistently seen in the footprint region that would be occupied by the leading ribosome's decoding center, ~15 nt upstream of the footprint 3' end. Notably, no specificity was visible for the region occupied by the upstream

ribosome. These findings were consistent with the idea of the leading ribosome
335 defining the pause site (with preference for specific codons) and the upstream
ribosome colliding in a sequence-independent fashion. Importantly, the com-
parison of the enrichment plots from the different footprint lengths allowed us
to propose a likely interpretation for the observed length heterogeneity. The
two major populations of 59-60 nt and 62-63 nt thus appeared to correspond
340 to ribosome collisions in which the upstream ribosome stacked onto the stalled
ribosome in two distinct states that differed by one codon (Supplemental Fig-
ure S5B). Conceivably, the 1 nt variation (i.e., 59 nt vs. 60 nt and 62 nt vs.
63 nt) corresponded to differences in trimming at the footprint 3' end. Using
this model, we aligned the main populations from the range of disome footprint
345 lengths (i.e., 58-60 nt and 62-63 nt, representing together about two-thirds of
all disome footprints) according to the predicted A-site of the paused, leading
ribosome (see STAR Methods) and used these A-site-corrected footprints for
the RUST pipeline.

Enrichment analysis showed marked selectivity for amino acids in the de-
350 coding center (P- and A-site) of the disome's leading ribosome (Figure 3A,
left panel). The magnitude of amino acid preference was considerably greater
than that seen for the monosome footprints (Figure 3A, middle panel), whereas
RNA-seq data lacked selectivity beyond the effects attributable to library gen-
eration enzymology, as expected (Figure 3A, right panel). Specific amino acids
355 stood out as preferred ribosome stall sites, irrespective of codon usage. The
most striking associations were, notably, the prominent overrepresentation of
aspartic acid (D) in both the P- and A-sites (Figure 3B), the enrichment of
isoleucine (I) in the A-site and its depletion from the P-site (Figure 3C), and
the enrichment of glycine (G) in the P-site (Figure 3D) of stalled ribosomes.
360 We next transformed the full amino acid analysis (Supplemental Figure S6)
to a position weight matrix representing the ensemble of positive and negative
amino acid associations with disome sites (Figure 3E). The specific enrichment
of acidic (D, E) and the depletion of certain basic amino acids (K, H) within the
decoding center of the leading ribosome suggested that amino acid charge may

365 be a relevant factor for ribosomal pausing. Of note, the observed amino acid
signature showed resemblance with ribosomal dwell times that were recently es-
timated through modelling of conventional mouse liver ribosome profiling data
(Gobet et al., 2019). In keeping, the analysis of amino acid associations in our
monosome data showed a similar, though in magnitude much reduced, pattern
370 of amino acid enrichment and depletion as the disomes (Supplemental Figures
S6 and S7A). Finally, we noticed that for certain amino acids, association with
disome sites was dependent on codon usage. For example, P-site Asparagin (N)
was strongly associated with pause sites only when encoded by AAT, but not by
AAC (Figure 3F); Lysine (K) was depleted at P-sites irrespective of the codon,
375 but the association of A-site Lys was highly dependent on codon usage, showing
either depletion (AAA) or enrichment (AAG) (Figure 3G).

We next investigated the association of pause sites with specific amino acid
combinations. Strong selectivity with regard to the 400 possible dipeptide mo-
tifs was apparent in the P- and A-sites of the leading ribosome (Figure 3H, left
380 panel). This effect was much weaker and absent for monosome and RNA data,
respectively (Figure 3H, middle and right panels). In the disome data, the en-
richment was highest, and independent of codon usage, for dipeptides consisting
of the most enriched single amino acids, i.e. Gly-Ile (GI), Asp-Ile (DI) and Gly-
Asp (GD) (Figure 3I-L). By contrast, the pausing of ribosomes at several other
385 dipeptides was strongly dependent on codon usage. In particular the presence of
a Lys (K) or Gly (G) in the A-site of the leading ribosome was associated with
codon selectivity (Figure 3M). For instance, the Asp-Lys (DK) dipeptide was
highly associated with disomes when encoded by GATAAG (Figure 3N; blue
trace); notably, with transcriptome-wide 910 cases of disome peaks observed on
390 the 2030 existing GATAAG positions (i.e. 44.8%), it was the 8th most disome-
prone dicodon out of the total 3721 (i.e., 61 x 61) possible dicodon combinations
(Supplemental Table S2; Figure 3P). By contrast, when encoded by GACAAA
(Figure 3N; black trace), disomes were observable on no more than 7.8% of sites
(272 out of 3529), ranking this dicodon at position 1419. The Gly-Gly (GG)
395 dipeptide represented a similar case (Figure 3O); of the 16 dicodon combina-

tions, GGAGGA (blue trace) was most strongly enriched (698/2407, i.e. 29.0% of sites showed disome peaks; rank 64), whereas GGCGGC (black trace) showed depletion from disome sites (92/1738, i.e. 5.3% of sites showed disome peaks; rank 2304).

400 In summary, the preference for codons, amino acids and amino acid combinations at the predicted P- and A-sites of the leading ribosome suggest that specific sequence signatures are an important contributor to the locations of collision events. Moreover, ribosome pausing that depends on codon usage opens the possibility to modulate the kinetics of translation elongation independently
405 of amino acid sequence.

Disome sites are related to structural features on the polypeptide

It was unlikely that the two factors associated with ribosomal collisions revealed above – i.e. high ribosomal flux (Figure 2) and specific amino acids/codons (Figure 3) – would suffice to provide the specificity required to discriminate
410 between the disome sites that were actually observable (e.g. the alleged "deterministic sites" in Figure 2I, L, N, P), as compared to other positions on the mRNA that were devoid of disome footprints despite similar codon composition. We therefore expected that additional features of the transcript and of the polypeptide would be critical in specifying ribosomal collision sites as well.
415 Our above findings indicated that the signal peptide represented one such element promoting stalling and stacking; in this context, it was interesting to note that even for SP-related stalling, the actual sites on which disomes were observable were in accordance with the generic characteristics of disome occurrence described above. Thus, disome density on SP-transcripts was also dependent on
420 TE (Figure 2Q) and the amino acid preference of disome sites at SP sequences (Supplemental Figure S7B) closely resembled that identified transcriptome-wide (Figure 3E).

In order to identify other protein traits associated with ribosomal stalling, we first assessed the relationship between disome sites and the electrostatic charge
425 of the nascent polypeptide. These analyses revealed, first, a strong association of

negatively charged amino acids with the decoding center of the stalled ribosome (Figure 4A, left panel). This was an expected outcome given the enrichment of Asp and Glu in the P- and A-sites (Figure 3E). Second, there was a broad stretch of positive charge on the nascent polypeptide that extended >20 codons upstream of the sequence actually occupied by the stalled and stacked ribosomes (Figure 4A, left panel). This marked charge association was specific to disome footprints; in monosome footprint and RNA data it was only weakly detectable and absent, respectively (Figure 4A, middle and right panels). These observations suggested that there was an interplay between the nascent polypeptide and the speed at which codons that were located substantially further downstream were translated (Figure 4B). This idea is fully consistent with previous work that has suggested that electrostatic interactions between a positively charged nascent peptide and the negatively charged lining of the exit tunnel are a major factor promoting local slowdown of elongating ribosomes (Charneski and Hurst, 2013).

We next explored the relationship between disome sites and the structure of the translated polypeptide. Using genome-wide peptide secondary structure predictions with the three categories, structured (i.e. α -helix or β -sheet), unstructured and unknown, we calculated position-specific observed-to-expected ratios. These analyses revealed that the decoding center of the downstream ribosome was enriched for codons predicted to lie in unstructured parts of the polypeptide, whereas structured amino acids were depleted (Figure 4C, left panel). Upstream and downstream of the stalled ribosome this pattern was inverted, with an increase in structured and a decrease in unstructured amino acids. The identical analysis on monosome footprints yielded associations that were, although qualitatively similar to those seen for disomes, severely blunted in their effect size (Figure 4C, middle panel). Moreover, these associations were absent from the RNA data (Figure 4C, right panel). These findings indicated a high degree of specificity for disome sites and, taken together, they were fully consistent with a model according to which there was a preference for pausing during the translation of unstructured polypeptide stretches that were preceded

and followed by structured regions (Figure 4D). We investigated this hypothesis more explicitly by retrieving the transcript regions encoding "structured-unstructured-structured" (s-u-s) polypeptide configurations transcriptome-wide
460 (N=9312). After re-scaling to allow for the global alignment of structured and unstructured areas, we assessed the relative disome distributions across the s-u-s-encoding regions. These analyses revealed that disomes were enriched within the 5' portion of the unstructured region, just downstream of the s-u boundary (Figure 4E, left panel). By comparing with distributions obtained from ran-
465 domizations of the disome peak positions within the same dataset, we could conclude that the observed disome enrichment was significantly higher than expected by chance. As before, a weak and no effect, respectively, were detectable in monosome footprint and RNA data (Figure 4E, middle and right panels). Finally, the position-specific analysis (i.e. without re-scaling) at the s-u bound-
470 ary indicated that disome sites were particularly enriched at the second codon downstream of the s-u transition (Supplemental Figure S8A, right panel). As an additional control for the specificity of these associations, we also analyzed the inverse configuration, u-s-u (Supplemental Figure S8D) and conducted all analyses on monosome footprint (Supplemental Figure S8B, E) and RNA data
475 (Supplemental Figure S8C, F) as well. Taken together, the analyses established that the most prominent enrichment was that of disome sites within the unstructured area of the s-u-s configuration, frequently directly after the s-u boundary. Visual inspection of individual examples of where disome-associated amino acids mapped within known protein structures confirmed this notion, as
480 shown for PSMA5, ALDH1A1, GAPDH, and EIF5A (Figure 4F-I).

In summary, we concluded that there was a direct link between the occurrence and positions of ribosomal pause sites and structural features of the nascent polypeptide. Translational pausing was thus more likely to occur while decoding negatively charged amino acids that were downstream of extended
485 positively charged regions of the polypeptide, and within unstructured areas downstream of structured regions.

Evolutionary conservation at disome sites suggests functional significance of pausing events

The above analyses did not allow distinguishing whether the observed ribo-
490 somal pauses were functionally important – for example to ensure independent
folding of individual protein domains, undisturbed from downstream nascent
polypeptide stretches – or whether they rather represented an epiphenomenon
of such processes. For example, the biosynthesis of proteins and their folding
could slow down translation and thus, as a downstream effect, lead to ribosome
495 pausing and collisions, without being of functional relevance for the preced-
ing folding event itself. We reasoned that in the case of a functional role, the
codons on which pausing occurred would show higher evolutionary conservation
than expected. Our first approach to this analysis consisted in using PhyloP
scores (with categories highly conserved, conserved, neutral) and calculating
500 their observed-over-expected ratios at disome sites, analogous to the above en-
richment analyses. Indeed, highly conserved codons were enriched at the P- and
A-sites of the downstream ribosome (Figure 5A). Conceivably, however, several
confounding factors likely affected this analysis, in particular the selectivity of
disome sites for specific amino acids (Figure 3E) and the association with un-
505 structured protein regions (Figure 4C). Moreover, we observed that especially
for mRNAs with high translation efficiency, transcripts that were overall more
highly conserved showed increased disome levels, whereas less conserved tran-
scripts had a tendency to be disome-poor (Figure 5B). To disentangle these
multiple, confounding biases, we thus used a different approach to analyze evo-
510 lutionary conservation. We first selected the most prominent deterministic sites
– i.e. pausing events that were not merely attributable to high ribosomal traffic
– in order to enrich for potential functionally relevant cases (Supplemental Table
S3). We then compared the PhyloP scores of these high-disome deterministic
sites (top 5650 peaks from 1185 genes) with the scores of matched control sites
515 that were derived from the same set of transcripts and had a similar overall
codon composition (Figure 5C). An increase in conservation at the P- and A-
site codons, as compared to up- and downstream codons, was clearly evident.

However, this effect was also present, and only slightly lower, at the control sites. This suggested that the enrichment seen in Figure 5A was to considerable extent
520 caused by the intrinsic bias at the P- and A-sites for specific codons/amino acids that were *per se* conserved. We therefore used linear regression analysis to determine what contribution the disome site made to overall conservation. Using conservation as the response and three explanatory variables (transcript conservation, codon conservation, presence of a disome peak, adjusted $R^2 = 0.22$,
525 $F(3, 3012510) = 2.84e + 05$, $p < 2e^{-16}$), our approach uncovered that the influence of the disome site on overall conservation was tangible and statistically significant ($B=0.106$, $\beta = 0.048$, $p < 2e^{-16}$), though small relative to that of general transcript and codon conservation (Supplemental Table S4). In conclusion, this analysis revealed a distinct signature of codon conservation associated
530 specifically with disome peaks. Although the overall contribution of disome occupancy to conservation was low when assessed over many genes ($N=6007$ in the regression model), it is conceivable that it makes an important contribution for individual genes or groups of genes.

We further analyzed whether the genes carrying the most prominent de-
535 terministic sites showed any specific enrichment that would indicate whether particular processes, pathways or co-translational events were especially prone to pausing. First, we searched whether structural data was available for the proteins with high disome sites. Out of the first ~ 50 genes in the list (Supplemental Table S3), about 20-25 structures (murine protein or orthologous protein
540 from another mammal) were available. Mapping the disome site amino acids onto the structures revealed that in most cases, these were located in unstructured regions, and very often directly at the structured-unstructured boundary (Supplemental Figure S9; Figure 4F-I).

Functional enrichment analysis (using the top-200 genes; Supplemental Ta-
545 ble S3) identified a very strong bias towards transcripts encoding proteins that were annotated as "co-factor and co-enzyme binding", that were involved in "oxidation-reduction processes", and that localized to mitochondria (Figure 5D; Supplemental Table S5). It is tempting to speculate that the integration

and/or covalent attachment of co-factors (a common feature of oxidoreductases)
550 onto polypeptides is coordinated co-translationally and dependent on transla-
tional pausing. Moreover, there is evidence for co-translational protein localiza-
tion/import to mitochondria (reviewed in Lesnik et al. (2015)). The enrichment
of deterministic disome peaks among transcripts encoding mitochondrially lo-
cated proteins may reflect a specific feature of the kinetics of translation for this
555 group of genes. Finally, visual inspection indicated that at least 7 transcripts
encoding proteins that are known to contain the rare amino acid selenocysteine
displayed a strong disome peak (*Glutathione peroxidase 4*, *Gpx4*; *Methionine-R-*
sulfonide reductase B1, *Msrb1*; *Selenoproteins F, K, P, S, and T*; Supplemental
Table S3). Interestingly, selenocysteine decoding is known to be particularly
560 slow (Stoytcheva et al., 2006; Howard et al., 2013).

Disome codon usage affects steady-state protein abundance

We speculated that in cases where proteins assemble into large multiprotein
complexes, poorly coordinated translation kinetics may affect the abundance of
the proteins due to changes to the efficiency at which proteins are incorporated
565 into the complex. Excess unincorporated protein may be subject to degradation,
mislocalization, aggregation or similar, which all may impact the steady-state
protein level. In the list of prominent deterministic sites (Supplemental Ta-
ble S3) we noted that several transcripts encoding ribosomal proteins showed
strong, discreet disome peaks. We selected two of them – *Rps5* and *Rpl35a*
570 (Figure 6A-B) – to assess whether a change in codon usage at the disome site
would influence the proteins’ steady-state accumulation. To this end, we cloned
the cDNAs in-frame with firefly luciferase in a lentiviral vector that allows in-
ternal normalization to renilla luciferase (Figure 6C). For *Rps5*, change of codon
usage at the Asp-Ile disome site from its natural GATATT (*Rps5*-wt), a codon
575 usage that is particularly highly disome-prone (46.6% of sites show disomes
transcriptome-wide; Supplemental Table S2), to the mutant version GATATC
(*Rps5*-mut2), that is transcriptome-wide disome-poorer (26.4% of sites have di-
somes) led to a significant change in steady-state protein output, although both

constructs encoded for precisely the same protein at the amino acid level.

580 Other variants of *Rps5* did not show an effect, although some of the codon usages, such as that in *mut1* and *mut3*, are also transcriptome-wide disome-poorer. Moreover, codon usage did not affect reporter levels for an analogous *Rpl35a* reporter (data not shown), indicating that identifying functionally important disomes sites and testing for such function will be one of the future
585 challenges.

In summary, however, it is encouraging that even in this simple assay that does not even test directly for changes to protein function, codon usage at a disome site affected gene expression output for an example we tested, *Rps5*. We anticipate that in the future, more sophisticated assays that directly probe for
590 functionality of the translated protein – for example enzyme assays, or genetic experiments that test the ability of disome site variants to rescue the loss-of-function of the endogenous gene – will allow identifying many functionally important cases of translational pausing.

Discussion

595 It has long been known that elongating ribosomes can pause in response to obstacles that they encounter, and most pauses are assumed to occur in a transitory fashion and to be resolved in a productive manner (Schuller and Green, 2018). For certain cases, there is good evidence that the translational pause is even an integral, possibly obligatory part of the mechanism of nascent polypeptide synthesis, as exemplified by the pausing seen on signal peptide-encoding
600 transcripts that require targeting to the secretory pathway (Wolin and Walter, 1988). Finally, in certain cases, pauses can be unresolvable and terminal, thus triggering a dedicated ribosome rescue program (Joazeiro, 2019). While a number of previous studies have attempted to infer pausing from monosome footprint
605 intensities, there is clear benefit in tracking stalled ribosomes using more direct evidence, such as specific footprint size variants (Guydosh and Green, 2014). Ribosome collisions are intrinsically linked to pausing, and the characteristics

of the disome footprints that we analyzed in this study indicate that they truly represent a steady-state snapshot of the translational pausing and collision status in mouse liver *in vivo*. A striking finding from our analyses using spike-based quantifications of disome vs. monosome signals is the sheer quantity of translating ribosomes that are trapped in the disome state. For a typical, highly translated mRNA, we estimate that $\sim 10\%$ of elongating ribosomes are affected by this phenomenon (Figure 2Q). While we consider many of these collisions to be rather an epiphenomenon of high ribosomal flux (i.e. "stochastic", though more likely to occur on certain codons/amino acids than on others) than an indication that they fulfill a downstream biological function or trigger a specific cellular response, the loss of monosome footprints into disomes (and even into higher order buildups, see e.g. trisome footprints in Figure 1A-B) poses challenges to the interpretation of existing ribosome profiling data. Simple monosome footprint-based quantifications very likely underestimate translation rates, especially for highly translated transcripts.

Which stalling events is our disome profiling method capturing, and which ones are missed? In yeast, stalls at truncated 3' ends of mRNAs have recently been shown to lead to a class of small monosome footprints of ~ 21 nt and, when an additional incoming ribosome stacks onto this stalled ribosome, of ~ 48 nt (Gyudosh and Green, 2014). Similar short footprints also occur in human cells (Wu et al., 2019), and a study in HeLa cells has suggested that both transient and hard stalls trigger an endonucleolytic cleavage that leads to short footprint species as well (Ibrahim et al., 2018). It is thus likely that such short footprints represent the more harmful pauses that provoke specific clearance pathways. The abundant ~ 60 nt footprints we describe here are distinct from the above not only in size, but likely also in the translational state that they reflect. They match the size reported to occur on SRP-related pausing *in vitro* (Wolin and Walter, 1988). Moreover, footprints of this size were also noted in the above yeast study (Gyudosh and Green, 2014). Although the authors did not further pursue this footprint species in greater detail, it is interesting to note that they reported a similar bimodal size distribution and a depletion from the first

codons post-initiation, just as seen in our liver data as well (Figure 1E-F). Taken
640 together, the association with signal peptides, the high steady-state abundance,
and the absence of signs of mRNA cleavage at the stall site, all suggest that
our disome method captures in particular the class of "benign" collisions from
resolvable stalling events, including possible programmed cases.

Liver disome footprints show distinct sequence characteristics which are
645 largely governed by the P- and A-site amino acids of the downstream ribosome
(Figure 3E). There is little specificity at the E-site, which is notable because
previous work that estimated pause sites from monosome footprints identified
pausing signatures with a strong E-site bias for proline (e.g. Ingolia et al. (2011);
Zhang et al. (2017); Pop et al. (2014). Due to their particular chemistry, prolines
650 (especially in a poly-proline context) are well-known for their difficult peptide
bond formation and they are slow decoding, leading to translational stalls that
can be resolved through the activity of eIF5A (Gutierrez et al., 2013). Structural
studies have suggested direct interactions of eIF5A with a vacant E-site that has
lost its empty tRNA before peptide bond formation has occurred between the
655 growing polypeptide at the P-site and the incoming charged tRNA at the A-site
(Melnikov et al., 2016; Schmidt et al., 2015). Apart from a minor signal in the
A-site codon (Figure 3E; Supplemental Figure S6), we do not see any prolines
associated with liver disome sites at all, which may be attributable to high ac-
tivity of eIF5A in this organ. We have noted that based on monosome footprint
660 RPKMs, eIF5A is indeed synthesized at very high levels that even exceed, for
example, those of the essential elongation factor eEF2 (data not shown). More-
over, it is curious that eIF5A is itself among the 200 genes with the strongest
disome peaks, occurring on a conserved Gly-Ile position (Figures 2P and 4I;
Supplemental Table S3). It is tempting to speculate that translational pausing
665 on *eIF5A* mRNA may be part of a mechanism designed to autoregulate its own
biosynthesis.

Beyond the lack of proline, the specific amino acid and dipeptide motifs that
we find enriched at the paused ribosomes (Figure 3; Supplemental Table S2)
show some resemblance as well as distinct differences to previous observations.

670 For example, Asp and Glu have been associated with presumed pause sites (In-
golia et al., 2011; Ibrahim et al., 2018), and Asp codons also figure among those
whose footprint signal increases strongest (apart from Pro) in cells deficient of
eIF5A (Pelechano and Alepuz, 2017; Schuller et al., 2017). The striking associ-
ation of pause sites with isoleucine is an unexpected outcome of our study, as
675 to our knowledge this amino acid is not typically reported among the top-listed
associations with paused ribosomes. For GI, DI, and a subset of NI dicodons
– and certain non-isoleucine dicodons as well – transcriptome-wide 35-50% of
such sites carry a strong disome footprint (Supplemental Table S2). These find-
ings suggests that on top of the simple three-letter codon table, a six-letter code
680 contributes to regulating the kinetics of translation by organizing the biosynthe-
sis of nascent polypeptides into segments separated by intermittent stop sites.
Our analysis identified thousands such sites transcriptome-wide (Supplemen-
tal Table S3) which, as an ensemble, likely reflect an array of different protein
biosynthetic phenomena whose common denominator is local ribosome slow-
685 down. Importantly, even on the global set, an association of disome sites with
structural features of the nascent polypeptide is evident (Figure 4). Conceiv-
ably, this indicates that one of the major reasons for pausing could lie in the
coordination of translation with the folding, assembly or structural modification
of the nascent polypeptide. Interestingly, there is compelling recent evidence
690 from yeast that many multiprotein complexes assemble in a co-translational
fashion (Shiber et al., 2018) and that the association of the individual subunits
involves translational pausing (identified from high monosome footprint peaks)
during the biosynthesis of the nascent polypeptides (Panasenکو et al., 2019).
Notably, the showcase example identified in the latter study are two proteins
695 of the yeast proteasome regulatory particle, Rpt1 and Rpt2, whose ribosomal
pausing leads to association of the translating ribosomes into heavy particles
("assemblysomes") in which the nascent peptides assemble into the multipro-
tein complex. Intriguingly, we find that also in mouse liver, a set of proteasomal
protein mRNAs carries high disome peaks, e.g. *Psmđ5* (Figure 2I) and more
700 than 10 other proteasome subunits (Supplemental Table S3).

In conclusion, we are confident that the disome profiling methodology that we present here is an important complementary technique to the already available ribosome profiling repertoire. Importantly, although the technique, just like conventional ribosome profiling, delivers a 'snapshot' of the translation status, the cellular disome state provides specific, new information on the kinetics of translation. It will be exciting to collect and analyze such data across different experimental setups in order to evaluate to what extent translational pausing represents an obligatory, potentially regulated event that contributes to physiological gene expression output. Through such datasets, and already through the extensive data we have collected and analyzed in the framework of this study, important questions are likely to become experimentally accessible, such as on the assembly of multiprotein complexes, on the co-translational attachment of protein co-factors, on the mechanics of biosynthesis of transmembrane proteins, or on the coupling of translation and RNA decay.

Acknowledgments

We thank the Lausanne Genomics Technologies Facility for high-throughput sequencing infrastructure, and Alex Tuck and members of the Gatfield lab for critical reading of the manuscript. DG acknowledges funding by the Swiss National Science Foundation through the instruments: National Centre of Competence in Research (NCCR) RNA & Disease (grant no. 141735) and individual grant 179190.

Author Contributions

ABA: conceptualization, investigation, data curation, formal analysis, validation, visualization, methodology, writing draft; AL: investigation, formal analysis, methodology; MDM: Investigation, formal analysis; RD: formal analysis, visualization; PJ: investigation, methodology; DG: conceptualization, resources, formal analysis, supervision, funding acquisition, visualization, project administration, writing draft.

Declaration of Interests

730 The authors declare no competing interests.

Main Figure Titles and Legends

Figure 1. Sequencing of disome footprints identifies transcriptome-wide ribosomal collisions

(A and B) Northern blot analysis of RNase I-treated mouse liver extracts
735 probed with radioactively labelled short oligonucleotides antisense to different
regions of *Albumin* (Alb) mRNA (A) and *Major urinary protein 7* (Mup7)
mRNA (B). Expected footprint sizes for monosomes, disomes and trisomes are
shown to the left of blots. Positions of probes (nt) relative to the CDS start site
of their respective transcripts are shown above each lane and also depicted as
740 blue boxes below the black bars representing CDS regions (top). Red boxes at
the 5' end indicate the coding region for signal peptide (SP) on each transcript.

(C) Graphical representation of experimental setup for sequencing of 60 nt
disome footprints.

(D) Bar-plots of percentages of reads from monosome and disome libraries
745 that were mapped to different sequence types: rRNA (gray), tRNA (golden),
genomic (green) and cDNA (teal for monosomes and brick red for disomes).
Percentages of unmapped reads are shown in blue.

(E) Histogram of insert size (nt) for reads that mapped to cDNA sequences
(monosomes: teal, disomes: brick red). A single mode for monosomes (29-30 nt)
750 and two modes for disomes (59-60 and 62-63 nt) are labeled above histograms.

(F) Density distribution of disome reads within 120 nt from the start or
120 nt from the stop codons reveals a 3-nt periodicity of disome footprints
within coding sequences. The meta-transcript analysis quantified the mean of
per-transcript normalized number of reads (monosomes: teal, disomes: brick
755 red) at each nucleotide based on the A-site prediction (small red boxes) as 15
nt downstream of the 5' ends of the footprints. Only transcripts from single
protein isoform genes with totalRNA-RPKM > 5 and CDS >400 nt (N = 5456)
were used (described in STAR Methods). Positions on footprints, corresponding
to the predicted E-, P-, and A-sites of ribosomes that presumably protected the
760 corresponding footprints, were shown in graphical depictions. The start and stop

codons were highlighted with small green boxes on a representative transcript (at the bottom).

Figure 2. Occurrence of disomes is associated locally with signal peptides, and globally with high volumes of translation

765 (A and C) Density distribution of disome footprints identify signal peptide (SP)-related pausing events. Meta-transcript analysis (see Figure 1F) quantified the mean normalized footprint densities of disomes (A) and monosomes (C) within 400 nt from the start or -400 nt from the stop codons of transcripts confirmed or predicted to code for SP (red, N = 713) or not (blue, N = 4743).

770 (B and D) Violin-plots showing the probability densities of length normalized proportions of footprints within the first 75 codons and the rest of CDS from transcripts with a signal peptide (red, N = 713) or without one (blue, N = 4743) for disomes (B) and monosomes (D).

(E) Scatter-plot illustrating the relationship between per-gene normalized 775 densities of disome and monosome footprints. The subset of all genes included in the analysis (N = 8626; black and red dots together) that code a signal peptide (SP) (N = 1119) are denoted by red dots. Kernel density estimates are plotted on the margins (monosome on x-, disome on y-axis) for datasets of all genes (black) and SP coding genes (red) (without an axis of ordinates). Deming 780 regression (errors-in-variables model to account for errors in both monosome and disome footprint estimates) lines are shown for all genes (black) and the SP coding subset (red). Regression slopes and their 95% confidence intervals (CI) are given in the top-left legends. Dashed gray line indicates the 1-to-1 slope.

(F - P) Distribution of normalized counts of monosome and disome foot- 785 prints (per nt) along transcripts of representative genes reveals presence of stochastic and deterministic sites. Upward y-axis of the bar-plots show the normalized read counts for disomes (brick red), while downward axis was used for monosomes (teal) and totalRNA (pink, pile-up). Transcript coordinates (nt) are shown on x-axis; the regions corresponding to respective CDS are shaded in gray. 790 If present, the signal peptide coding regions are indicated with small red boxes

along the x-axis. Plots are shown for *Adhesion G protein-coupled receptor G3* (*Adgrg3*), *Transferrin receptor* (*Tfrc*), *Proteasome 26S Subunit, Non-ATPase 4* (*Psm4*) and *5* (*Psm5*), *Aldehyde dehydrogenase 1a1* (*Aldh1a1*), *Pyruvate kinase liver and red blood cell* (*Pklr*) and *Eukaryotic translation initiation factor*
795 *5A* (*Eif5a*) in **(F - P)**, respectively.

(Q) Box-plots illustrating the estimated amounts of ribosomes that were retained in disomes as a percentage of all translating ribosomes for different groups of genes in mouse liver samples. Box-and-whiskers were drawn for all genes, detectably expressed in spike-in experiment (All, gray, N = 7375), sub-
800 sets that codes for SP (SP, red, N = 892) and not (no SP, blue, N = 6483) and stratified into 8 groups based on the octiles of the translation efficiency (TE) calculated using all genes which had the following right-closed interval boundaries (-5.41, -1.23, -0.77, -0.47, -0.23, -0.04, 0.17, 0.47, 3.17), depicted as increasing TE below the x-axis. Width of each box is proportional to the number of data
805 points it represents.

Figure 3. Disome sites show specific enrichment of amino acids and codons

(A) Position-specific enrichment analysis of proximal sequences reveals selectivity for amino acids in the decoding center of pausing ribosomes. Normal-
810 ized ratios of observed-to-expected occurrences (y-axis, log-scaled) of nucleotide triplets, grouped by the amino acid they code (color codes in the right inset), are plotted for each codon position relative to the estimated A-site (0 at x-axis) of the leading ribosome of disomes (left), or of the individual ribosome in the case of monosomes (middle). For total RNA (right), position 0 denotes the mid-
815 point of the reads. Ratios above and below 1 suggest enrichment and depletion, respectively. The vertical gray bars indicate the likely positions of the 5' and 3' ends of the read inserts for different library types. In addition to the A-site, positions corresponding to P-sites are also marked by vertical dashed lines.

(B - D) Position-specific enrichment plots of sequences coding for representative amino acids at and around pause sites identified by disomes. Similar to
820

(A), however, the triplets were not combined by the amino acids, instead shown individually (color codes in the right inset) for aspartic acid (Asp), isoleucine (Ile) and glycine (Gly), respectively, in **(B - D)**.

(E) Position weight matrix of sequence triplets grouped by amino acids illustrates enrichment and depletion specific amino acids within the decoding center
825 of the leading ribosome of the disomes. Position-specific weighted \log_2 -likelihood scores were calculated from the observed-to-expected ratios (A). Enrichments and depletions were represented with positive and negative scores, respectively. Height of each single-letter amino acid character is determined by its absolute
830 score. At each codon position, the order of letters was sorted by the absolute scores of the corresponding amino acids, decreasing towards the 0-line. Amino acid letters are colored by their hydrophobicity and charges (color codes at bottom). The disome pair and the their footprint are depicted graphically at the top. The gray zones at the extremities of the footprint denote the likely
835 positions/regions of the 5' and 3' ends of the read inserts.

(F and G) Similar to (B - D), for asparagin (Asn) **(F)** and lysine (Lys) **(G)**.

(H) Position-specific enrichment plots of sequences coding for dipeptides. Similar to (A), however, instead of triplets and single amino acids, 6mers coding
840 for a pair of amino acids (dipeptides) were used to calculate the observed-to-expected ratios for all possible dipeptides. Color code is not given due to vast number of dipeptides.

(I - L) Similar to (B - D), however, enrichment of individual 6mers are shown for representative dipeptides: Gly-Ile **(I)**, Asp-Ile **(K)** and Gly-Asp **(L)**.

(M) Enrichment and codon selectivity of all amino acid combinations corresponding to the predicted P- and A-sites of the leading ribosome. Identities of the amino acids at the P- and A-site are resolved vertically and horizontally, respectively. The area of disks represents enrichment of disome sites at the dipeptide signature. The color of disks represents the codon selectivity for a
845 dipeptide signature, calculated as the difference between the max. and min. enrichment ratios (log) of all 6mers coding for that dipeptide.
850

(**N and O**) Similar to (I - L), for Asp-Lys (**N**) and Gly-Gly (**O**).

(**P**) Codon selectivity at certain dipeptides. Disome occupancy for all di-codon combinations (3721) were plotted in descending order. Occupancies were
855 calculated as the raw percentage of sites observed with a disome to all present sites for a given 6mer (dicodon) across the studied transcriptome. Annotated are two pairs of 6mers coding for Asp-Lys or Gly-Gly to illustrate the rather large differences in their disome occupancies (blue vs black fonts for high vs. low occupancy, respectively) as inferred in (N and O).

860 **Figure 4. Disome site positions are related to the charge and secondary structure of the nascent polypeptide**

(**A**) Position-specific enrichment analysis of proximal sequences reveals association of positive charges in the nascent polypeptide with pausing ribosomes. Average charge of 3 consecutive amino acids were stratified into 5 charge groups
865 (interval boundaries and color codes on the left). Normalized ratios of observed-to-expected occurrences (y-axis, log-scaled) of charge groups were plotted at the center position of the tripeptide relative to the estimated A-site (0 at x-axis) of the leading ribosome of disomes (left panel), or of the individual ribosome in the case of monosomes (middle). For total RNA (right), position 0 denotes
870 the midpoint of the reads. The vertical gray bars indicate the likely positions of the 5' and 3' ends of the read inserts for different library types. Positions corresponding to P- and A-sites are also marked by vertical dashed lines.

(**B**) Graphical representation of the electrostatic interactions between the leading ribosome and the nascent peptide chain. Associations of negatively
875 charged residues (blue) with the P-A sites and a stretch of positively charged residues (red) with the exit tunnel are depicted. Corresponding codons on the mRNA (series of filled rectangles at the bottom) are colored similarly based on the charge of the amino acids they code for.

(**C**) Association between disome sites and the structure of the translated
880 polypeptide. Based on the UniProt database, each position of translated peptides was labeled according to their structural information: 'structured' for

α -helix or β -sheet, 'unstructured' and 'unknown'; β -turns were excluded. Normalized ratios of observed-to-expected occurrences (y-axis, log-scaled) of structural categories were plotted relative to the estimated A-site (0 at x-axis) of the leading ribosome of disomes (left), or of the individual ribosome in the case of monosomes (middle). See (A) for other elements.

(D) Graphical model depicting a preference for pausing during the translation of unstructured polypeptide stretches (orange) that were preceded and followed by structured regions (purple).

(E) Enrichment of disome sites within the unstructured stretches of polypeptides that are preceded and followed by structured regions. Structured (min. 3 aa, up to 30th position) - unstructured (min. 6, max. 30 aa) - structured (min. 3 aa, up to 30th position) regions were identified transcriptome-wide. Positions across regions were scaled to the length of the unstructured region and centered to its start, such that the start and the end of the unstructured region would correspond to 0 and 1, respectively (x-axis). Kernel density estimates (thick black lines) were calculated for peaks across normalized positions weighted with their normalized counts, estimated at the A-site of the leading ribosome for disomes (left, N = 9312), A-site of the monosomes (center, N = 21140) or center of total RNA reads (right, N = 24632). Confidence intervals for the kernel densities, which were calculated by randomly shuffling (N = 10000) peaks within each transcript, are shown by gray shaded regions: darkest at the center, 50% (median) to outward, 25%, 12.5%, 5%, 2.5% and 1%.

(F - I) Three-dimensional structure of four individual proteins with disome site amino acids highlighted. PSMA5, structure for the *H. sapiens* homologue (PDB ID: 5VFT); ALDH1A1, structure for the *H. sapiens* homologue (4WJ9); GAPDH, structure for the *H. sapiens* homologue (4WNC), corresponding residues at position 65-66; murine EIF5A (5DLQ). The strongest disome site of each transcript is highlighted in red.

Figure 5. Evolutionary conservation at disome sites

(A) Association of evolutionary highly conserved codons with the P- and A-

sites of disome sites revealed by position-specific enrichment analysis of proximal sequences. Along coding regions, PhyloP conservation scores were grouped into three categories: neutral - blue, [-3, 3), conserved - orange, [3, 5) and highly conserved [5,). Normalized ratios of observed-to-expected occurrences (y-axis, log-scaled) of conservation categories were plotted relative to the estimated A-site (0 at x-axis) of the leading ribosome of disomes (left), or of the individual ribosome in the case of monosomes (middle). For total RNA (right), position 0 denotes the midpoint of the reads. See Figure 4A for other elements.

(B) Box-and-whiskers illustrating the estimated percentages of ribosomes that were pertained in disomes for groups of transcripts with different overall evolutionary conservation. Groups included all detectably expressed genes (All, gray, N = 7375), which were stratified into 4 groups (N = 2270 or 2271 for each) (color code at the top) based on the quartiles of average PhyloP scores with following right-closed boundaries: -0.585, 2.327, 3.356, 4.239, 6.437. x-axis is same as in Figure 2Q. Width of each box is proportional to the number of data points it represents.

(C) Box-and-whiskers of PhyloP scores around deterministic disome sites. PhyloP scores were plotted at 20 codon positions (x-axis) surrounding prominent high-disome sites (N = 5650, brick red). The position of the leading ribosome of the disome, together with the E-, P- and A-sites were shown on top in a graphical representation. For control sites (pink), from the same pool of transcripts (N = 1185), random stretches of 20 aa were selected (N = 5909) such that 6mer frequencies corresponding to the PA positions would be very similar to that of high-disome sites used but without containing strong disome peaks. Significance in median disome-site PhyloP score difference between codon positions was tested with Dunn's method (FDR adjusted p-value $\leq \alpha/2$, $\alpha = 0.05$). Number of significant differences (out of 19 comparisons at each codon position) were given below the box-and-whiskers. At the P-codon (*), disome site PhyloP scores were significantly higher than they were in any other position, except the A-codon. At the A-codon (*^{*}), they were significantly higher than they were in 8 other positions.

(D) Functional enrichment analysis of top 200 genes containing most prominent deterministic sites. Five terms with the highest $-\log_{10}(p_{adj})$ values (horizontal bars) are shown from each Gene Ontology (GO) group: GO molecular function, GO cellular component, GO biological process.

Figure 6. Testing functional importance of disome sites on ribosomal proteins

(A) Upper: Transcript plot for *Rps5* showing the distribution of normalized disome peaks along the transcript coordinates. Shaded area highlights the CDS and the prominent disome peak at positions 32-33, corresponding to Asp-Ile (DI), is shown with a red arrow. Lower: Same for *Rpl35a* with disome peak at positions 44-45 (Gly-Lys) highlighted. See Figure 2F for further details.

(B) Three-dimensional protein structures of human RPS5 (PDB ID: 5VYC) and RPL35A (PDB ID: 5t2c). The position of the conserved disome site amino acids are highlighted in red.

(C) Upper: Schematic of the lentiviral reporter used to probe for the effect of synonymous disome site mutations on steady-state protein abundance. *Rps5* cDNA is fused in-frame to firefly luciferase and transcribed by a bidirectional promoter that also drives the control gene, Renilla luciferase. Lower: Effect of synonymous disome site mutations (mut1-mut5) on Firefly/Renilla ratios, expressed relative to the wild-type reporter which was internally set to 100% in each experiment (N=2-10). Mut2 vs. wild-type: p=0.008, Student's t-test.

Supplemental Figure Titles and Legends

965 **Figure S1. High-order ribosome protected fragments are highly re-**
970 producible under various assay conditions

(A) Northern blot similar to that shown in Figure 1A (probe Mup₇₅₈₋₈₁), but from experiments in which extract preparation and RNase I digestion was performed at different temperatures and with harsher detergent conditions, as indicated.

(B) Similar to (A), but under conditions in which the concentration of RNase I was varied, as indicated. Probes Mup₇₅₈₋₈₁ (left panel) and Alb₇₁₋₁₀₁ (right panel) were used to detect possible changes in disome footprints across conditions.

975 (C) Similar to (A) and (B), but using a different nuclease, micrococcal nuclease (MNase). Probes Alb₇₁₋₁₀₁ (left panel), Mup₇₅₈₋₈₁ (middle panel) and Mup₇₂₉₈₋₃₂₀ (right panel) were used to detect variations in disome footprints across conditions. Note that although MNase produced somewhat different patterns than RNase I, disome footprints were still readily detectable.

980 **Figure S2. Mapping characteristics of disome reads**

(A) Pie-charts of percentages of reads from monosome, disome and total RNA libraries that were mapped to different sequence types (rRNA, human rRNA, mt-tRNA, tRNA, mouse cDNA and mouse genome) or were unmmapped. Color codes are given in top right legend.

985 (B) Read distribution within 5' UTRs, CDS, and 3 UTRs for monosome (teal), disome (brick red), and total RNA (pink) data compared with the distribution expected by chance, which is determined by the feature sizes (gray; N = 7413). Note the enrichment of disomes reads within CDS and the depletion from UTRs, similar to that of monosome reads.

990 **Figure S3. Signal peptide and translational efficiency explain some portion of the observed disome sites**

(A) Density distribution of disome reads within 120 nt from the start or 120 nt from the stop codons reveals a 3-nt periodicity of disome footprints within coding sequences. Similar to Figure 1F, except for disomes, the meta-transcript analysis is aligned relative to the predicted A-site of the leading, rather than the stacked ribosome.

(B) Density distribution of disome footprints identify the translational pausing due to signal peptide (SP) recognition. Similar to Figure 2A, except, the meta-transcript analysis is aligned relative to the predicted A-site of the leading, rather than the stacked ribosome.

(C) Time dependent changes in translation efficiencies of ribosomal proteins result in changes in occurrence of disomes. Kernel density estimates of the difference in relative disome densities ($\log(\text{disome density}/\text{monosome density})$) between ZT12 and ZT2 were found to be significantly different for ribosomal proteins (RP, $N = 57$, dashed green line) and others ($N = 8558$, black line) by the two-sample Kolmogorov-Smirnov test (K-S test, $D = 0.1911$, $p = 0.032$). RP genes were identified to have increased translation efficiencies at ZT12 compared to ZT2 (Janich et al., 2015).

(D) Difference in translational efficiencies explains only a small portion of the variation in relative disome densities among transcripts. Relative disome density ($\log(\text{average normalized disome counts} / \text{average normalized monosome counts})$) for all transcripts in the dataset ($N = 8626$) was regressed (black line) on the translational efficiency (TE, $\log(\text{average normalized monosome counts} / \text{average normalized total RNA counts})$). Statistics on the coefficient and the regression are given in upper part of the scatter-plot. TE explained significantly a small (2.7%) portion of the variance. Dashed line marks the $b = 0$ line.

Figure S4. Experimental setup for spike-ins

(A) Graphical representation of experimental setup for sequencing of monosome and disome footprints spiked with pre-synthesized 30 and 60 nt RNA oligonucleotides.

(B) Northern blots of spike-in oligonucleotide mixes to assess the apparent

ratio of the 60 nt oligos to 30 nt oligos. Same radioactively 5' labelled DNA oligo could hybridize to both 30 nt and 60 nt RNA oligos, however, two probe molecules could hybridize to a single 60 nt oligo at the same time, therefore
1025 expected signal ratios were around 2.

Figure S5. Empirical identification of the offsets for estimation of the A-sites of the leading ribosome of the disome pair

(A) Position-specific information content for different size classes (55 - 64 nt) of disome footprints combined with their frame (5' position relative to the main
1030 CDS's open reading frame - 0, 1 or 2) revealed the optimal offsets for estimating the A-site of the leading ribosome. Position-specific information content was calculated using the Kullback-Leibler divergence scores of observed-to-expected ratios of codon analysis (similar to Figure 3A but without any A-site estimation - only using the 5' ends of the footprints) as described elsewhere (O'Connor
1035 et al., 2016). For each size group, the KL plots were drawn separately for three frame offsets (color code at the top of the figure). For combinations of footprint size and frame, where information content could be resolved at nucleotide level within the expected region of the decoding center (2 - 3 peaks in KL), offsets from the footprint's 5' end were calculated to the most probable position of
1040 the A-sites (colored rectangles). Frequencies of each size group are given at the right side in million reads. For each plot, 5' and 3' ends of the footprints were marked with vertical dashed lines and the region occupied by the footprint was shaded in a gray tone.

(B) Graphical model illustrating the different configurations of disomes evidenced by the two major populations of 59-60 nt and 62-63 nt footprints. Based
1045 on (A), the difference between these two size groups is whether the ribosomes were collided completely (top), or a small gap of a single codon was left between the two (bottom).

Figure S6. Enrichment of amino acids and codons at disome sites.

1050 Identical to the analysis in main Figure 3A-D, but for the complete set of amino acids, and for disome, monosome and RNA data. See figure legend for

Figure 3A for details.

Figure S7. Position weight matrix for disomes, monosomes and RNA data, and specifically on the signal peptide.

1055 (A) Position weight matrix of sequence triplets grouped by amino acids illustrates enrichment and depletion specific amino acids within the decoding center of the leading ribosome of the disomes. At the top of the panel, the same data as in main Figure 3E (disomes) is shown. Middle and lower parts of the panel depict the identical analysis for monosome and RNA-seq data,
1060 respectively. See figure legends to Figure 3E for details. Of note, the figure shows that monosome footprints had a similar, though in magnitude massively reduced preference for amino acids compared to the disomes. No specificity was found in total RNA, as expected.

(B) As in panel (A) and in Figure 3E, but the position weight matrix was
1065 calculated only from the disomes that were found over codons 8-75 of signal peptide-containing transcripts, i.e. over the positions where the disomes related to SRP recruitment are located. Interestingly, even in this area the global pattern of amino acids at which disomes were preferentially found, corresponded to the pattern seen transcriptome-wide.

1070 **Figure S8. Analysis of disome sites at and around junctions between structured and unstructured regions**

(A - C) On the left, same plots in Figure 4E were redrawn to facilitate
comparison. Data from Figure 4E were analyzed within a +/- 30 codon window around positions that corresponded to junctions between structural regions, such
1075 as structured-to-unstructured junction (left plots) or unstructured-to-structured (right plots). All junction regions were aligned, such that first codon at the junction was labeled as 0 on the x-axis (codon positions). Average densities (y-axis) of disome (A), monosome (B) footprints or total RNA reads (C) at each position within the window were plotted (red dots). Data from randomized
1080 peaks (N = 10000) were shown with box-and-whiskers at each codon position.

(D - F) Same as (A - C), but the structural configuration was reversed as a control: unstructured - structured - unstructured. In these kind of regions, a decrease in the middle structured section was observed, consistent with Figure 4E and (A - C). Analysis was performed for disome (D), monosome footprints (E) and total RNA reads (F).
1085

Figure S9. Examples of disome sites in in transcripts/proteins from the list of deterministic sites

Each panel is comprised of two parts: the three-dimensional structure of the protein (from mouse or, if not available, a related mammalian species, with the disome site amino acids highlighted in red and with a shaded circle) and transcript plots showing the distribution of disome (brown), monosome (green) and total RNA (pink) signals along the transcript. Shaded areas correspond to the CDS (for UTRs, only the boundaries were plotted). The protein structures have been generated using the following data (PDB ID in parenthesis):
1090

- 1095 (A) Homologue of *Ctsd* from *H. sapiens* (1LYW).
- (B) Homologue of *Dynlrb1* from *H. sapiens* (6F1Z), corresponding residues at positions 93-94.
- (C) Homologues of *Fh1* from *H. sapiens* (5UPP), corresponding residues at positions 58-59.
- 1100 (D) *Fth1* structure from *M. musculus* (6S61).
- (E) Homologue of *Gpd1* from *H. sapiens* (6E8Y).
- (F) Homologue of *Mrps17* from *S. scrofa* (5AJ3).
- (G) *Mrsb1* structure from *M. musculus* (2KV1).
- (H) Homologue of *Nars* from *H. sapiens* (5XIX), corresponding residues at
1105 positions 476-477.
- (I) Homologue of *Ndufb6* from *B. taurus* (5LDW).
- (K) Homologue of *Nqo2* from *H. sapiens* (1ZX1).
- (L) Homologue of *Pah* from *R. norvegicus* (1PHZ).
- (M) *Rdx* structure from *M. musculus* (3X23).
- 1110 (N) Homologue of *Sub1* from *H. sapiens* (4USG).

- (O) *Sult1d1* structure from *M. musculus* (2ZPT).
- (P) Homologue of *Tdo2* from *H. sapiens* (4PW8).
- (Q) Homologue of *Tkt* from *H. sapiens* (4KXU).

STAR Methods

1115 **Lead contact and materials availability**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Gatfield (david.gatfield@unil.ch).

Experimental model and subject details

1120 Extracts from 12-week-old male C57BL/6 mice were the same as reported previously (Janich et al., 2015), with experiments approved by the Veterinary Office of the Canon Vaud (authorization VD2376 to DG). NIH3T3 and HEK293FT cells were same cell lines as described in Janich et al. (2015). Culture conditions: DMEM; 10% FCS, 1% penicillin/streptomycin, all from Invitrogen; 37°C; 5%
1125 CO₂).

Experimental methods details

Northern blot

The northern blot protocol has been described in Gatfield et al. (2009). Radioactively labelled oligonucleotides: Alb_{1–22} ggagaaaggttaccacttcat, Alb_{71–101} c
1130 gatgggcgatctcactcttgtgtgcttctc, Alb_{131–165} gagatactgggaaaaggcaatcaggactag
g, Alb_{1099–1120} gatcagcaggcatggtgtcatgc, Alb_{1805–1827} ttaggctaaggcgtctttgc
atc, Mup_{71–21} cagcagcagcagcatcttcat, Mup_{758–81} gttccttcccgtagaactagcttc,
Mup_{7298–320} gtattgaatccatcatagctcac, Mup_{7563–584} tcattctcgggcctggaggcag,
Mup_{7688–708} tcagtgagacaggatggaatg. Please note that the lower part of the
1135 northern blot panels shown in Supplemental Figure S1B was also used in our
previous publication (Janich et al., 2016).

Library generation (monosome footprints, disome footprints, RNA)

The initial libraries used in this study (without spike-ins) for monosome footprints and RNA corresponded to the time points ZT0, 2 and 12 from our previous study (Janich et al., 2015). Per timepoint, two biological replicates were
1140

used (total 6 independent samples); every sample was the pool of liver lysate from two mice. Disome footprints were cut from the same gel as the monosome footprints of the initial study. For the spike-in experiments, independent libraries from frozen mouse liver lysates were prepared. For the general library preparation protocol, ribosome-protected mRNA fragments were generated from
1145 the lysates by RNase I digestion as previously described (Janich et al., 2015); micrococcal nuclease (Supplemental Figure S1C) was from New England Biolabs. For the excision of footprints from 15% urea-polyacrylamide gels, single strand RNA oligonucleotides of 26 nt and 34 nt (for monosome footprints) and
1150 of 52 nt and 69 nt (for disome footprints) served as size markers for excision of footprints. After fragment purification with miRNeasy RNA Extraction kit (217004 Qiagen), 5 μ g fragmented RNA was used for ribosomal RNA removal using Ribo-Zero Gold rRNA Removal Kit (MRZG12324 Illumina) according to Illumina's protocol for TruSeq Ribo Profile (RPHMR12126 Illumina). 2 μ l of
1155 RNA spike-in mix, containing three synthetic 30 nt RNA oligonucleotides (sequences: AAUACCACCCCAUGAACGCUGCACACACG, AACUACCGACUCAUCCCAUCUUGCCAGUAC, CUAUACUUACGAACCAGACGAAUCCUUG) and three 60 nt length oligos (sequences AAUACCACCCCAUGAACGCUGCACACACG, AA
1160 CUACCGACUCAUCCCAUCUUGCCAGUACAACUACCGACUCAUCCCAUCUUGCCAGUAC, CUAUACUUACGAACCAGACGAAUCCCUUGC
UAUACUUACGAACCAGACGAAUCCCUUG), at 0.016 fmol/ μ l, were added to the purified, rRNA-depleted RNA samples. Sequencing libraries were generated according to Illumina's TruSeq Ribo-Profile protocol with minor modifications.
1165 Monosomes and disomes were treated as independent libraries. cDNA fragments were separated on a 10% urea-polyacrylamide gel and gel slices between 70-80 nt for monosomes and 97-114 nt for disomes were excised. The PCR-amplified libraries were size selected on an 8% native polyacrylamide gel. Monosome libraries were at \sim 150 bp and disome libraries at \sim 180 bp. Parallel RNA-seq
1170 libraries were prepared essentially following the Illumina protocol; briefly, after total RNA extraction using miRNeasy RNA Extraction kit (Qiagen), ribosomal

RNA was depleted using Ribo-Zero Gold rRNA (Illumina), and sequencing libraries generated from the heat-fragmented RNA as described (Janich et al., 2015). Of note, 3 μ l of the same RNA spike mix as above were added to the total RNA after heat fragmentation (during the ice incubation step). All libraries were sequenced in-house on Illumina HiSeq 2500.

Cloning, lentiviral production, luciferase assays

For the generation of the Rps5 dual luciferase (Firefly/Renilla) reporter plasmid, *Rps5* CDS (without stop codon) was PCR-amplified from mouse cDNA using oligos Rps5CDS-F, aaaggatccgccaccATGACTGAGTGGGAAGCAGCC ACACCAG and Rps5CDS-R, ttggatccactagtGCGGTTAGACTTGGCCACA CGCTCCAGT, digested with BamHI and cloned upstream and inframe of luciferase into BamHI-opened dual luciferase vector prLV1 (Du et al., 2014); this vector is suitable for lentiviral production), and validated by sequencing. Disome site mutants were generated by site-directed mutagenesis with the primers: Rps5mut1-up, GATGACGTGCAGATCAACgacataTCTCTGCAGGATTAC ATTG; Rps5mut1-low CAATGTAATCCTGCAGAGAtatgtcGTTGATCTGC ACGTCATC; Rps5mut2-up, GATGACGTGCAGATCAACgacatcTCTCTGC AGGATTACATTG; Rps5mut2-low CAATGTAATCCTGCAGAGAgatgtcG TTGATCTGCACGTCATC; Rps5mut3-up, GATGACGTGCAGATCAACgacattTCTCTGCAGGATTACATTG; Rps5mut3-low CAATGTAATCCTGCAGAGaatgtcGTTGATCTGCACGTCATC; Rps5mut4-up, GATGACGTGC AGATCAACgataTCTCTGCAGGATTACATTG; Rps5mut4-low CAATGTAATCCTGCAGAGAtatcGTTGATCTGCACGTCATC; Rps5mut5-up, GATGACGTGCAGATCAACgataTCTCTGCAGGATTACATTG; Rps5mut5-low CAATGTAATCCTGCAGAGatcGTTGATCTGCACGTCATC. All mutants were verified by sequencing.

Plasmids were used to produce lentiviral particles in HEK293FT cells with envelope pMD2.G and packaging psPAX2 plasmids, and viral transduction of NIH3T3 cells, were performed following published protocols (Salmon and Trono, 2007). 1-2 weeks after lentiviral transduction, cells were lysed in passive lysis

buffer and luciferase activities were quantified using DualGlo luciferase assay system and a GloMax 96 Microplate luminometer (all from Promega). Firefly/Renilla luciferase (FL/RL) of the Rps5 wt plasmid were internally set to 100% in each experiment, and mutant FL/RL ratios expressed relative to wt.

Computational methods details

Preprocessing of Sequencing Reads

Initial quality assessment of the sequencing reads was conducted based on Illumina pipeline's (v1.82) preliminary quality values such as the percentage of clusters passed filtering (%PF clusters) and the mean quality score (PF clusters). Adapter sequences were removed with the `cutadapt` utility (Martin, 2011) and following arguments: `-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC --match-read-wildcards`. Trimmed read sequences were filtered by their size using an in-house Python script to conform the following inclusive ranges: [45,70] for disome footprints, [26,35] for monosome footprints, and [21,70] for total RNA reads. Finally, the reads were filtered for quality using the `fastq_quality_filter` tool from the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) with the following arguments: `-Q33 -q 30 -p 80`.

Mapping of Footprints to Mouse Genome

A similar sequential mapping strategy was adapted as described in Janich et al. (2015). The preprocessed insert sequences were mapped sequentially to following databases: mouse rRNA, human rRNA, mt-tRNA, mouse tRNA, mouse cDNA from Ensembl mouse database release 91 Flicek et al. (2013) and, finally, mouse genomic sequences (Genome Reference Consortium GRCm38.p2). With the exception of the final mapping against genomic sequences, `bowtie` version 2.3.0 Langmead and Salzberg (2012) was used with the following parameters:

```
-p 2 -L 15 -k 20 --no-unal
```

After each alignment, unmapped reads were used in the succeeding mapping. For each sequence, only valid alignments with maximum alignment scores were

1230 kept. For further analysis, only alignments against mouse cDNA were used,
unless specifically stated otherwise.

In parallel to the sequential mapping strategy, preprocessed total RNA sequences were also directly aligned against the mouse genome (GRCm38.p2). Alignments against genome sequence databases were performed using the STAR
1235 mapper version 2.5.3a (Dobin et al., 2012) with the following parameters:

```
    --runThreadN 6 --genomeDir=mouse/star/Mmusculus.GRCm38.91  
    --readFilesCommand zcat --genomeLoad LoadAndKeep  
    --outSAMtype BAM SortedByCoordinate Unsorted  
    --alignSJDBoverhangMin 1 --alignIntronMax 1000000  
1240 --outFilterType BySJout --alignSJoverhangMin 8  
    --limitBAMsortRAM 15000000000
```

The output of this alignment was then processed with `StringTie` version 1.3.3b (Pertea et al., 2015) to estimate the number of fragments per kilobase of exon per million fragments mapped (FPKM) for each transcript (Ensembl mouse
1245 database release 91), with the following parameters:

```
    -p 8 -G Mmusculus.GRCm38.91.gtf -A gene_abund.tab  
    -C cov_refs.gtf -B -e
```

The outputs were parsed with an in-house Python script to identify transcripts which had an FPKM >0.2 and an isoform abundance fraction >0.05 in at least
1250 2 samples. A database of expressed transcripts ($N_{genes} = 19508$, $N_{transcripts} = 24927$) was used in further analysis. Among those, genes that were estimated to have a single expressed isoform were annotated as **single transcript genes** ($N = 9711$). For genes with multiple transcript isoforms, the transcript, whose exons were inclusive of all others, was used whenever possible ($N = 548$).

1255 **Quantification of mRNA and Footprint Abundances**

Abundance of total RNA reads and monosome or disome footprints was estimated per locus as described in Janich et al. (2015). Separate counts were obtained for whole gene, UTRs and CDS. Only reads that were mapped uniquely

to a single gene and only to transcripts that were identified to be expressed (see
1260 Mapping of Footprints to Mouse Genome) were used. Exclusively for the analy-
sis of ribosomal proteins (Figure S3C), this criterion was slightly relaxed to also
include multireads that were mapping to a single protein coding locus. Tran-
scripts which did not have at least 10 counts in at least one third of the samples
were excluded. For all further analysis, reads that mapped to CDS regions were
1265 used, unless stated otherwise. A total of 8626 loci had above threshold read
counts within the CDS for all read types: total RNA, monosome and disome.

Read counts of total RNA and footprints were normalized with upper quan-
tile method of R package `edgeR` v3.16.5 Robinson et al. (2010). For increased
comparability between datasets, RPKM values were calculated as the number
1270 of reads per 1000 bases per geometric mean of normalized read counts per mil-
lion. Genes that had an average total RNA RPKM >5 were designated as
robustly expressed. Combined with the single transcript genes (see Mapping
of Footprints to Mouse Genome), robustly expressed single transcript genes (N
 $= 6007$) were used for analyses where inclusion of genes with multiple expressed
1275 isoforms was not possible (e.g meta-transcript analysis). Normalized footprint
densities were calculated as the \log_2 -ratio of footprint-RPKM to total RNA-
RPKM per gene, for disomes and monosomes. For the latter, this quantity is
also called translational efficiency (TE). In mouse liver, TEs were shown to be
stable over time-points around the day (Janich et al. (2015)); disome densities
1280 were similarly stable between the samples (ZT0, ZT2, and ZT12) and therefore
treated as replicates, unless stated otherwise.

Spike-in Normalization and Global Quantification of Ribosomes Re- tained in Disomes

Random 30 and 60 nt long RNA oligonucleotide sequences were designed follow-
1285 ing these criteria: (i) have a GC % similar to that of mouse liver translome
(mean was 52.05, 5% and 95% were 42.2 and 62.6, respectively), (ii) should
be void of potential hairpin structures and self-dimerization (using `ViennaRNA`
package 2.3.5, Lorenz et al. (2011)), (iii) should not be highly similar to mouse or

Drosophila transcriptome and genome, (iv) should not contain certain sequences
1290 at the extremities which we were identified as highly biased in our analyses (GG,
GC, CC, CG, CA, GA, TG, AC), (v) should not contain stop codons and (vi)
60mers were designed as 2 x 30mer repeats. Out of 35 possible candidates, 3
sets with different GC% were selected: 43, 50 and 56. Spike reads were mapped
and processed similarly to all other reads. To avoid counting degradation prod-
1295 ucts of the 60mers as 30mers, we devised a two-step counting algorithm. First,
spike read distributions were inspected on total RNA reads to assess possible
degradation and define proper size limits. The GC56 spike was eliminated from
further analysis due to fragmentation; for others [24,31] and [45,60] inclusive
size filters were used for 30- and 60mers, respectively. In addition to the size
1300 filtering, the presence/absence of the junction of the 2 x 30mer repeats were
identified for all spike reads. 30mers were included if they did not have a junc-
tion, and 60mers only if they did. Spike counts were first normalized for library
size with upper-quantile method and spike-in normalization factors were cal-
culated as 60mer / 30mer ratios per sample to correct the experimental biases
1305 between the disome and monosome counts. The spike-in normalization factors
were nearly identical for triplicate biological replicates (mean = 2.495, SD =
0.028). The spike-normalized counts of disomes and monosomes were then used
to estimate the percentage of ribosomes that were identified within disomes to
the whole, taking into account that each disome represented two ribosomes.

1310 **Observed-to-Expected Ratios For Proximal Sequence Features**

Calculation of observed-to-expected ratios for sequence features proximal to
footprint sites was performed as described in O'Connor et al. (2016) with mod-
ifications. Briefly, the

Calculation of observed-to-expected ratios for sequence features proximal
1315 to footprint sites was performed following the principles of Ribo-seq Unit Step
Transformation method (O'Connor et al., 2016). First, a Heaviside step func-
tion was applied to individual features (codon, amino acid, 6mer, dipeptide,
charge, secondary structure, PhyloP conservation categories) along each CDS,

such that a feature at a position was given a score of 1 or 0 depending on
1320 whether the footprint density at that position exceeded the average of the cor-
responding CDS. A margin of 30 nt were excluded from each end of the CDS.
Then, a typically 50-codon wide window (80-codon wide for certain analysis
such as charge in Figure 4A), was moved along the CDS regions at 3 nt steps,
except for analyses that required single nt resolution. Window positions were
1325 labeled relative to RUST scores, 0-position marking the score. The scores were
either not offset (5') or A-site offset (see Estimation of A-site Positions). At
each iteration, position specific occurrence of features was counted and associ-
ated with if there was a RUST score in that window. Present, observed and
expected values of each feature at each window position were calculated as sums
1330 over all windows. When necessary, Kullback-Leibler divergence scores were cal-
culated using the observed-to-expected ratios of all features (O'Connor et al.,
2016). Enrichment was calculated as the observed-to-present ratio normalized
to expected. All analyses were performed with in-house Python (creation of
data matrices) and R software (visualization and statistical analysis). Features
1335 that were based on (discrete) sequence information (nucleotide or amino acid
sequence) were created simply using the letters of such sequences in different
word sizes (such as CCT or proline for single; CCTCCA or proline-proline for
two-word). Other discrete data, such as secondary structure, were also analyzed
similarly. Features that were based on continuous numeric data, were first strat-
1340 ified into discrete levels. For example, PhyloP conservation scores were grouped
into three categories: neutral $[-3, 3)$, conserved $[3, 5)$ and highly conserved $[5,$
 $)$. Visualization of complex RUST ratios was facilitated using \log_2 transformed
position specific enrichment matrices with the `ggseqlogo` package for R (Wagih,
2017) and converting them sequence logos.

1345 **Estimation of A-site Positions**

The A-sites of the monosomes (RPF) were calculated identically as described in
Janich et al. (2015). For disomes, for initial analyses we used a similar approach
to estimate the A-site of the upstream ribosome in the disome pair as 15 nt from

the 5' end of the footprints. This approach was suitable for exploratory analyses
1350 (e.g meta-transcript analysis) for facilitating the comparability to monosome
results. In other analyses, we used an empirical method to estimate the A-site
of the leading (downstream) ribosome within the disome pair. In order to infer
the optimum offsets for different lengths of footprints, we first split the disome
1355 footprints by their size, from 55 to 64 nt. Within each size group, footprints
were further split into 3 classes based on their open reading frame relative
to that of the main CDS. For each group, position-specific (relative to their
5' ends at nucleotide resolution) information content matrices were calculated
using the Kullback-Leibler divergence scores (O'Connor et al., 2016) of observed-
1360 to-expected ratios of codon analysis (see Calculation of Expected-to-observed
Ratios For Proximal Sequence Features). For combinations of footprint size and
open reading frame, where the position of PA sites could be identified as highest
information positions (with 2 peaks 3 nt apart from each other) around 40 -
50 nt downstream of the 5' ends of the footprints, exact offsets were calculated
as the distance of the deduced A-site from the 5' end. Following offsets for 58,
1365 59, 60, 62 and 63 nt long disome footprints on different open reading frames
were used, respectively: [45, 44, 43], [45, 44, 46], [45, 44, 46], [48, 47, 46], [48,
47, 49]. Total RNA reads were offset with different methods to be consistent
with the dataset they were being compared to: by their center (general), +15
(when compared to monosomes, also selecting a similar size range of 26-35 nt)
1370 or disome offsetting (selecting a size range of 58-63).

Meta-transcript Analysis

Meta-transcript analyses were performed on robustly expressed single transcript
genes that had a CDS region larger than 400 nt ($N = 5456$). Firstly, footprint
positions were determined with appropriate A-site estimation (see Estimation of
1375 A-site Positions), then footprint counts were normalized to the total number of
footprints per transcript. Mean normalized footprint densities were plotted for
the first or last 400 nt of CDS plus a small region from the adjacent UTRs. For
analysis of signal peptide (SP) genes, transcripts were annotated as SP or no-SP

based on the Signalp protein feature from Ensembl Database v91. To calculate
1380 the probability densities of length normalized proportions of footprints within
the first 75 codons and the rest of CDS, for each transcript, footprints within
each portion were counted separately and normalized to library size as usual
and in addition to the size of their respective counting region. Then length-
normalized counts per region (first 75 codons vs rest of CDS) were expressed
1385 as a proportion to their sums, so that when footprints have similar densities
between the two regions, normalized proportions would be around 0.5. The
analysis was repeated for SP and no-SP genes using either disome or monosome
footprints.

Analysis of Footprint Densities in Relation to Peptide Secondary 1390 Structures

An in-house Python script was used to extract annotated secondary structures of
peptides (UniProt Database (UniProtConsortium, 2018) release-2018_06) map-
ping them to the corresponding codon positions along CDS. This informa-
tion is either used for analysis of observed-to-expected ratios (see Observed-
1395 to-Expected Ratios For Proximal Sequence Features), or studying the distribu-
tion of footprint (disome or monosome) densities across regions with pre-defined
structural compositions such as structured-unstructured-structured (s-u-s). To
this end, we have extracted coordinates of regions that included a stretch of
structured (min. 3 aa, up to 30th position), followed by an unstructured stretch
1400 (6 to 30 aa), and finally concluded with a structured stretch (min. 3 aa, up
to 30th position) or the reverse of this configuration (u-s-u) with similar size
restrictions. Positions of normalized footprint peaks (normalized to transcript's
mean footprint count) across regions were scaled to the length of the middle
portion (unstructured portion in the case of s-u-s) and centered to its start,
1405 such that the start and the end of the middle region would correspond to 0
and 1, respectively. Distribution of footprint densities across such regions was
analyzed by kernel density estimates which were weighted with normalized foot-
print peak densities. Significance of density probability functions were evaluated

with randomized sampling. For each transcript, keeping the structures identical, peaks were randomly shuffled ($N = 10000$) and confidence intervals for the kernel densities were calculated. Total RNA reads were similarly analyzed as a control.

Analysis of Evolutionary Conservation at Disome Sites

Evolutionary conservation of sites were evaluated using the PhyloP scores (Polard et al., 2010), that were computed from alignments of 59 vertebrate genomes to the mouse genome (phyloP60way) and were retrievable from the Conservation tracks in the UCSC Genome Browser (Haeussler et al., 2018). When required, the PhyloP scores were stratified in 5 levels : highly accelerated [$5, -5$), accelerated [$-5, -3$), neutral [$-3, 3$), conserved [$3, 5$), highly conserved [$5, \infty$), of which the presence of the first two were negligible within CDS regions. As a control to disome sites, we have devised a random sampling algorithm that would pick random two-codon positions (6mer) (i) at frequencies that were similar to that these 6mers were observed with disome peaks (Figure 3M, P, Supplemental Table S2), (ii) without containing a disome site (disome footprint density \leq mean transcript disome footprint density) and (iii) from the same pool of transcripts that the disome peaks in question were identified from. These criteria ensured to have a control pool that was similar in codon-bias and transcript level conservation to that of the disome set. Test for median difference between PhyloP scores at different codon positions was performed by Dunn's test using the `dunn.test` R package controlling for false discovery rate. For regression analysis of conservation scores, a 6-codon wide window (to correspond to -1, E-, P-, A-, +1, +2 sites, Table S4) was used to collect position-specific conservation scores, and presence or absence of a disome peak (peak density $>$ mean transcript density) at the the corresponding A position of the window iterating over the robustly expressed single transcripts ($N = 6007$). From resulting dataset, a linear model was used for each position to explain the PhyloP scores at that position with three explanatory variables: mean transcript PhyloP score, mean codon PhyloP score and presence/absence of a disome peak at the A site. Effect sizes were

evaluated with Cohen's f^2 using the `sjstats` R package.

1440 **Mapping of Disome Amino Acids onto Protein Three-dimensional Structures**

Structure models for target proteins were downloaded from the Protein Data Bank (PDB, <http://www.rcsb.org/>) or Protein Data Bank Europe (PDBe, www.ebi.ac.uk/pdbe/). Image rendering was performed with PyMol (DeLano, 1445 2002). For proteins where no murine structure was available, data from other related mammals (mostly from *H. sapiens*) was used instead. The identification of the residues with high disome signal in the non-mouse protein was performed by manual comparison of the two protein sequences (the original for mouse and the target from the other mammal).

1450 **Functional Enrichment Analysis of Genes with Prominent Disome Peaks**

Deterministic disome peaks were defined as prominent peaks that were not necessarily a result of high levels translational activity. To identify such peaks, library size normalized disome peaks (normalized peak count > 5) along each 1455 transcript were normalized to the mean monosome count of that transcripts. To avoid very noisy peaks, transcripts that had a mean monosome count fewer than 5 were excluded. For each transcript up to 5 peaks that had the highest monosome-normalized scores were collected. Finally, peaks were sorted in descending order of the normalized scores (Supplemental Table S4). Top 200 genes 1460 (identified by their Ensembl IDs) from this list were subsequently submitted to functional enrichment analysis using the `g:GOST` tool of web-based `g:Profiler` software and database platform (Raudvere et al., 2019). Statistically significantly enriched terms within three Gene Ontology (GO) groups - molecular function, cellular component and biological process - were identified (Supplemental Table S5). False discovery was controlled by the default method, `g:SCS`, 1465 to an experiment-wide threshold of $\alpha=0.05$. As a background, list of all genes identified to have above threshold levels of total RNA, monosome and disome reads in the current study was used ($N = 8626$).

Data Access

¹⁴⁷⁰ The raw sequencing data and processed quantification data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>).

Supplemental Table Titles

Supplemental Table S1: Sequencing and mapping information.

Supplemental Table S2: Amino acid enrichment at disome site (by dicodon).

¹⁴⁷⁵ **Supplemental Table S3:** Transcripts with prominent ('deterministic') disome peaks.

Supplemental Table S4: Modelling of conservation on disome sites.

Supplemental Table S5 Enrichment analyses for top-200 genes from Supplementary Table S3.

1480 **References**

- Charneski, C.A., Hurst, L.D., 2013. Positively charged residues are the major determinants of ribosomal velocity. *PLOS Biology* 11, 1–20.
- Dana, A., Tuller, T., 2012. Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLOS Computational Biology* 8, 1–11.
- 1485 Dao Duc, K., Song, Y.S., 2018. The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. *PLOS Genetics* 14, 1–32.
- Darnell, A.M., Subramaniam, A.R., OShea, E.K., 2018. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Molecular Cell* 71, 229 – 243.e11.
- 1490 DeLano, W.L., 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography* 40, 82–92.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- 1495 Doring, K., Ahmed, N., Riemer, T., Suresh, H.G., Vainshtein, Y., Habich, M., Riemer, J., Mayer, M.P., OBrien, E.P., Kramer, G., Bukau, B., 2017. Profiling ssb-nascent chain interactions reveals principles of hsp70-assisted folding. *Cell* 170, 298 – 311.e20.
- 1500 Du, N.H., Arpat, A.B., De Matos, M., Gatfield, D., 2014. Micrnas shape circadian hepatic gene expression on a transcriptome-wide scale. *eLife* 3, e02510.
- Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L.,
- 1505

- Garcia-Giron, C., Gordon, L., Hourlier, T., Hunt, S., Juettemann, T., Kae-
haeri, A.K., Keenan, S., Komorowska, M., Kulesha, E., Longden, I., Maurel,
T., McLaren, W.M., Muffato, M., Nag, R., Overduin, B., Pignatelli, M.,
Pritchard, B., Pritchard, E., Riat, H.S., Ritchie, G.R.S., Ruffier, M., Schus-
1510 ter, M., Sheppard, D., Sobral, D., Taylor, K., Thormann, A., Trevanion, S.,
White, S., Wilder, S.P., Aken, B.L., Birney, E., Cunningham, F., Dunham, I.,
Harrow, J., Herrero, J., Hubbard, T.J.P., Johnson, N., Kinsella, R., Parker,
A., Spudich, G., Yates, A., Zadissa, A., Searle, S.M.J., 2013. Ensembl 2013.
Nucleic Acids Research 41, D48–D55.
- 1515 Gamble, C.E., Brule, C.E., Dean, K.M., Fields, S., Grayhack, E.J., 2016. Ad-
jacent codons act in concert to modulate translation efficiency in yeast. Cell
166, 679 – 690.
- Gatfield, D., Le Martelot, G., Vejnar, C.E., Gerlach, D., Schaad, O., Fleury-
Olela, F., Ruskeep, A.L., Oresic, M., Esau, C.C., Zdobnov, E.M., Schibler, U.,
1520 2009. Integration of microrna mir-122 in hepatic circadian gene expression.
Genes & Development 23, 1313–1326.
- Gobet, C., Weger, B., Marquis, J., Martin, E., Gachon, F., Naef, F., 2019.
Modeling ribosome dwell times and relationships with trna loading and codon
usage in mammals. bioRxiv (posted May 6, 2019) doi:10.1101/551838.
- 1525 Gutierrez, E., Shin, B.S., Woolstenhulme, C., Kim, J.R., Saini, P., Buskirk, A.,
Dever, T., 2013. eif5a promotes translation of polyproline motifs. Molecular
Cell 51, 35 – 45.
- Guydosh, N., Green, R., 2014. Dom34 rescues ribosomes in 3 untranslated
regions. Cell 156, 950 – 962.
- 1530 Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney,
B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., Gibson, D.,
Diekhans, M., Clawson, H., Casper, J., Barber, G.P., Haussler, D., Kuhn,
R.M., Kent, W., 2018. The UCSC Genome Browser database: 2019 update.
Nucleic Acids Research 47, D853–D858.

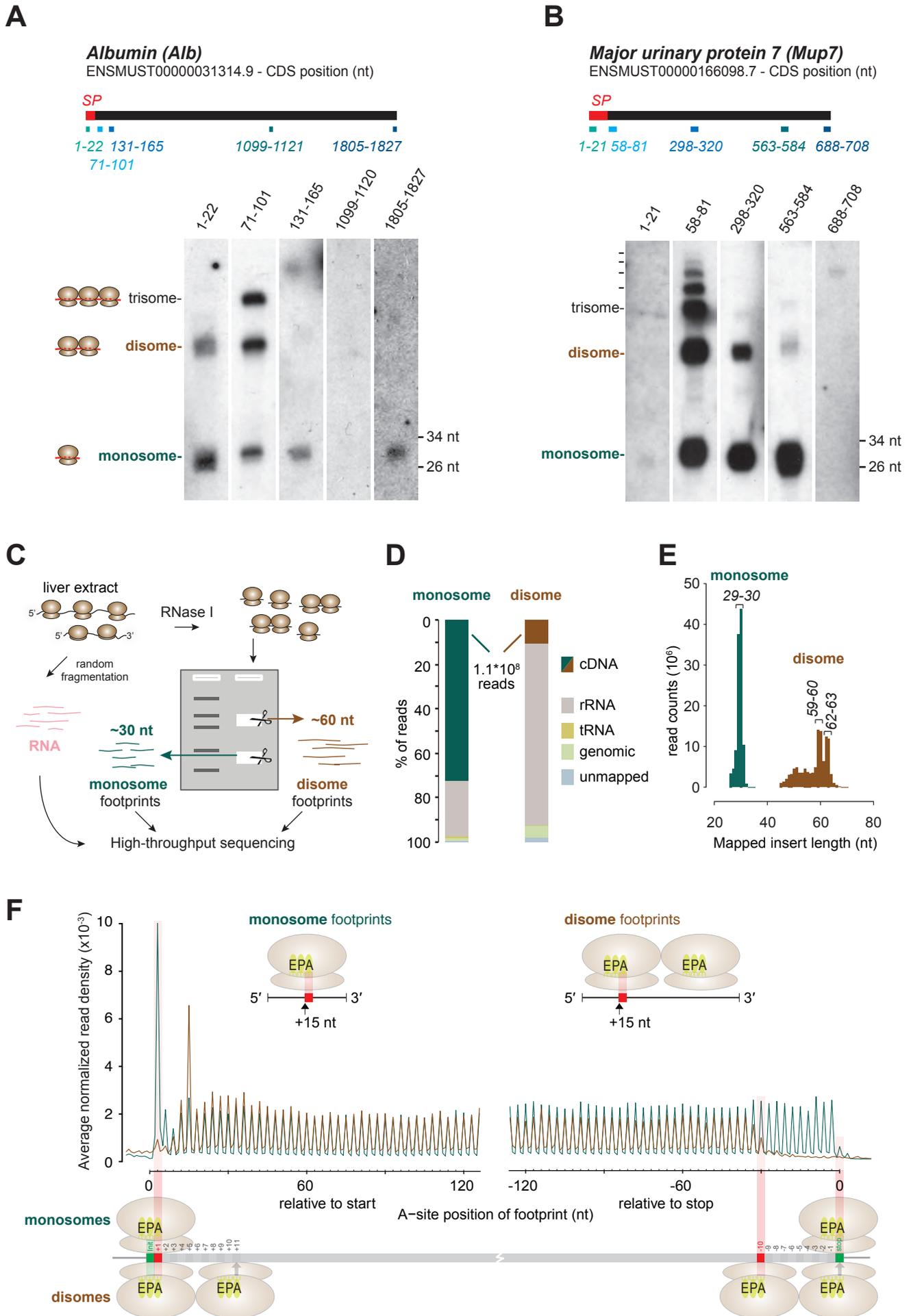
- 1535 Han, Y., Gao, X., Liu, B., Wan, J., Zhang, X., Qian, S., 2014. Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. *Cell Research* 24, 842–851.
- Hinnebusch, A.G., 2014. The scanning mechanism of eukaryotic translation initiation. *Annual Review of Biochemistry* 83, 779–812.
- 1540 Howard, M.T., Carlson, B.A., Anderson, C.B., Hatfield, D.L., 2013. Translational redefinition of uga codons is regulated by selenium availability. *Journal of Biological Chemistry* 288, 19401–19413.
- Ibrahim, F., Maragkakis, M., Alexiou, P., Mourelatos, Z., 2018. Ribothrypsis, a novel process of canonical mrna decay, mediates ribosome-phased mrna endonucleolysis. *Nature Structural & Molecular Biology* 25, 302–310.
- 1545 Ingolia, N., Lareau, L., Weissman, J., 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., Weissman, J.S., 2009. 1550 Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Ingolia, N.T., Hussmann, J.A., Weissman, J.S., 2019. Ribosome profiling: Global views of translation. *Cold Spring Harbor Perspectives in Biology* 11, a032698.
- 1555 Ivanov, I.P., Shin, B.S., Loughran, G., Tzani, I., Young-Baird, S.K., Cao, C., Atkins, J.F., Dever, T.E., 2018. Polyamine control of translation elongation regulates start site selection on antizyme inhibitor mrna via ribosome queuing. *Molecular Cell* 70, 254 – 264.e6.
- Janich, P., Arpat, A., Castelo-Szekely, V., Lopes, M., Gatfield, D., 2015. 1560 Ribosome profiling reveals the rhythmic liver translome and circadian clock regulation by upstream open reading frames. *Genome Research* 25, 1848–1859.

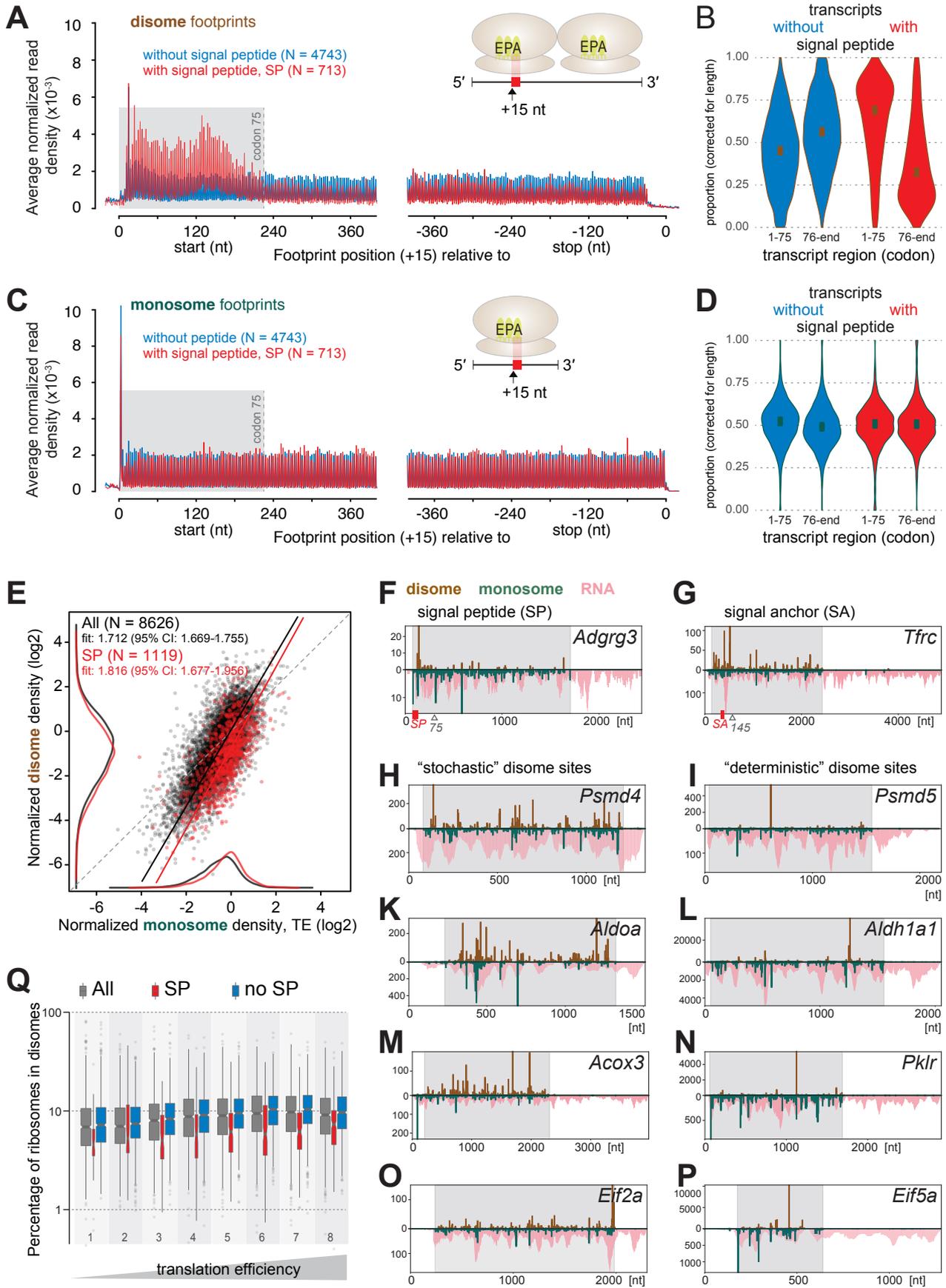
- Janich, P., Arpat, A.B., Castelo-Szekely, V., Gatfield, D., 2016. Analyzing the temporal regulation of translation efficiency in mouse liver. *Genomics Data* 8, 41 – 44.
- 1565
- Joazeiro, C.A.P., 2019. Mechanisms and functions of ribosome-associated protein quality control. *Nature Reviews Molecular Cell Biology* 20, 368–383.
- Johnstone, T., Bazzini, A., Giraldez, A., 2016. Upstream orfs are prevalent translational repressors in vertebrates. *EMBO J* 35, 706–723.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods* 9, 357–U54.
- 1570
- Lesnik, C., Golani-Armon, A., Arava, Y., 2015. Localized translation near the mitochondrial outer membrane: An update. *RNA Biology* 12, 801–809.
- Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. Viennarna package 2.0. *Algorithms for Molecular Biology* 6, 26.
- 1575
- Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., Hegde, R.S., 2010. A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* 466, 1120–1124.
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12.
- 1580
- Melnikov, S., Mailliot, J., Shin, B.S., Rigger, L., Yusupova, G., Micura, R., Dever, T.E., Yusupov, M., 2016. Crystal structure of hypusine-containing translation factor eif5a bound to a rotated eukaryotic ribosome. *Journal of Molecular Biology* 428, 3570 – 3576.
- 1585
- O'Connor, P.B.F., Andreev, D.E., Baranov, P.V., 2016. Comparative survey of the relative impact of mrna features on local ribosome profiling read density. *Nature Communications* 7, 12915.

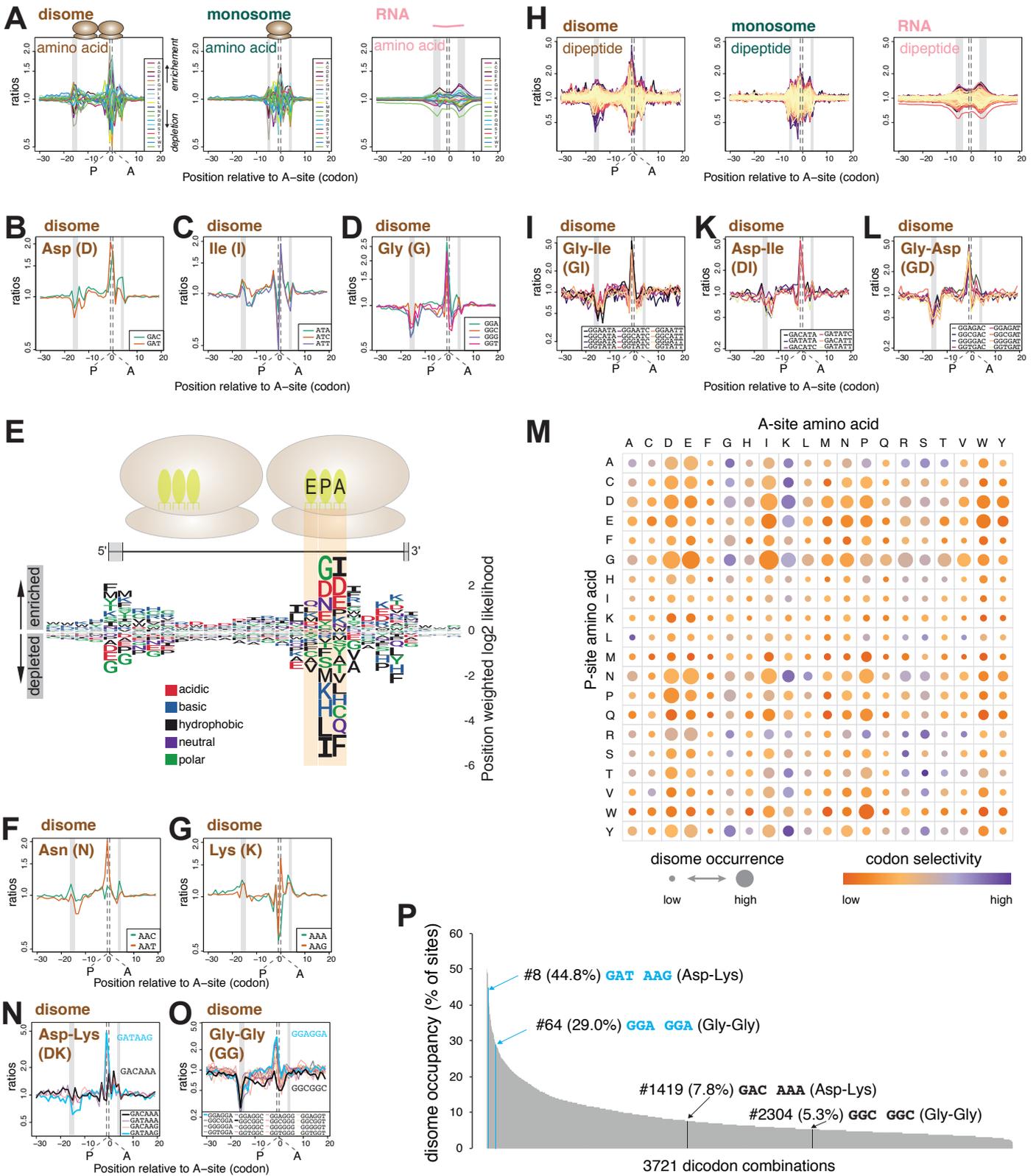
- Panasenko, O.O., Somasekharan, S.P., Villanyi, Z., Zagatti, M., Bezrukov, F.,
1590 Rashpa, R., Cornut, J., Iqbal, J., Longis, M., Carl, S.H., Pea, C., Panse,
V.G., Collart, M.A., 2019. Co-translational assembly of proteasome subunits
in not1-containing assembliesomes. *Nature Structural & Molecular Biology*
26, 110–120.
- Pelechano, V., Alepuz, P., 2017. eIF5A facilitates translation termination glob-
1595 ally and promotes the elongation of many non polyproline-specific tripeptide
sequences. *Nucleic Acids Research* 45, 7326–7338.
- Pertea, M., Pertea, G., Antonescu, C., Chang, T., Mendell, J., Salzberg, S.,
2015. Stringtie enables improved reconstruction of a transcriptome from rna-
seq reads. *Biotechnology* 33, 290–295.
- 1600 Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., Siepel, A., 2010. Detection of
nonneutral substitution rates on mammalian phylogenies. *Genome research*
20, 110–121.
- Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S.,
Koller, D., 2014. Causal signals between codon bias, mrna structure, and the
1605 efficiency of translation and elongation. *Molecular Systems Biology* 10, 770.
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo,
J., 2019. g:Profiler: a web server for functional enrichment analysis and
conversions of gene lists (2019 update). *Nucleic Acids Research* 47, W191–
W198.
- 1610 Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a bioconductor
package for differential expression analysis of digital gene expression data.
Bioinformatics 26.
- Salmon, P., Trono, D., 2007. Production and titration of lentiviral vectors. *Curr
Protoc Hum Genet* Chapter 12, Unit 12.10.
- 1615 Schmidt, C., Becker, T., Heuer, A., Braunger, K., Shanmuganathan, V., Pech,
M., Berninghausen, O., Wilson, D.N., Beckmann, R., 2015. Structure of the

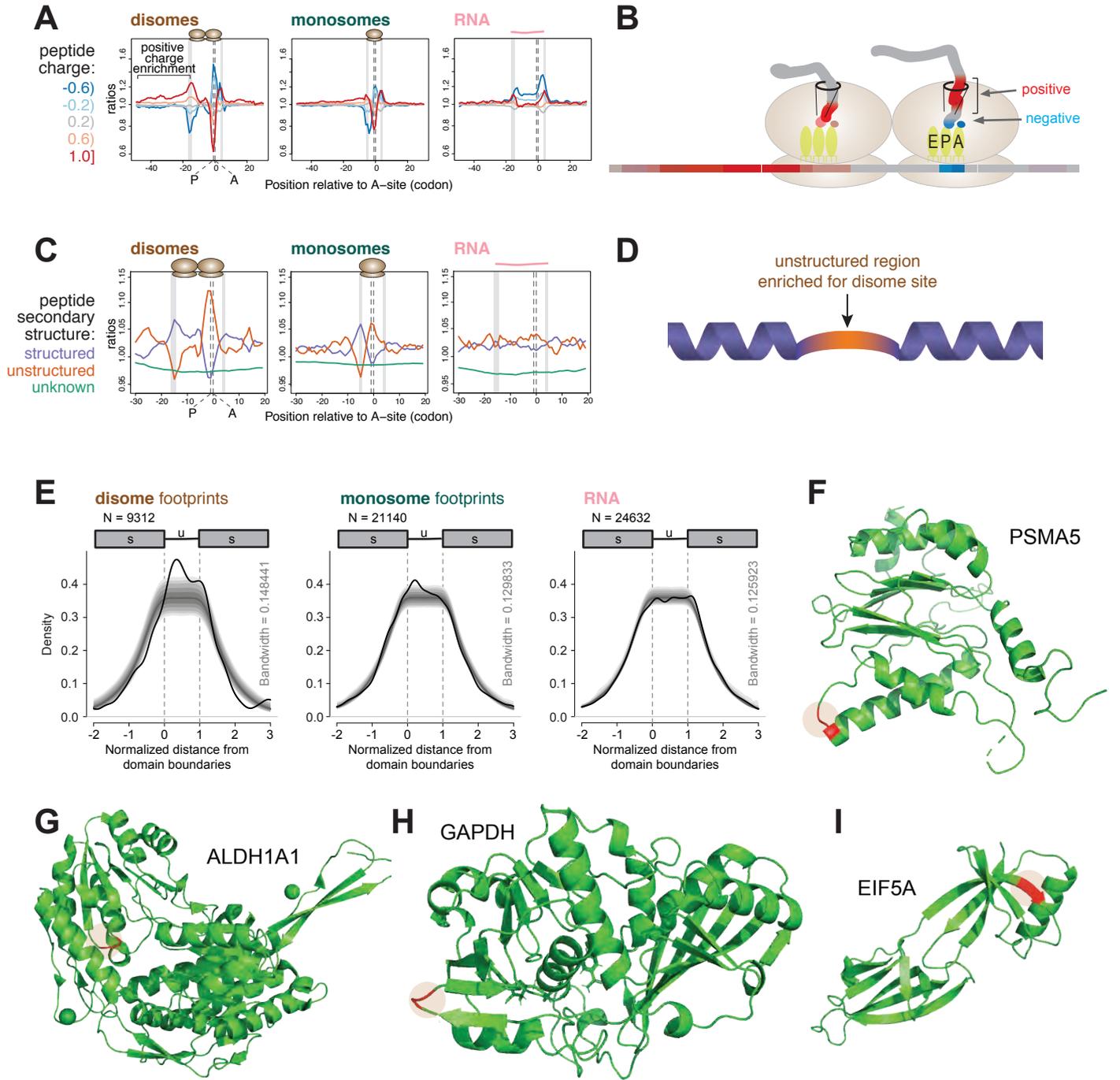
- hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome. *Nucleic Acids Research* 44, 1944–1951.
- Schuller, A.P., Green, R., 2018. Roadblocks and resolutions in eukaryotic translation. *Nature* 19, 526–541.
- 1620 Schuller, A.P., Wu, C.C.C., Dever, T.E., Buskirk, A.R., Green, R., 2017. eif5a functions globally in translation elongation and termination. *Molecular Cell* 66, 194 – 205.e5.
- Shiber, A., Doring, K., Friedrich, U., Klann, K., Merker, D., Zedan, M., Tippmann, F., Kramer, G., Bukau, B., 2018. Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature* 561, 268–272.
- 1625 Sinturel, F., Gerber, A., Mauvoisin, D., Wang, J., Gatfield, D., Stubblefield, J.J., Green, C.B., Gachon, F., Schibler, U., 2017. Diurnal oscillations in liver mass and cell size accompany ribosome assembly cycles. *Cell* 169, 651–663.
- 1630 Stoytcheva, Z., Tujebajeva, R.M., Harney, J.W., Berry, M.J., 2006. Efficient incorporation of multiple selenocysteines involves an inefficient decoding step serving as a potential translational checkpoint and ribosome bottleneck. *Molecular and Cellular Biology* 26, 9177–9184.
- UniProtConsortium, T., 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* 46, 2699–2699.
- 1635 Wagih, O., 2017. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647.
- Wolin, S., Walter, P., 1988. Ribosome pausing and stacking during translation of a eukaryotic mrna. *EMBO J* 7, 3559–3569.
- 1640 Wu, C.C.C., Zinshteyn, B., Wehner, K.A., Green, R., 2019. High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Molecular Cell* 73, 959 – 970.e5.

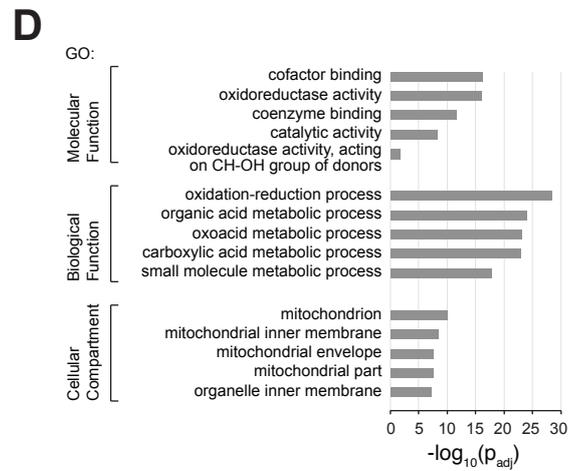
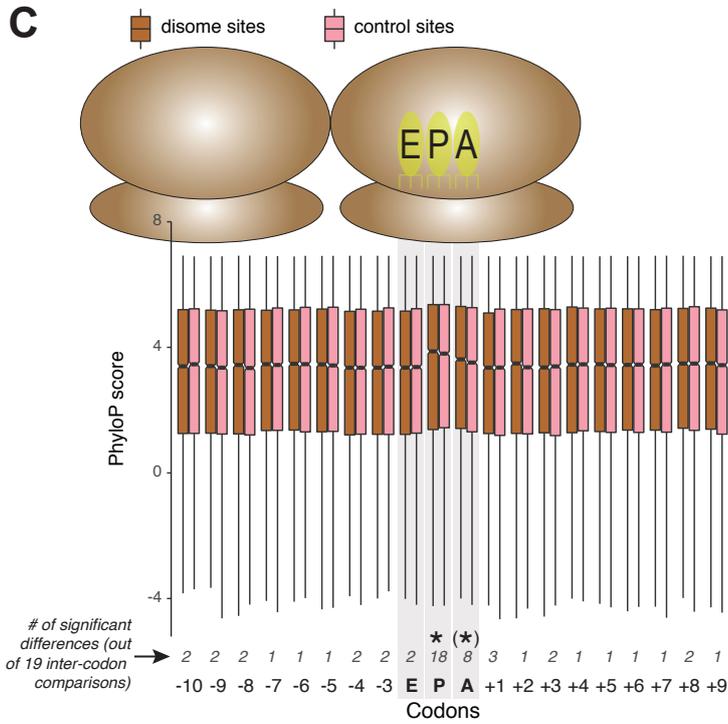
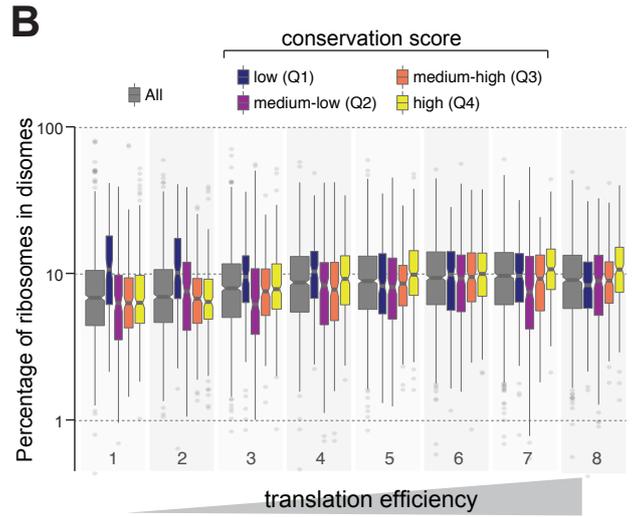
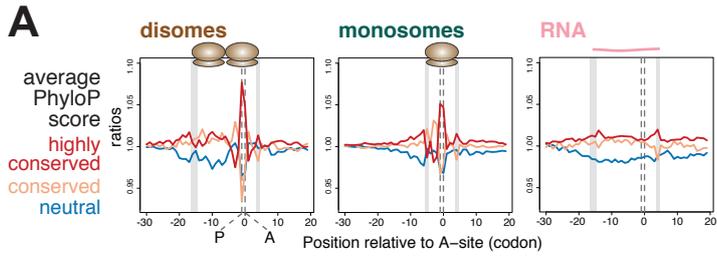
- 1645 Yanagitani, K., Kimata, Y., Kadokura, H., Kohno, K., 2011. Translational pausing ensures membrane targeting and cytoplasmic splicing of xbp1u mrna. *Science* 331, 586–589.
- Yordanova, M.M., Loughran, G., Zhdanov, Alexander V.and Mariotti, M., Kiniry, S.J., OConnor, P.B.F., Andreev, D.E., Tzani, I., Saffert, P., Michel, A.M., Gladyshev, V.N., Papkovsky, D.B., Atkins, J.F., Baranov, P.V., 2018. Amd1 mrna employs ribosome stalling as a mechanism for molecular memory formation. *Nature* 553, 356–360.
- 1650 Zerial, M., Melancon, P., Schneider, C., Garoff, H., 1986. The transmembrane segment of the human transferrin receptor functions as a signal peptide. *EMBO J* 5, 1543–1550.
- Zhang, S., Hu, H., Zhou, J., He, X., Jiang, T., Zeng, J., 2017. Analysis of ribosome stalling and translation elongation dynamics by deep learning. *Cell Systems* 5, 212 – 220.e6.

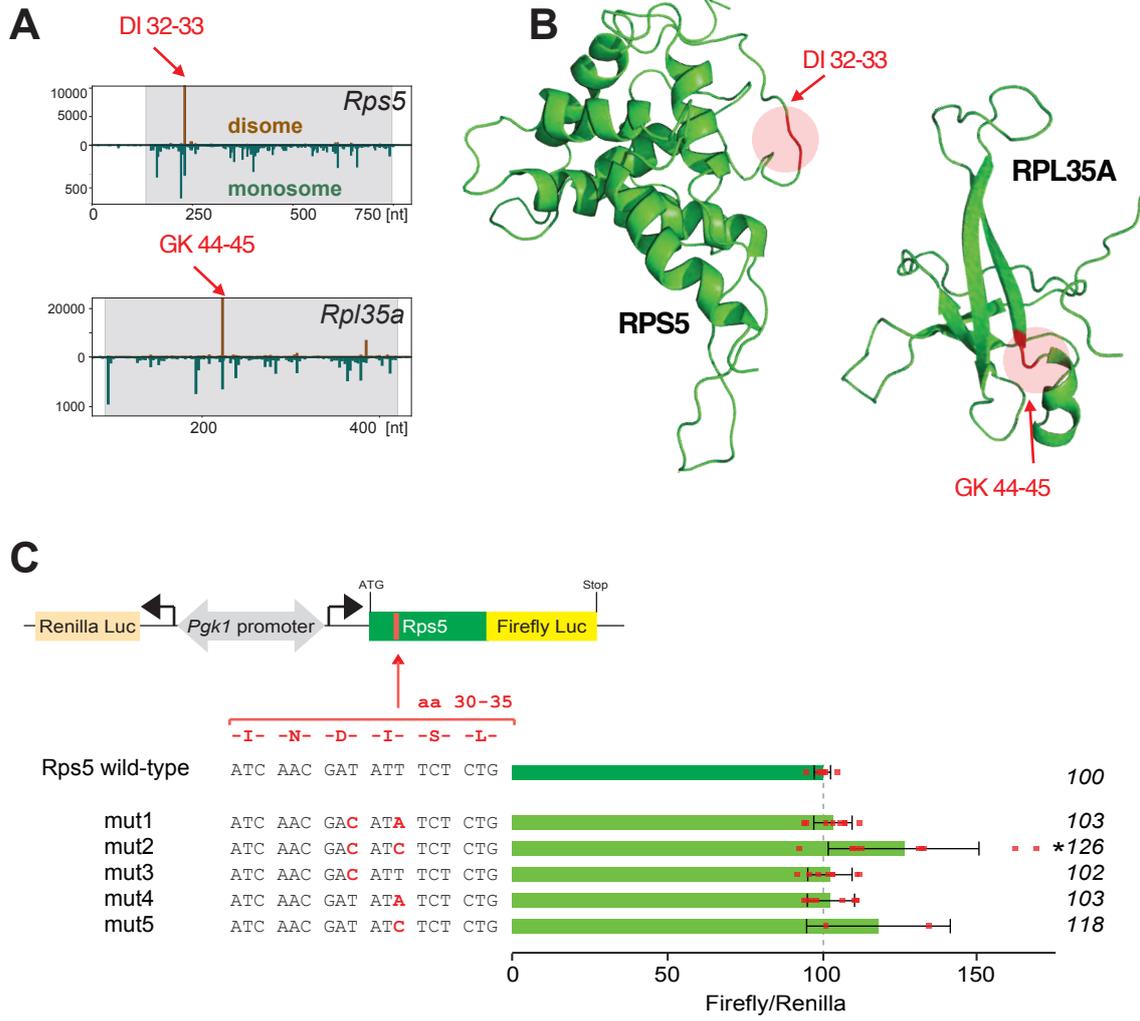




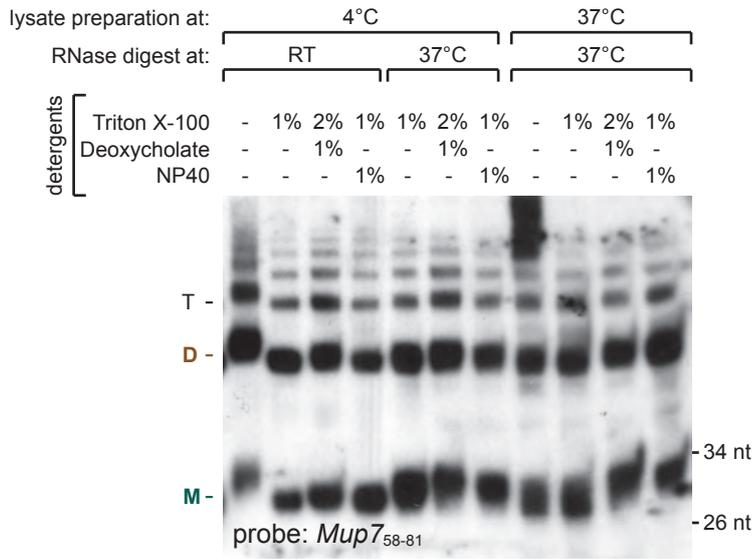




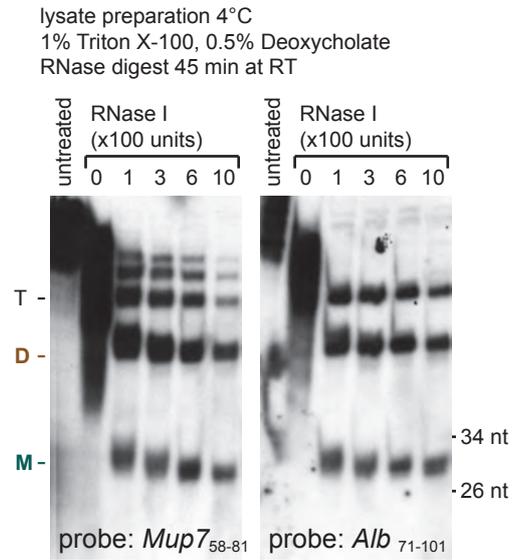




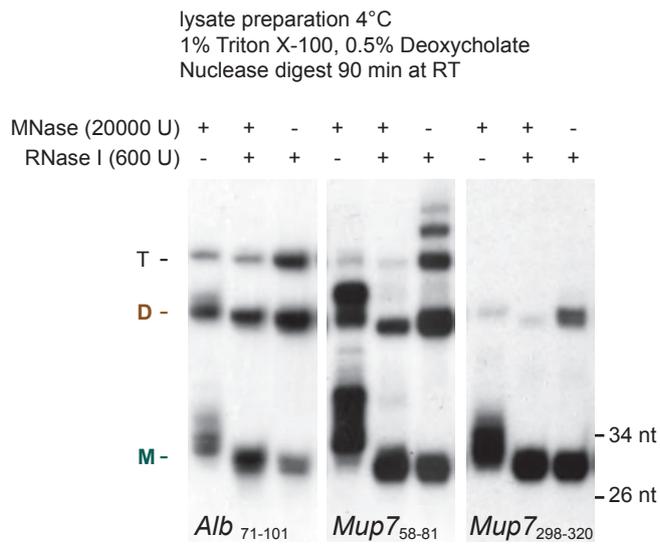
A

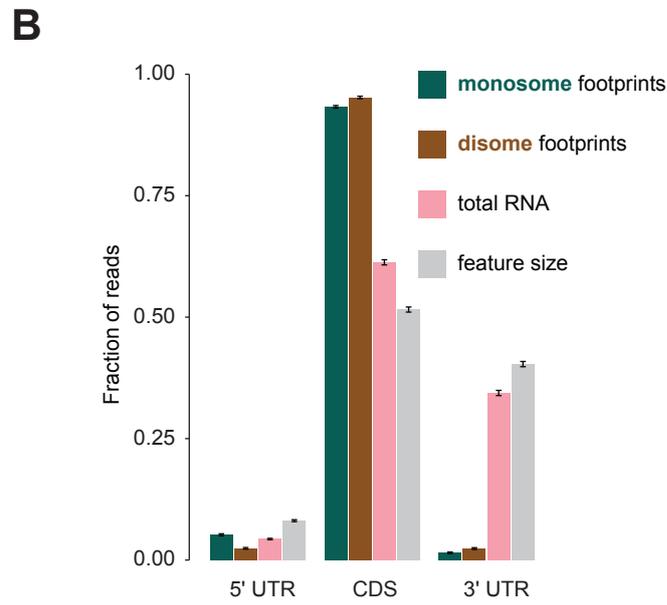
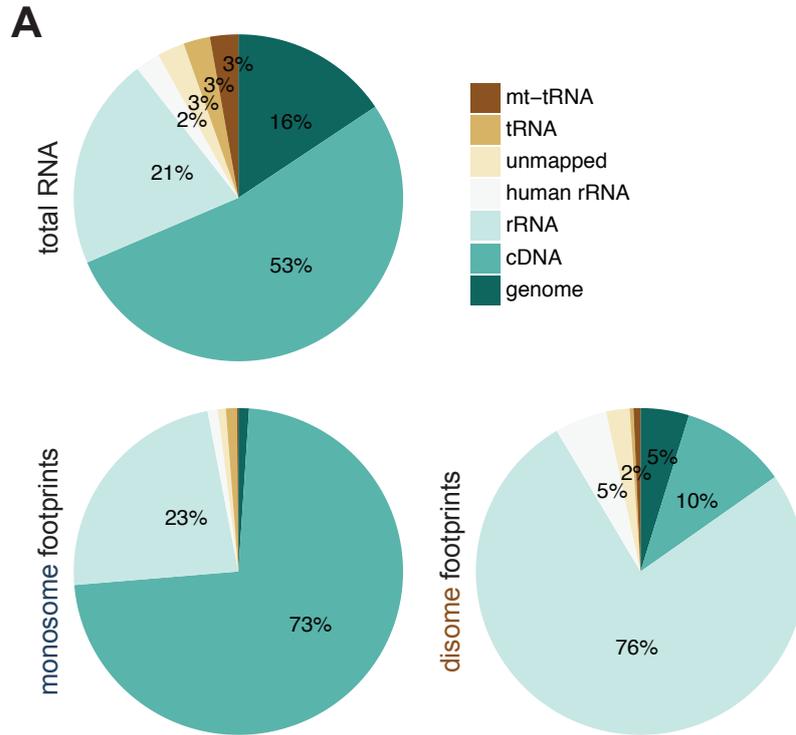


B

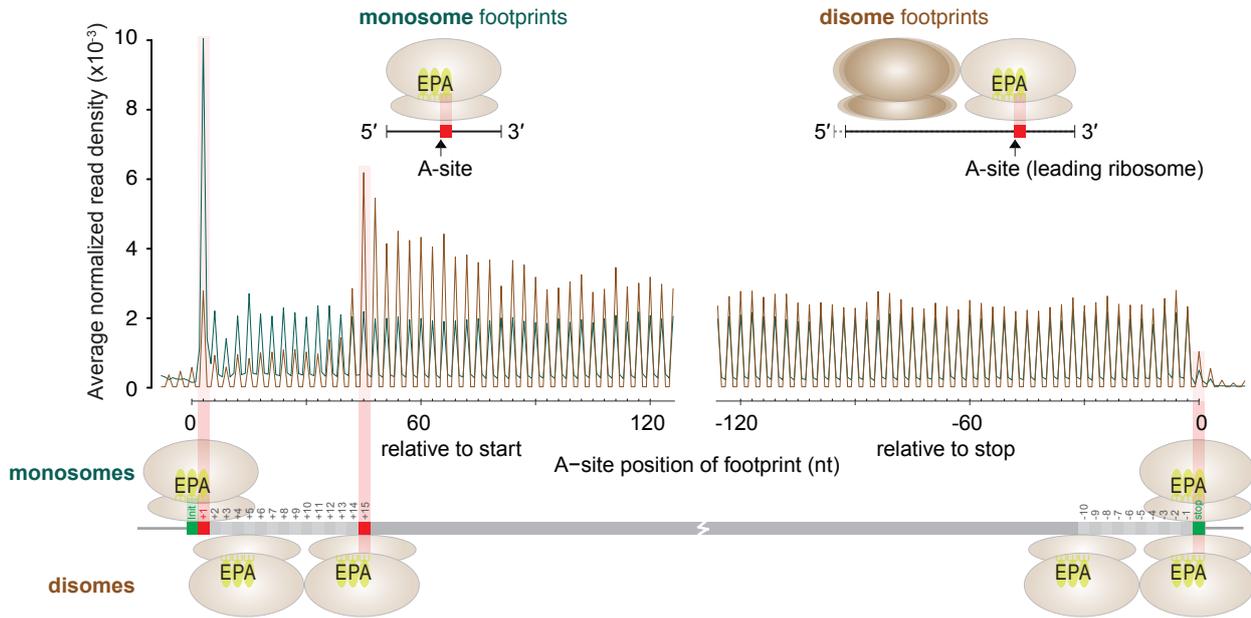


C

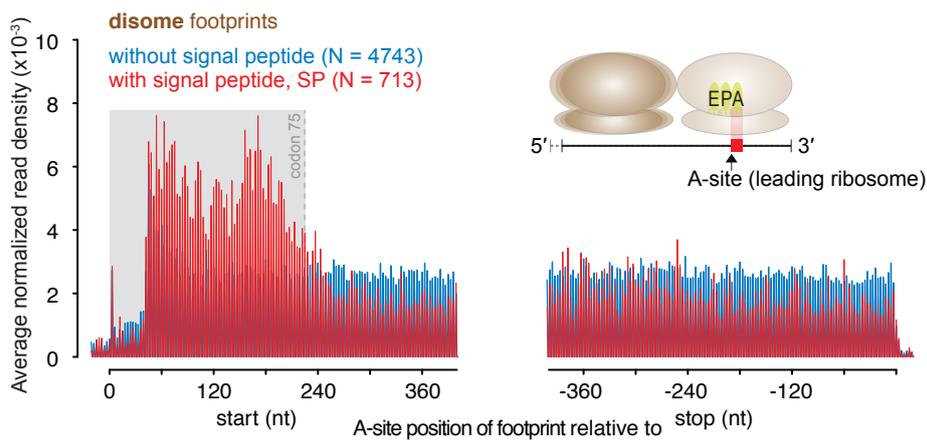




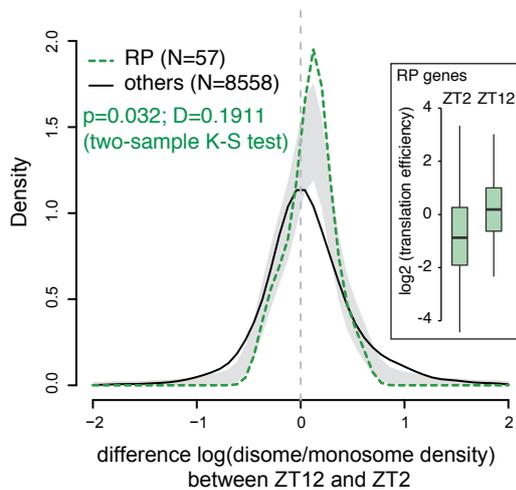
A



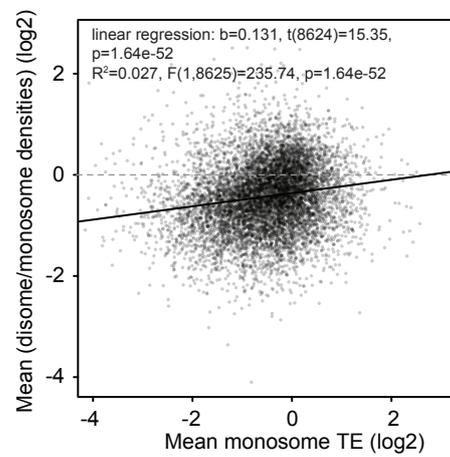
B

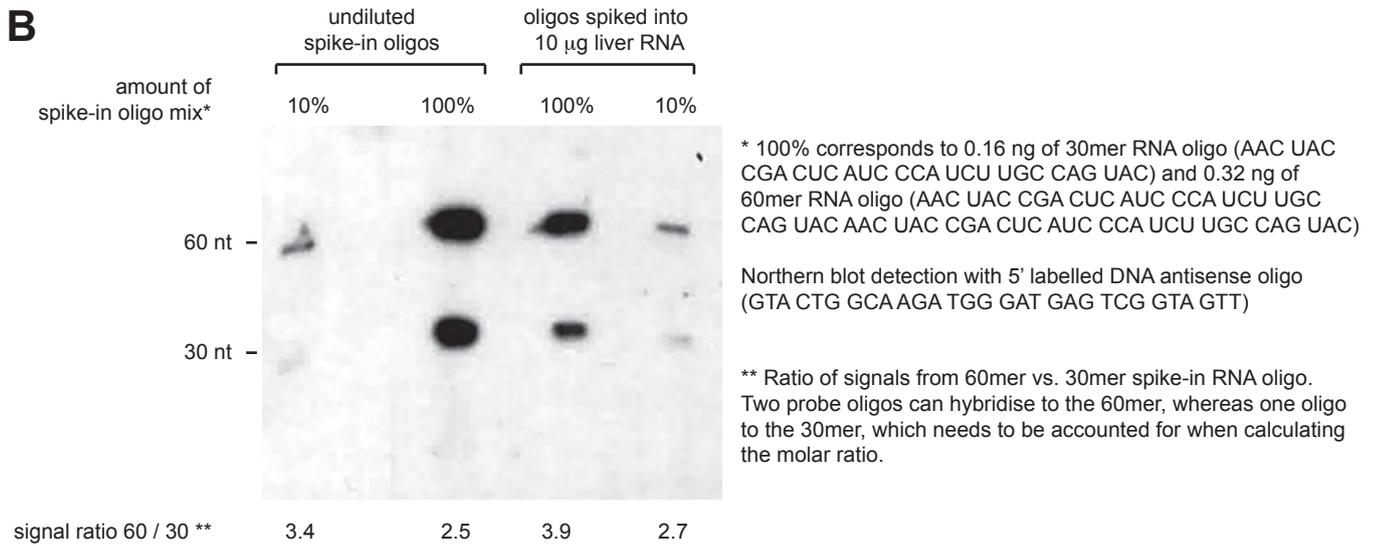
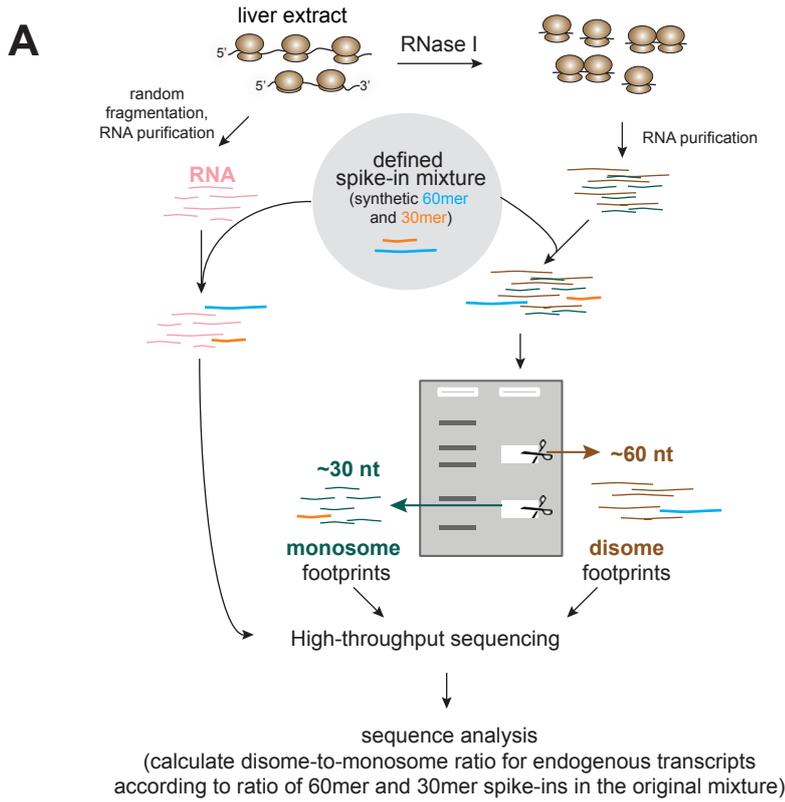


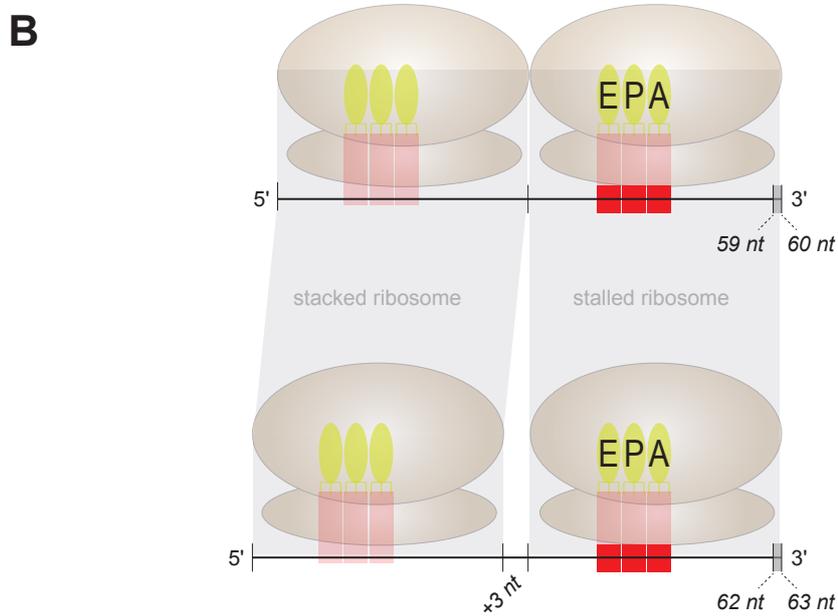
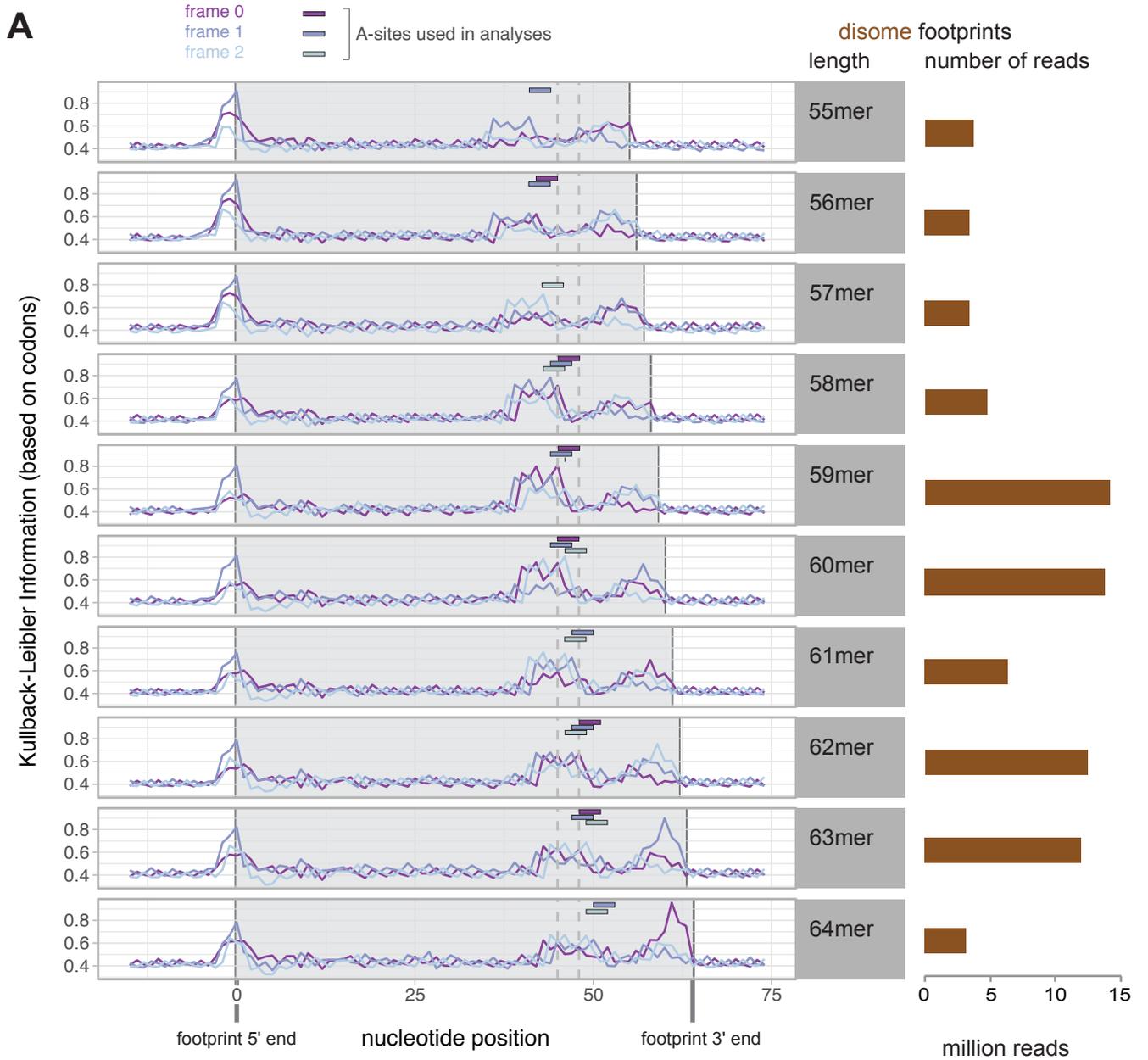
C

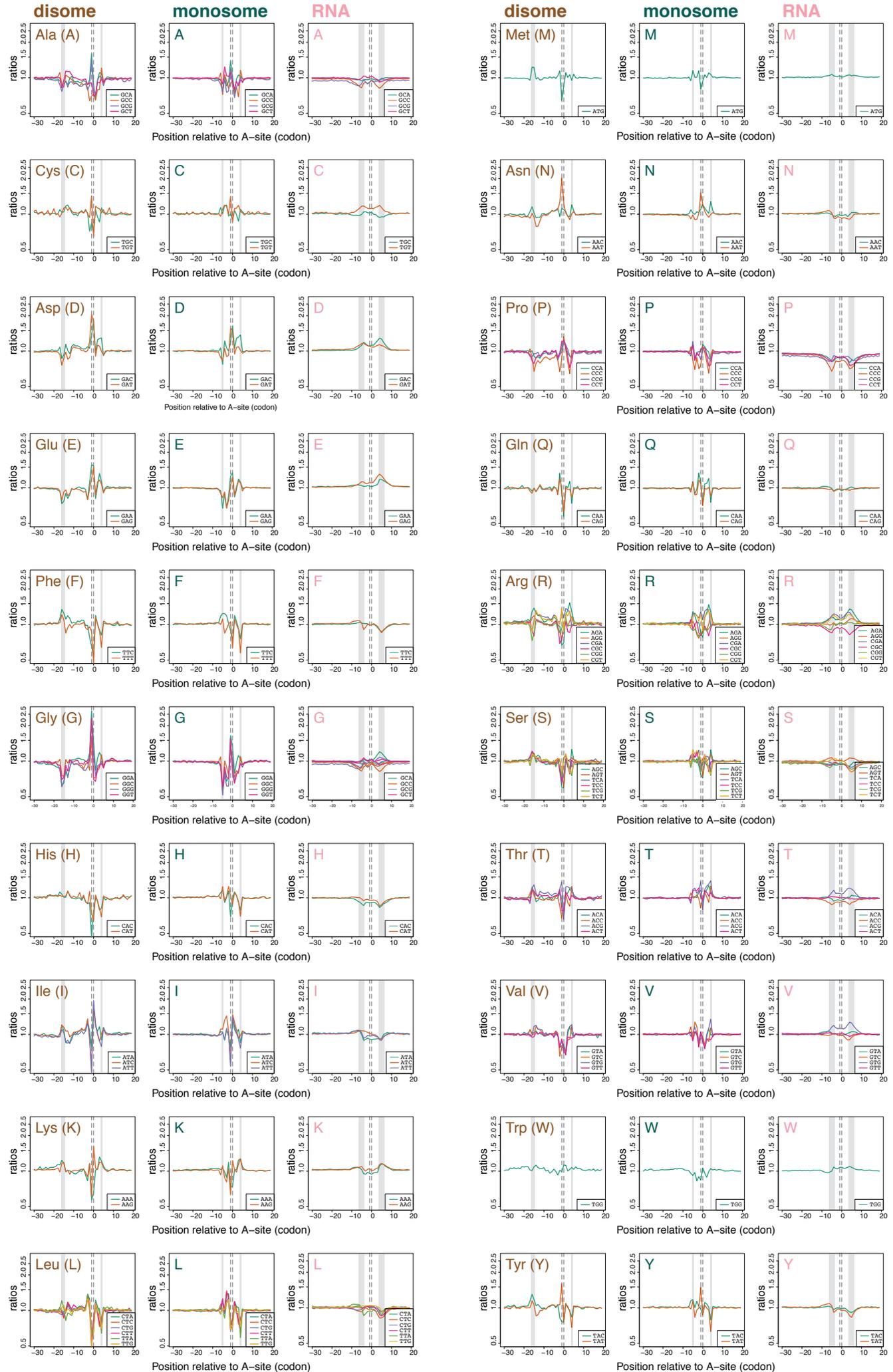


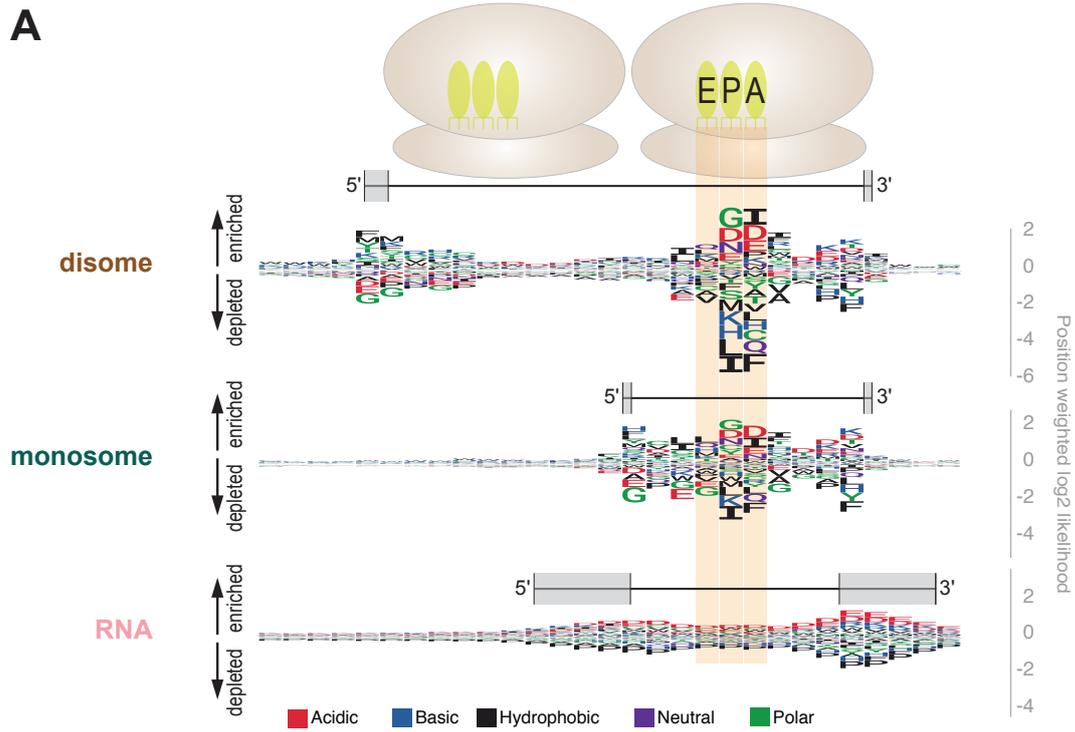
D



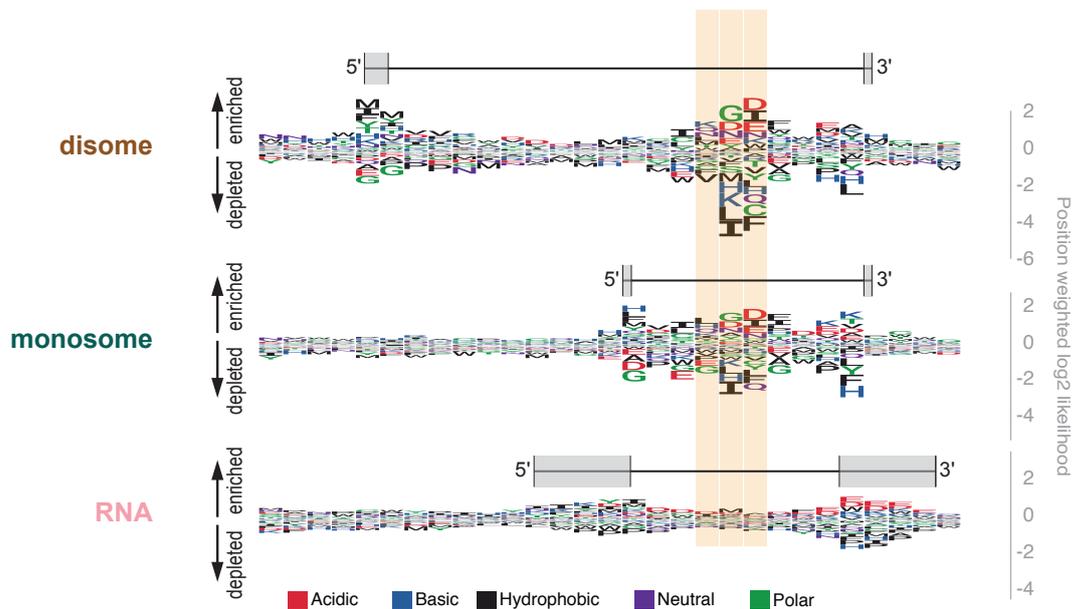






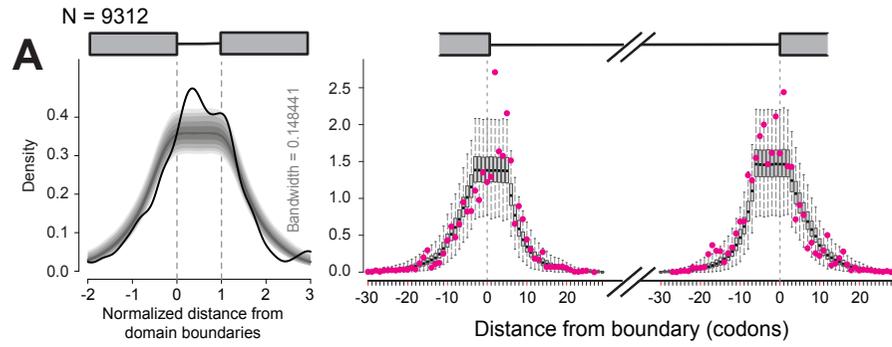


B SP-transcripts (analysis over codons 8-75)

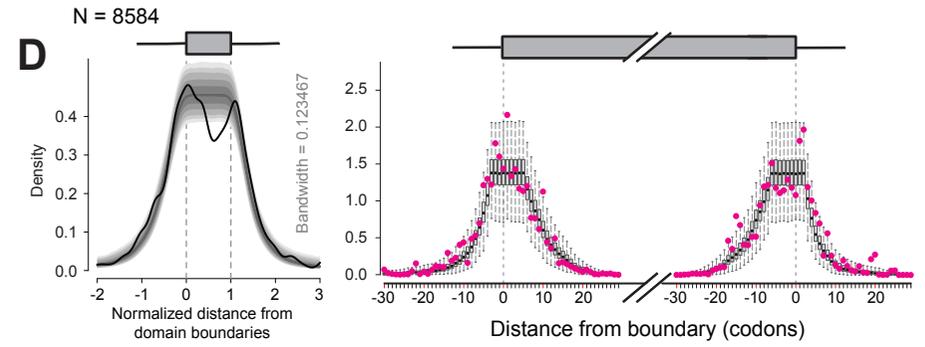


disome footprints

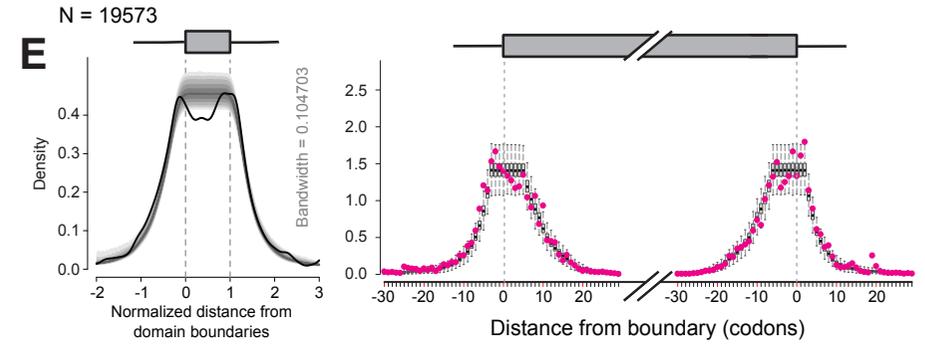
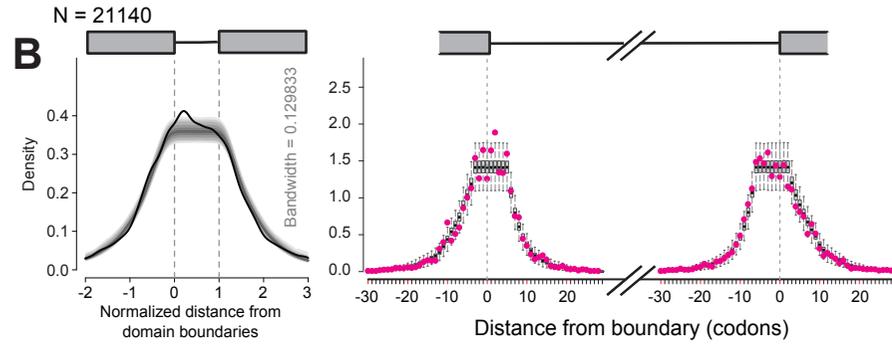
standard configuration



reverse configuration



monosome footprints



total RNA reads

