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

Alaaddin Bulak Arpat,[1,2] Angélica Liechti,[1] Mara De Matos,[1] René Dreos,[1,2] Peggy Janich,[1] and David Gatfield[1]

[1]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland; [2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Translation initiation is the major regulatory step defining the rate of protein production from an mRNA. Meanwhile, the impact of nonuniform 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 is indicative of deterministic pausing signals. Pause site association with specific amino acids, peptide motifs, and nascent polypeptide structure is suggestive of programmed pausing as a widespread mechanism associated with protein folding. Evolutionary conservation at disome sites indicates functional relevance of translational pausing. Collectively, our disome profiling approach allows unique insights into gene regulation occurring at the step of translation elongation.

[Supplemental material is available for this article.]

The translation of messenger RNA (mRNA) to protein is a central step in gene expression. Knowledge of this process has exploded since the emergence of ribosome profiling (Ribo-seq), a technique based on the high-throughput sequencing of the ∼30-nt mRNA footprints that are buried inside the translating ribosome and protected from the nuclease treatment used to digest the mRNA regions that are not occupied by ribosomes (Ingolia et al. 2009). A plethora of studies have built on the quantitative, transcriptome-wide, and nucleotide-resolved information that Ribo-seq provides to gain insight into a variety of aspects of protein biosynthesis (for review, see Ingolia et al. 2019). This includes the annotation of coding sequences, the study of differential translation, 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 Ribo-seq data supports the longstanding notion that, of the four distinct phases defining translation (initiation, elongation, termination, ribosome recycling), the commitment of the ribosome to initiate is rate-limiting for the overall process (Hinnebusch 2014). It is assumed that the quantity of elongating ribosome footprints (i.e., the species mainly captured by conventional Ribo-seq methodology) is proportional to initiation rate and to overall protein biosynthesis. Elongating ribosome footprint distribution across protein-coding sequences (CDS) is distinctly nonuniform, which has been attributed to variations in ribosome decoding speed and dwell times (Ingolia et al. 2011). Integrating footprint reads across the CDS is thought to correct for local variation in footprint density, allowing for accurate estimates of relative translation efficiencies per gene (TEs, calculated as CDS-mapping footprint reads normalized to RNA abundance). 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. It remains an intrinsic limit of the technique that it delivers static snapshots 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/no protein biosynthesis), and of another transcript with strong, productive flux of elongating ribosomes (high protein biosynthesis), the actual footprint snapshots that would be seen by Ribo-seq 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 2014; Diament et al. 2018).

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 been shown to be involved in protein localization to membranes (Mariappan et al. 2010; Yanagitani et al. 2011), in start codon selection (Ivanov et al. 2018), and in the regulation of full-length protein production (Yordanova et al. 2018). It is tempting to extrapolate from such individual examples to general roles for elongation pausing that cells could 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

are sensed and handled is a highly active field of current research (for review, see Joazeiro 2019).

An early Ribo-seq study in mouse embryonic stem cells (mESCs) already addressed the question of how to extract potential pause sites from footprint data, identifying thousands of alleged pauses within CDS sequences and on termination codons (Ingolia et al. 2011). In combination with quantitative modeling approaches, subsequent studies have identified parameters that can impinge on local translation speed and pausing (for review, see Schuller and Green 2018). These include specific amino acids (Charneski and Hurst 2013), codon pairs (Gamble et al. 2016), tRNA availability (Guydosh and Green 2014; Darnell et al. 2018), RNA secondary structures (Pop et al. 2014; Zhang et al. 2017), and nascent peptide folding (Döring et al. 2017) and exit tunnel interactions (Charneski and Hurst 2013; Dao Duc and Song 2018). However, to what extent translational pausing occurs in vivo in a mammalian system, what the pause site characteristics are, 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, to directly reveal the sites where two ribosomes collide. The deep analysis of these ~60-nt "disome footprints" provides insights into elongation pausing transcriptome-wide.

Results

Disome footprint sequencing allows transcriptome-wide mapping of ribosomal collisions

A critical step in ribosome profiling is the quantitative conversion of polysomes down to individual, footprint-protecting monosomes. For the setup of Ribo-seq in mouse liver for a previous study (Janich et al. 2015), we monitored the efficiency of RNase I-mediated footprint generation by northern blot. Radioactively labeled oligonucleotide probes antisense to the highly abundant albumin (*Alb*) and major urinary protein 7 (*Mup7*) mRNAs indeed revealed the expected ~30-nt monosome footprints (Fig. 1A,B). Moreover, some probes detected additional higher-order bands whose estimated sizes corresponded to multiples of monosome footprints (i.e., ~60 nt, ~90 nt, etc.). These bands were particularly prominent with probes annealing to the CDS just downstream of where the signal peptide (SP) was encoded (see two and approximately five higher-order bands for probes Alb_{71-101} and $Mup7_{58-81}$, respectively). We initially suspected that suboptimal nuclease treatment had caused the incomplete collapse of polysomes to monosomes and hence tested other reaction conditions. However, neither modified temperature or detergent concentrations during extract preparation and nuclease treatment (Supplemental Fig. S1A), nor higher RNase I activity (Supplemental Fig. S1B), nor a different nuclease altogether, micrococcal nuclease (Supplemental Fig. S1C), were able to quantitatively collapse the higher-order bands to monosome footprints. 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 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 queuing of subsequent incoming ribosomes, was related to the recruitment of the signal recognition particle (SRP) to the SP. We wished to determine whether our higher-order footprints reflected a similar phenomenon and would allow detecting ribosomal pause and collision sites transcriptome-wide and in vivo. We selected a subset of samples from our previously collected mouse liver time series (Janich et al. 2015), corresponding to three time points at the beginning of the daily light (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 (Fig. 1C). Libraries were sequenced sufficiently deeply to obtain $>10^8$ cDNA-mapping reads per footprint species (Fig. 1D; Supplemental Fig. S2A; Supplemental Table S1). Monosome footprints showed the expected length and mapping features, i.e., the majority were 29–30 nt in size (Fig. 1E), and they were enriched on CDS and depleted from untranslated regions (UTRs) (Supplemental Fig. S2B). The observed stronger depletion from 3′ UTRs than from 5′ UTRs was expected given that 5′ UTRs harbor considerable translational activity on upstream open reading frames (uORFs). Disome footprints showed two distinct length populations at 59–60 nt and 62–63 nt (Fig. 1E) that resembled the bimodal pattern that has been observed in yeast (Guydosh and Green 2014). The mapping to transcript regions was similar to that of monosome footprints, albeit with a stronger depletion from 5′ UTRs (Supplemental Fig. 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 translating ribosomes simultaneously. Reduced levels in 5′ UTR disome footprints were thus consistent with the hypothesis that they reflected ribosomal collisions.

We next analyzed footprint frame preference and distribution along the CDS. To this end, we mapped the predicted ribosomal aminoacyl-tRNA acceptor site (A-site) codon of each monosome footprint (i.e., nucleotides 15–17 for 29- to 30-nt footprints) (see Janich et al. 2015) onto the metatranscriptome. 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 placed over the +1 codon) and ending at the termination codon (Fig. 1F). Moreover, the profile showed previously reported features, including elevated and reduced ribosome densities at the start and stop codons, respectively, as well as increased occupancy of the +5 codon, which has been interpreted to reflect a pause occurring between initiation and elongation commitment (Han et al. 2014). For an equivalent analysis on disome footprints, we aligned them to the CDS according to their +45-nt position, corresponding to the alleged A-site of the leading ribosome (Fig. 1F; see Supplemental Fig. S3A for a +15 alignment). Disome footprints also showed transcriptome-wide 3-nt periodicity. At the CDS 3′ end, footprint coverage ended at the position expected when the disome's leading ribosome would occupy the termination codon. Maximal disome footprint abundance was found near the CDS 5′ end, at a position corresponding to a disome formed from lagging and leading ribosomes on the +5 and +15 codons, respectively (Fig. 1F; Supplemental Fig. S3A). Further upstream, i.e., on the first few codons post-initiation, disomes were distinctly depleted. Distribution over the remainder of the CDS was overall rather uniform, with some 5′-to-3′ decrease. Finally, a small, local increase in footprints that would correspond to a leading ribosome on the initiation codon and a lagging ribosome at the −10 codon in the annotated 5′ UTR presumably reflected translated uORF codons (marked with a blue arrow in Supplemental Fig. S3A).
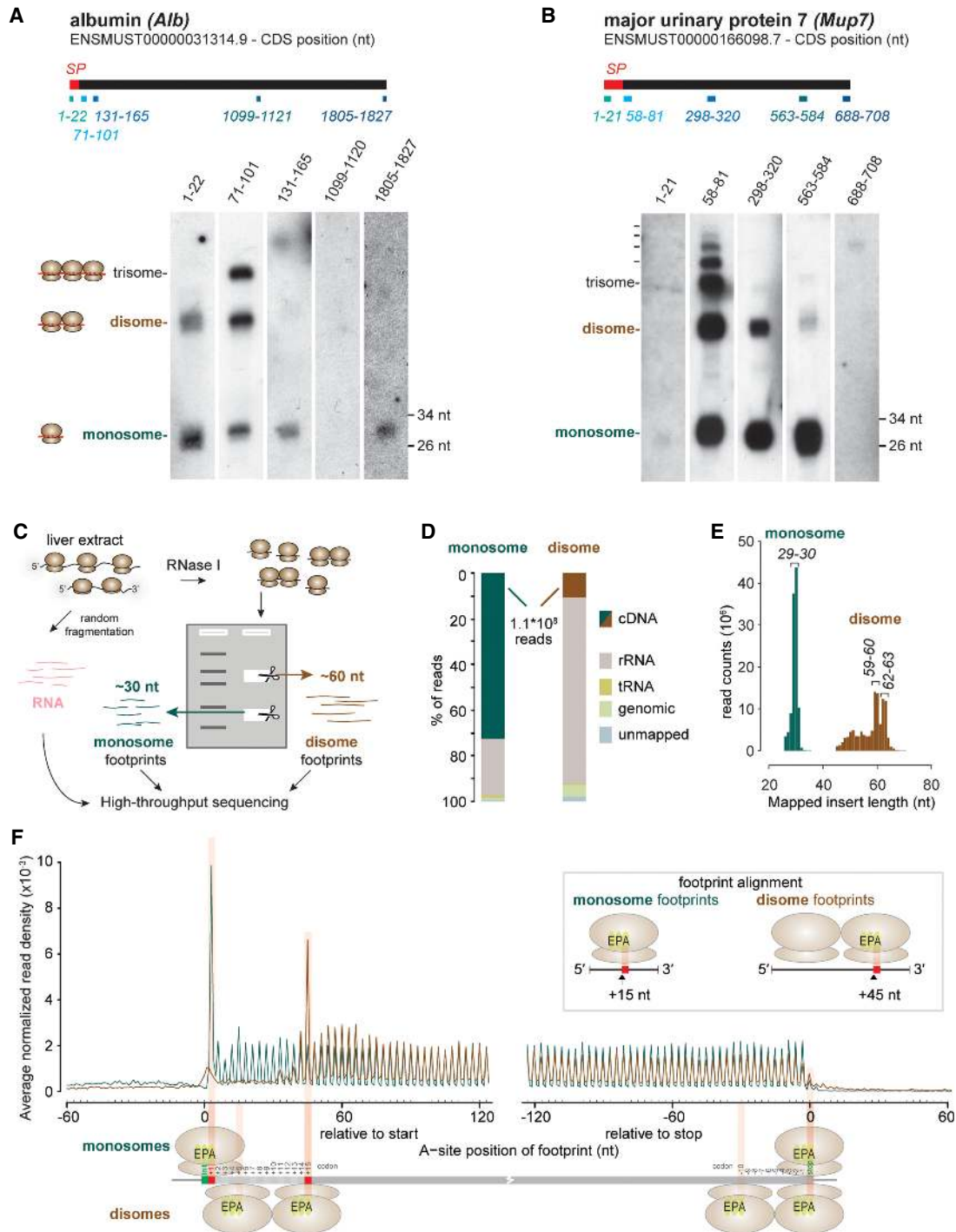
Figure 1. Sequencing of disome footprints identifies transcriptome-wide ribosomal collisions. (A,B) Northern blot analysis of RNase I-treated mouse liver extracts using probes antisense to *Alb* (A) and *Mup7* mRNA (B). Expected footprint sizes for monosomes, disomes, and trisomes are shown to the *left* of blots. Positions of probes (nt) relative to the annotated CDS start sites on the indicated transcripts are shown *above* each lane and depicted as blue boxes *below* the CDS (black bar). The CDS region encoding the signal peptide (SP) is marked in red. (C) Schematic of experimental setup for sequencing of ~60-nt disome footprints. (D) Proportion of reads from monosome and disome libraries that mapped to different sequence types: rRNA (gray), tRNA (golden), genomic (green), and cDNA/mRNA (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/mRNA sequences (monosomes: teal, disomes: brick red). A single mode for monosomes (29–30 nt) and two modes for disomes (59–60 and 62–63 nt) are labeled *above* histograms. (F) Density distribution of footprint reads within 120 nt from the start or −120 nt from the stop codons reveals 3-nt periodicity of footprints within coding sequences. The metatranscript analysis quantified the mean of per-transcript normalized number of reads (monosomes: teal, disomes: brick red) at each nucleotide based on the A-site prediction (15 nt and 45 nt downstream of the 5′ end of monosome and disome footprints, respectively). Transcripts from single protein isoform genes with total RNA-RPKM > 5, CDS > 400 nt, and UTRs of >180 nt (N = 4994) were used. The predicted E-, P-, and A-sites of ribosomes that presumably protected the corresponding footprints are shown in graphical depictions. Start/stop codons are highlighted (green) on a representative transcript *below*.

Taken together, these findings were consistent with the hypothesis that the ~60-nt higher-order bands represented footprints originating from translated mRNA that was protected by two adjacent ribosomes. It would appear that, transcriptome-wide, the alleged ribosomal collisions could occur at most CDS positions, although the likelihood of stacking onto a downstream ribosome would seem reduced immediately post-initiation. Sterical constraints to initiation—for example, extra space that may be required to allow the formation of the initiation complex with its associated initiation factors—have been proposed to play a role in a similar phenomenon in yeast (Guydosh and Green 2014).

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

Given the correlation between pausing and SRP recruitment reported in vitro (Wolin and Walter 1988), we next assessed footprint densities for transcripts encoding signal peptide-containing proteins (SP transcripts; $N = 713$) versus non-SP transcripts ($N = 4743$). SP transcripts showed a distinct buildup of disome footprints toward the 5′ end of the CDS that extended virtually to the codon 75 position Wolin and Walter had described for preprolactin mRNA, whereas downstream of this region, disome densities were reduced (Fig. 2A,B; Supplemental Fig. S4). These features were absent from monosome footprint data (Fig. 2C,D). We concluded that disome footprint profiling was able to capture the previously described translational pausing and stacking events (Wolin and Walter 1988).

We next analyzed monosome and disome footprints per gene. Many transcripts that were detectable at the monosome footprint level showed robust disome coverage, allowing for quantification of both footprint species across a large portion of the expressed genome ($N = 8626$ genes). We first computed the ratio of CDS-mapping footprint to RNA-seq reads per gene. For monosome footprints, this ratio is frequently referred to as "ribosome density," reflecting a transcript's relative translation efficiency. Similarly, 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 densities were positively correlated with TEs (Fig. 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 (Fig. 2B; Supplemental Fig. S3C). Second, the steepness of the fit in the double-log plot in Figure 2E was ~1.7, i.e., much greater than 1, indicating a power relation between disome and monosome densities. Conceivably, increased ribosomal flux on mRNAs was associated with an even higher relative increase in ribosomal collisions. This relationship between disome and monosome footprint levels was not only observable across different transcripts but also for a given transcript at different TEs. Thus, we analyzed mRNAs encoding ribosomal proteins (RPs), which show prominent, feeding-dependent daily rhythms in TE (Sinturel et al. 2017). Using two time points from our data sets 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 approximately twofold increase in monosome TE between the two time points (Supplemental Fig. S3D). Taken together, these findings suggested that—at least in part—disome footprints were related to high ribosomal traffic ("traffic jams")

(Diament et al. 2018). In that case, one might hypothesize that the actual sites of ribosomal crowding 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. 2020)—would be expected to bias the collision sites as well. In contrast, however, the observation that signal peptides represented 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 Fig. S3E), suggested that beyond the alleged "stochastic" sites, more specific, "deterministic" signals and stall sites existed, too.

To gain a sense of whether these different scenarios (i.e., stochastic/deterministic disome sites) truly existed, we first visually inspected various individual transcript examples (Fig. 2F–O; see also Supplemental Fig. S5 for transcript plots stratified by individual ZT libraries). To begin with, we noted that individual SP transcripts exhibited the expected disome patterns. As shown for the case of *Adgrg3*—whose annotated signal peptide spans amino acids (aa) 1–18—disome footprint coverage was elevated upstream of codon ~75 and was lower and dispersed over the remainder of the CDS (Fig. 2F). Similarly, *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 (Fig. 2G), indicating a direct relationship between the positions of disome buildup and of the signal sequence. We next examined individual non-SP transcripts for the presence of the alleged stochastic and deterministic sites. For example, the transcripts encoding two 26S proteasome subunits, *Psmd4* (Fig. 2H) and *Psmd5* (Fig. 2I), showed distinct patterns of disome distribution that were consistent with our expectations for stochastic and deterministic sites, respectively. *Psmd4* thus showed disome coverage at numerous positions along the CDS, yet a specific, dominant site was apparent for *Psmd5*. Many other transcripts showed such patterns with distinct dominant sites as well, e.g., *Aldh1a1* (Fig. 2K), *Pklr* (Fig. 2M), and *Eif5a* (Fig. 2O). Dispersed disome patterns similar to *Psmd4*, as well as mixed cases combining broad coverage with specific dominant sites were frequent, too, e.g., *Aldoa* (Fig. 2J), *Acox3* (Fig. 2L), and *Eif2a* (Fig. 2N). 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 anticorrelation—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; 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 a monosome footprint depletion.

An obvious consequence of elongating ribosomes getting diverted into disomes is that conventional (monosome) Ribo-seq data sets will likely underestimate the number of translating ribosomes per transcript, in particular for mRNAs with high TE. We wished to quantify this effect. Because our existing monosome and disome footprint data sets originated from independent libraries (Fig. 1C), they could not be normalized relative to each other. We therefore sequenced new libraries from liver samples to which, early in the protocol, we had added defined quantities of synthetic 30-mer and 60-mer RNA spike-ins (Supplemental Fig. S6A,B), allowing for a quantitative realignment of monosome and disome
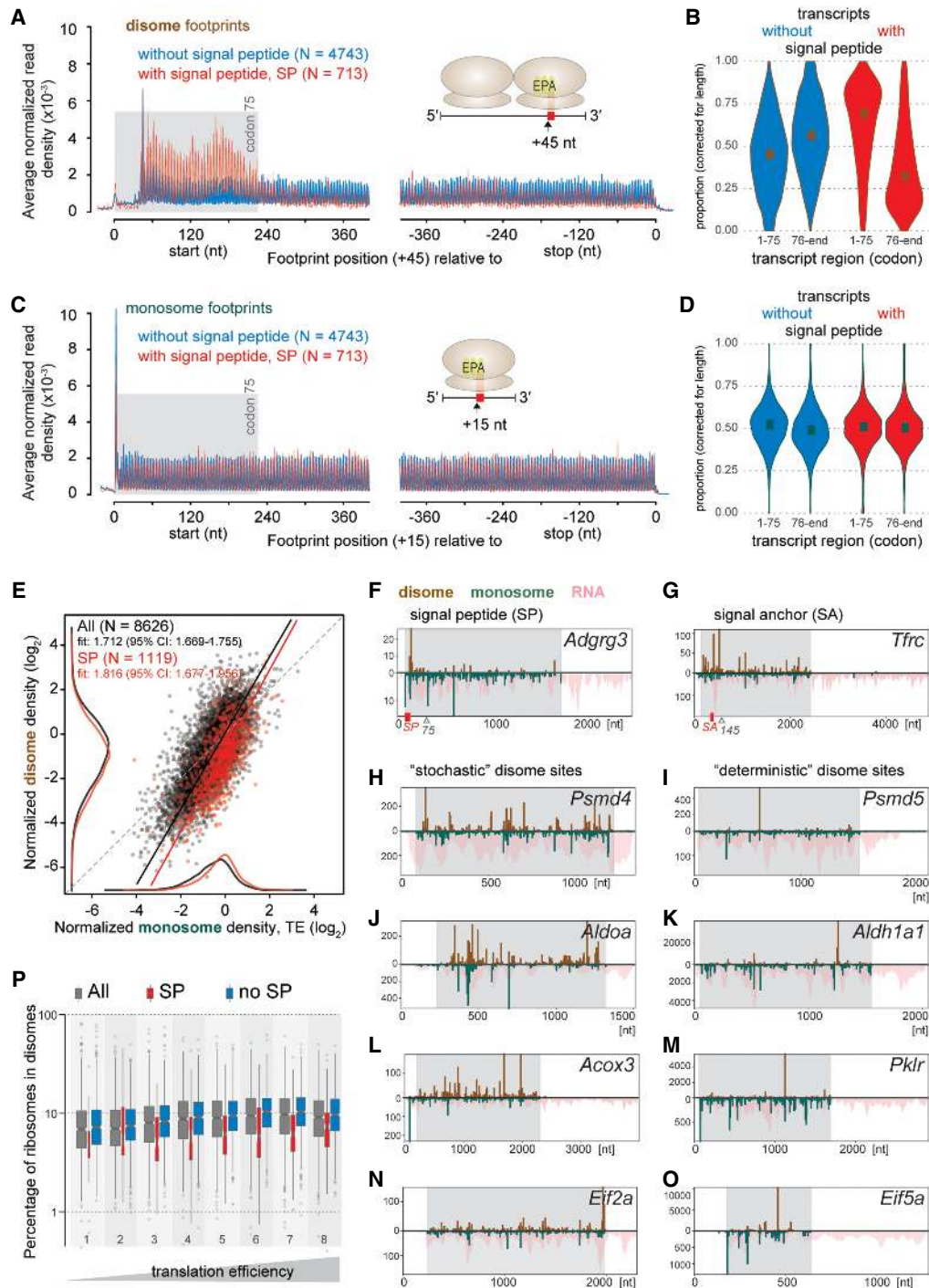
Figure 2. Disomes are associated locally with signal peptides and globally with high volumes of translation. (*A,C*) Density distribution of disome foot-prints identify signal peptide (SP)-related pausing events. Metatranscript analysis quantified the mean normalized footprint densities of disomes (*A*) and monosomes (*C*) within 400 nt from the start or −400 nt from the stop codons of transcripts encoding SPs (red, *N* = 713) or not (blue, *N* = 4743). (*B,D*) Violin-plots show the probability densities of length-normalized proportions of footprints within the first 75 codons and the rest of CDS from transcripts with (red, *N* = 713) or without (blue, *N* = 4743) SP for disomes (*B*) and monosomes (*D*). (*E*) Scatterplot of the relationship between per-gene normalized densities of disome and monosome footprints. All genes (*N* = 8626) were marked red or black depending on if they coded a SP (*N* = 1119) or not, respectively. Kernel density estimates are plotted on the margins (monosome on *x*-, disome on *y*-axis) for data sets of all genes (black) and SP coding genes (red) (without an axis of ordinates). Deming regression (errors-in-variables model) 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–O*) Distribution of normalized counts of monosome and disome footprints along transcripts of representative genes confirms stochastic versus deterministic sites. The upward *y*-axis of the bar-plots shows the normalized read counts for disomes (brick red), while the downward *y*-axis was used for monosomes (teal) and total RNA (pink, pile-up). Transcript coordinates (nt) are shown on the *x*-axis; CDS regions are shaded in gray. If present, SP or signal anchor (SA) regions are indicated as red boxes along the *x*-axis. Plots show: *Adgrg3*, *Tfrc*, *Psmd4*, *Psmd5*, *Aldoa*, *Aldh1a1*, *Acox3*, *Pklr*, *Eif2a*, and *Eif5a* in *F–O*, respectively. (*P*) Box-plots illustrate the estimated proportion of ribosomes retained in disomes as a percentage of all translating ribosomes for different groups of genes. Box-and-whiskers were drawn for all genes detectable in the spike-in experiment (gray, *N* = 7375), subsets that code for SP (red, *N* = 892) or not (blue, *N* = 6483) and stratified into eight groups based on the octiles of the TE calculated from all genes, with 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 graph. Width of each box is proportional to the number of data points it represents.

footprint data. This approach revealed that, for transcripts with high TE, typically ∼10% of translating ribosomes were in disomes (Fig. 2P). This proportion decreased with decreasing 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 indicated that disome footprints indeed resulted from ribosomal collisions between a downstream, slow decoding event and an upstream ribosome stacking onto the (temporarily) stalled ribosome.

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. We adapted a method developed for the analysis of monosome-based footprint data, termed Ribo-seq Unit Step Transformation (RUST), which calculates observed-to-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 was nonrandom as well, which was, however, not specific to footprints and also found in RNA-seq data. It was thus concluded that (1) A- and P-site codon identity and (2) the sequence-specificity of the enzymes used for library construction were the main factors dictating footprint frequency at a given mRNA location (O'Connor et al. 2016).

Before applying the RUST pipeline to the disome footprints, we first needed to investigate the origins of their bimodal length distribution (Fig. 1E) and determine which footprint nucleotides likely corresponded to the ribosomal E-, P- and A-sites. To this end, codon enrichment analyses conducted individually for the different disome footprint sizes (Supplemental Fig. S7A) resulted in profiles resembling the reported RUST profiles for monosome data (O'Connor et al. 2016). Increased information content at the footprint boundaries reflected the aforementioned library construction biases. 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. We did not notice any selectivity in the region occupied by the upstream ribosome or at the boundary between the ribosomes. These findings fit the model that the leading ribosome defined the pause site (with preference for specific codons) and an upstream ribosome colliding sequence independently. Furthermore, the comparison 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 to ribosome collisions in which the upstream ribosome stacked onto the stalled ribosome in two distinct states that differed by one codon (Supplemental Fig. S7B). Conceivably, the 1-nt variation (59 nt vs. 60 nt; 62 nt vs. 63 nt) corresponded to different trimming at the footprint 3′ end. Using this model, we aligned the main populations from the range of footprint lengths (i.e., 58–60 nt and 62–63 nt; together approximately two-thirds of all disome footprints) according to the predicted A-site of the paused, leading ribosome. Using these corrected A-site predictions on the metatranscriptome

analyses led to an improvement of the 3-nt periodic signal of the disome footprints (Supplemental Fig. S3B; cf. Fig. 1F). We used these A-site-corrected footprints for the RUST pipeline.

Enrichment analyses revealed marked amino acid selectivity in the P- and A-sites of the disome's leading ribosome (Fig. 3A, left panel). The magnitude of amino acid preference was greater than that seen for monosome footprints (Fig. 3A, middle panel), and RNA-seq data only showed the expected biases from library enzymology (Fig. 3A, right panel). Specific amino acids stood out as preferred ribosome stall sites, irrespective of codon usage. Strong associations were, in particular, the prominent P- and A-site over-representation of aspartic acid (Fig. 3B), the enrichment of isoleucine in the A-site and its depletion from the P-site (Fig. 3C), and the enrichment of glycine in the P-site (Fig. 3D) of paused ribosomes. We transformed the full amino acid analysis (Supplemental Fig. S8) into a position weight matrix representing the ensemble of positive and negative amino acid associations with disome sites (Fig. 3E). The 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 was a relevant factor for ribosomal pausing. Moreover, we noticed that for certain amino acids, association with disome sites was dependent on codon usage. For example, P- site asparagine was strongly associated with pause sites only when encoded by AAT, but not by AAC (Fig. 3F); lysine was depleted at P-sites irrespective of codon usage, but at the A-site either depleted (AAA) or enriched (AAG) (Fig. 3G). We also compared the reproducibility of enrichment patterns. Individual analyses of the 59–60 nt and 62–63 nt footprints resulted in near-identical weight matrices (Supplemental Fig. S9A,B), indicating that footprint size did not discriminate pausing events of different specificity. Moreover, across the six independent biological samples (Supplemental Figs. S10, S11) and for the independent libraries from the spike-in experiment (Supplemental Fig. S9C), the decoding center of the paused ribosome showed similar enrichment. Finally, we realized that the observed amino acid signatures showed resemblance with ribosomal dwell times that were recently estimated through modeling of conventional mouse liver Ribo-seq data (Gobet et al. 2020). Indeed, our monosome footprint data, too, showed similar patterns of amino acid enrichment and depletion, though much reduced in magnitude (Supplemental Figs. S8, S11, S12A).

We next investigated the association of pause sites with specific amino acid combinations. Strong selectivity with regard to the 400 possible dipeptide motifs was apparent in the P- and A-sites of the leading ribosome (Fig. 3H, left panel). This effect was much weaker and absent for monosome and RNA data, respectively (Fig. 3H, middle and right panels). In the disome data, the enrichment was highest, and independent of codon usage, for dipeptides consisting of the most enriched single amino acids, i.e., Gly-Ile, Asp-Ile, and Gly-Asp (Fig. 3I–K). In contrast, the pausing of ribosomes at several other dipeptides was strongly dependent on codon usage. In particular, the presence of lysine or glycine in the A-site of the leading ribosome was associated with codon selectivity (Fig. 3L). For instance, the Asp-Lys dipeptide was highly associated with disomes when encoded by GATAAG (Fig. 3M, blue trace); with transcriptome-wide 910 cases of disome peaks observed on the 2030 existing GATAAG positions (i.e., 44.8%), it was the eighth most disome-prone dicodon out of the total 3721 (i.e., 61×61) possible dicodon combinations (Supplemental Table S2; Fig. 3O). In contrast, when encoded by GACAAA (Fig. 3M, black trace), disomes were observable on no more than 7.8% of sites (272 out of 3529), ranking this dicodon
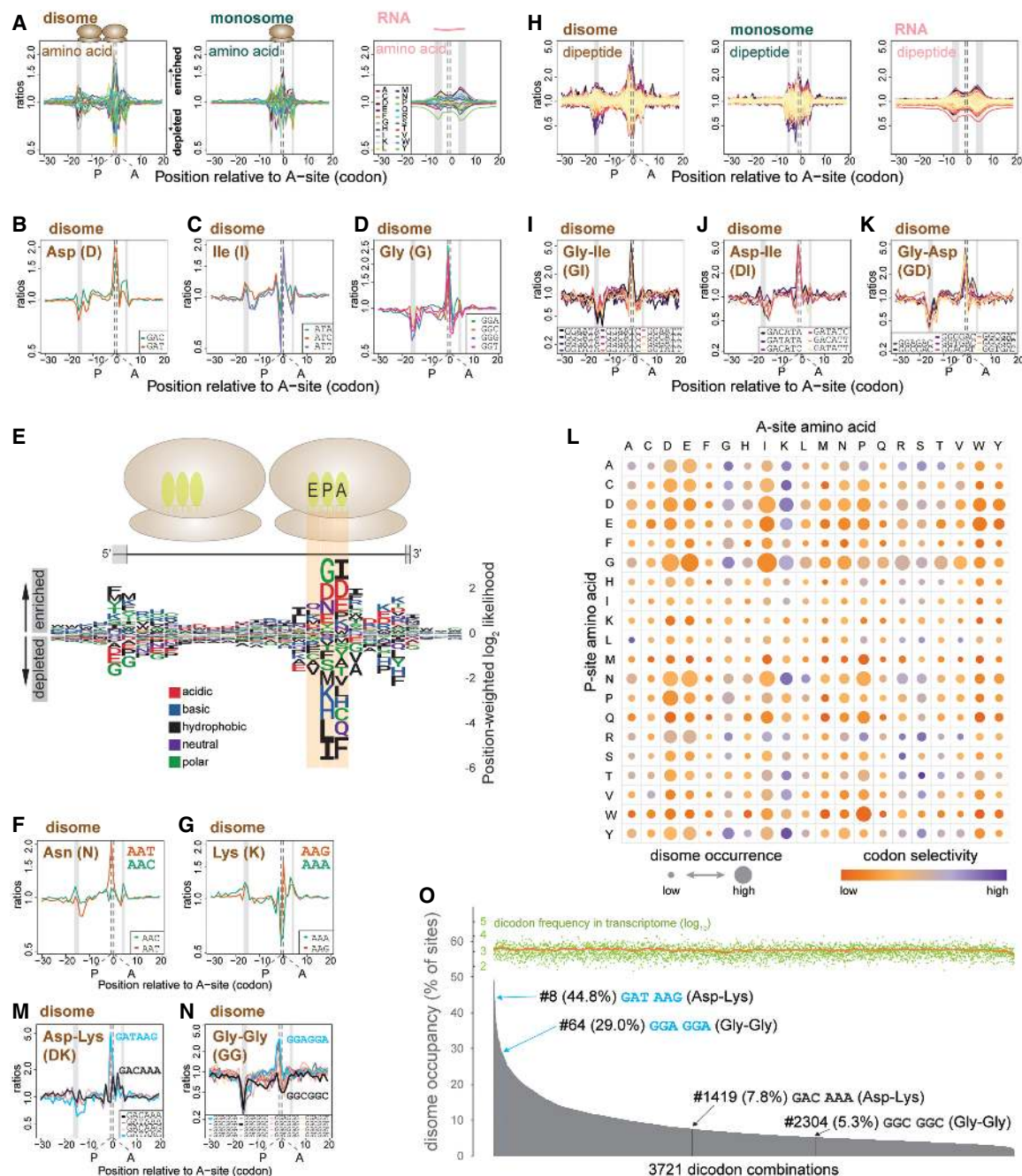
Figure 3. Disome sites show specific amino acid and codon enrichment. (*A*) Position-specific enrichment analysis reveals selectivity for amino acids in the decoding center of paused ribosomes. Normalized ratios of observed-to-expected occurrences (*y*-axis, log-scaled) of nucleotide triplets, grouped by the amino acid they code (*inset* in *right* plot), 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 midpoint of the reads. Ratios above and below 1 suggest enrichment and depletion, respectively. The vertical gray bars indicate the positions of the 5′ and 3′ ends of the read inserts for different library types. A- and P-sites are 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 *A*, yet triplets were not combined into amino acids but instead shown individually (*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 of specific amino acids within the decoding center of the leading ribosome of the disomes. Position-specific weighted log$_2$-likelihood scores were calculated from the observed-to-expected ratios (*A*). Enrichment and depletion carry positive and negative scores, respectively. Height of each single-letter amino acid character is determined by its absolute score. At each codon position, letters were sorted by the absolute scores of the corresponding amino acids, in descending order. Letters are colored by amino acid hydrophobicity and charge. The ribosome pair and their footprint are depicted graphically at the *top*, with gray zones at the extremities of the footprint denoting the spread of 5′ and 3′ ends of the read inserts. (*F,G*) Similar to *B*, for asparagine (Asn) (*F*) and lysine (Lys) (*G*). (*H*) Position-specific enrichment plots for dipeptides. Similar to *A*, but instead of triplets and single amino acids, 6-mers coding 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–K*) Similar to *B*, showing enrichment of individual 6-mers for dipeptides Gly-Ile (*I*), Asp-Ile (*J*), and Gly-Asp (*K*). (*L*) Enrichment and codon selectivity of all amino acid combinations at the predicted P- and A-sites of the leading ribosome. Identities of amino acids at the P- and A-sites are resolved vertically and horizontally, respectively. Disk area and color represent enrichment of disome sites and codon selectivity, respectively. Codon selectivity is calculated as the difference between the max. and min. enrichment ratios (log) of all 6-mers coding for a given dipeptide. (*M,N*) As in *I–K*, for Asp-Lys (*M*) and Gly-Gly (*N*). Disome-prone and disome-poor codon usages are marked in blue and black, respectively. (*O*) Relative disome occupancy by dicodon. Disome occupancy for the 3721 dicodon combinations was plotted in descending order. Occupancies were calculated for a given 6-mer (dicodon) as the raw percentage of sites with disome to all present sites (with + without disome) across the studied transcriptome. The frequency of sites is shown at the *top* of the graph colored in lime (moving average trend line in orange). Annotated are two pairs of 6-mers from panels *M* and *N*, coding for Asp-Lys or Gly-Gly, which show large differences in disome occupancies depending on codon usage (blue vs. black for high vs. low occupancy, respectively).

at position 1419. The Gly-Gly dipeptide represented a similar case (Fig. 3N); of the 16 dicodon combinations, 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).

In summary, the preference for codons, amino acids and dipeptides at the predicted P- and A-sites of the leading ribosome suggested that specific sequence signatures are an important contributor to the locations of collision events. Pausing that depends on codon usage opens the possibility to modulate the kinetics of translation elongation independently of amino acid coding potential; globally, disome-prone dicodons were indeed slightly less abundant than expected by chance (Supplemental Fig. S13).

Disome sites are related to structural features of the nascent polypeptide

The two factors revealed so far—high ribosomal flux (Fig. 2) and specific amino acids/codons (Fig. 3)—would likely not provide enough specificity to discriminate between the disome sites that were actually observable, as compared to mRNA positions that were devoid of disome footprints despite similar codon composition. We therefore expected that additional features would be critical in specifying ribosomal collision sites. Our above findings indicated that the signal peptide represented one such element promoting stalling and stacking. In this context, we noted that even for SP-related stalling, the actual sites on which disomes were observable were in accordance with the generic features identified above. Thus, disome density on SP-transcripts was dependent on TE (Fig. 2P), and the amino acid preference of disome sites at SP sequences (Supplemental Fig. S12B) closely resembled that identified transcriptome-wide (Fig. 3E).

To identify other protein features associated with ribosomal stalling, we first assessed the relationship between disome sites and the electrostatic charge of the nascent polypeptide. These analyses revealed, first, a strong association of negatively charged amino acids with the decoding center of the stalled ribosome (Fig. 4A, left panel), as expected from the Asp and Glu enrichment (Fig. 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 (Fig. 4A, left panel; red shaded area). This association was specific to disome footprints (it was only weakly detectable and absent, respectively, from monosome footprint and RNA data) (Fig. 4A, middle and right panels), and it continued far upstream of the footprint, ruling out that it was an effect of sequence bias in library generation. These findings indicated an interplay between the nascent polypeptide and the speed at which codons located substantially further downstream were translated (Fig. 4B). This idea is consistent with previous work showing that electrostatic interactions between a positively charged nascent peptide and the negatively charged lining of the exit tunnel promote local slow-down of elongating ribosomes (Charneski and Hurst 2013).

We next explored the influence of nascent polypeptide structure. Using genome-wide peptide secondary structure predictions with the three categories, structured (α-helix, 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 (Fig. 4C, left panel). Upstream and

downstream of the stalled ribosome, this pattern was inverted, with an increase in structured and a decrease in unstructured residues. The identical analyses on monosome and RNA data yielded associations that were weak (although qualitatively similar) (Fig. 4C, middle panel) and absent (Fig. 4C, right panel), respectively, suggesting high specificity. These findings were 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 (Fig. 4D). To investigate this hypothesis more directly, we retrieved the transcript regions encoding "structured-unstructured-structured" (s-u-s) polypeptide configurations transcriptome-wide (N = 9312). After rescaling 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 (Fig. 4E, left panel). By comparing with distributions obtained from randomizations of the disome peak positions within the same data set, we could conclude that the observed disome enrichment was significantly higher than expected by chance. As before, weak and no effects, respectively, were detectable in monosome footprint and RNA data (Fig. 4E, middle and right panels). Finally, the position-specific analysis (without rescaling) at the s-u boundary indicated that disome sites were particularly enriched at the second codon downstream of the s-u transition (Supplemental Fig. S14A, right panel). As an additional control for the specificity of these associations, we analyzed the inverse configuration, u-s-u (Supplemental Fig. S14D) and conducted all analyses on monosome footprint (Supplemental Fig. S14B,E) and RNA data (Supplemental Fig. S14C,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 residues mapped within known protein structures confirmed this finding, as shown for PSMA5, ALDH1A1, GAPDH, and EIF5A (Fig. 4F–I).

In summary, we concluded that there was a direct link between ribosomal pause sites and structural features of the nascent polypeptide. Translational pausing was more likely to occur while decoding negatively charged amino acids that were downstream of extended positively charged regions of the polypeptide, and within unstructured areas downstream of structured regions. These associations are suggestive of connections between elongation pausing and protein folding and assembly.

Disome sites are enriched within distinct transcript groups and are associated with previously documented translational pauses

Only a few translational pauses have been documented in the literature so far. Among them is the pausing associated with SRP recruitment (Wolin and Walter 1988) that is recapitulated in the disome data (Fig. 2A–G). We examined whether our analyses could provide insights into other known pausing events and whether specific groups of transcripts, processes, pathways, or cotranslational events were especially prone to pausing. We first selected the most prominent deterministic sites—i.e., pausing events that were not merely attributable to high ribosomal traffic—to enrich for potential functionally relevant pauses (top 5650 disome peaks from 1185 genes) (Supplemental Table S3). These strong disome sites showed high correspondence across all six independent biological samples (Supplemental Fig. S15), indicating high
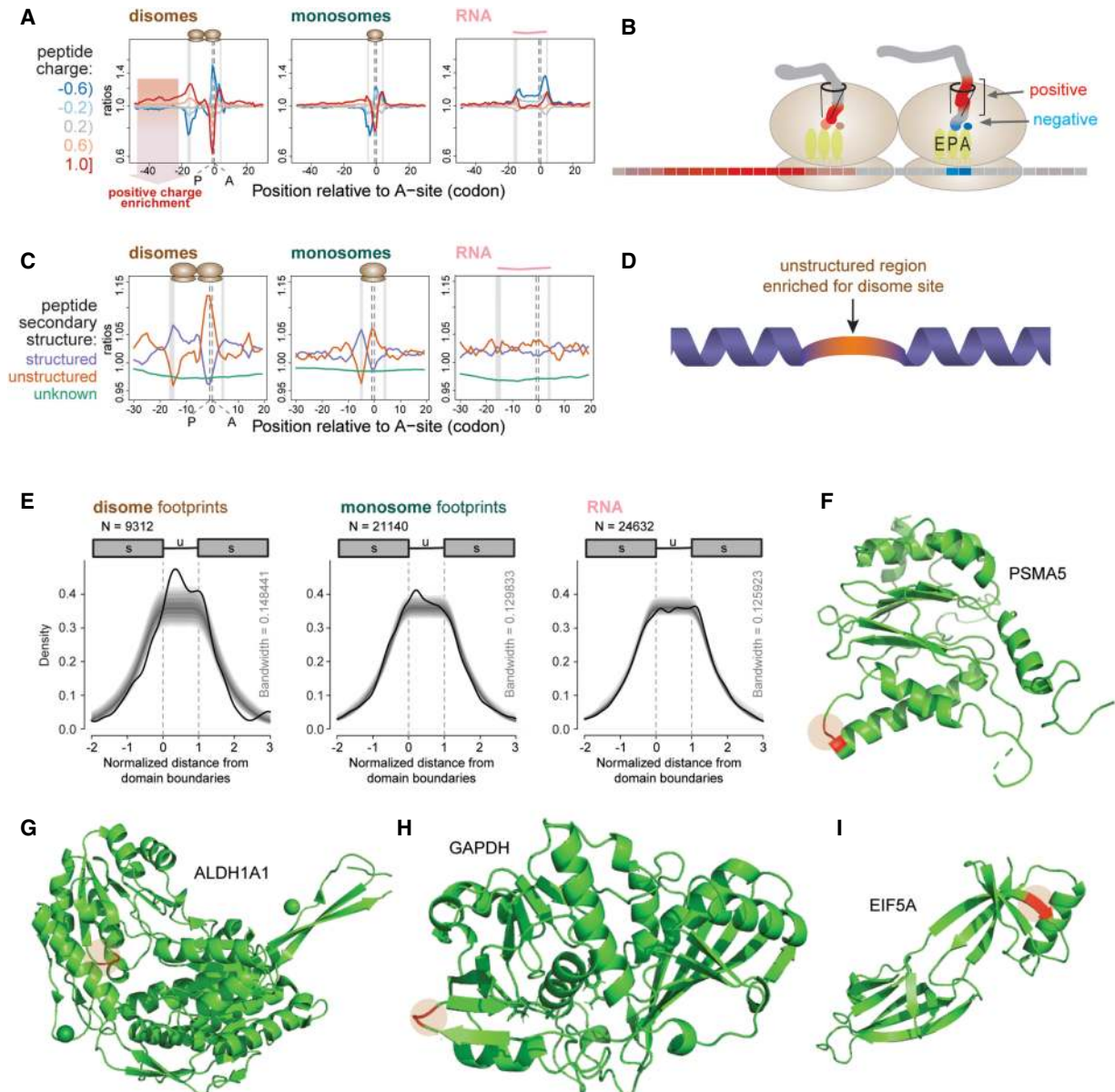
Figure 4. Disome site positions are related to nascent polypeptide charge and secondary structure. (*A*) Position-specific enrichment analysis reveals association with positive charge in the nascent polypeptide. Average charge of three consecutive amino acids was stratified into five charge groups (interval 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 (disomes, *left* panel), or of the individual ribosome (monosomes, *middle* panel). RNA is shown in the *right* panel. The red shaded area in the disome panel (*left*) marks the extended stretch of positive charge upstream of pause sites. See Figure 3A for general plotting features. (*B*) Schematic of the electrostatic interactions between the leading ribosome and the nascent peptide chain. Associations of negatively charged residues (blue) with the P- and A-sites and a stretch of positively charged residues (red) within the exit tunnel is depicted. (*C*) Association between disome sites and the nascent polypeptide structure. Based on the UniProt structural annotation, each position of translated peptides was labeled "structured" for α-helix or β-sheet, "unstructured", or "unknown"; β-turns were excluded. See Figure 3A for general plotting features. (*D*) Schematic depicting a preference for pausing during the translation of unstructured polypeptide stretches (orange) that are 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 aligned to its start, such that start and 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*), A-site of the monosomes (*center*), or center of total RNA reads (*right*). The density lines drop naturally towards the extremities, as the data matrices were normalized and aligned to the unstructured region and lower numbers of data points are expected to be observed at increasing distance from the boundaries. Confidence intervals for the kernel densities, which were calculated by randomly shuffling (*N* = 10,000) peaks within each transcript, are shown by gray shaded regions (and allow estimating statistical significance of the signal): darkest at the center, 50% (median) to outward, 25%, 12.5%, 5%, 2.5%, and 1%. (*F–I*) Three-dimensional structures of proteins with disome site amino acids highlighted. Human PSMA5 (PDB ID: 5VFT) (*F*); human ALDH1A1 (4WJ9) (*G*); human GAPDH (4WNC), corresponding residues at aa 65–66 (*H*); murine EIF5A (5DLQ) (*I*). The positions of the strongest disome sites are shown in red.

reproducibility. First, we searched whether structural data was available for the proteins with prominent disome sites. Out of the first ~50 genes in the list, around 20–25 structures (from mouse or mammalian orthologs) were available from published data. 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 Fig. S16; Fig. 4F–I). These findings were consistent with the idea that efficient pausing may be important for the structural integrity and folding of nascent polypeptides.

We next explored whether the pausing phenomenon affected specific pathways or functions. Among the genes with prominent disome peaks (top 200 genes from Supplemental Table S3), the analysis revealed a strong bias for categories "cofactor and coenzyme binding," "oxidation-reduction processes," and "mitochondria" (Fig. 5A; Supplemental Table S4). It is tempting to speculate that the integration and/or covalent attachment of cofactors (a common feature of oxidoreductases) onto polypeptides is coordinated cotranslationally and that the mechanism of biogenesis employs translational pausing. The enrichment of transcripts encoding mitochondrial proteins may reflect a specific feature of their translational kinetics, possibly related to cotranslational protein localization/import to mitochondria (Lesnik et al. 2015).

Next, we inspected notable specific cases of translational pausing. We spotted nine transcripts specifying proteins that contained the rare amino acid selenocysteine (Sec/U) among the translatome-wide most prominent disome peaks (Supplemental Table S3; Supplemental Fig. S17A). Six additional Sec-codon-containing transcripts were expressed in liver but showed varying degrees of disome occurrence (Supplemental Fig. S17B) and did not feature among the most prominent peaks translatome-wide. Sec is encoded by the UGA stop codon, whose reinterpretation involves the 3′ UTR-located Selenocysteine Incorporation Sequence element (SECIS) (Vindry et al. 2018). Selenocysteine decoding is slow,

and ribosomal collisions during prolonged dwelling of elongating ribosomes on Sec codons is a plausible scenario (Howard et al. 2013). In several cases, such as Selenok (Fig. 5B) and Sephs2 (Fig. 5C), the disome peak indeed coincided with the Sec codon. In other cases, however, there was no correspondence between disome peak and Sec codon (Supplemental Fig. S17A,B). We inspected whether specific RNA elements could be responsible for the differences in disome location but did not observe any such association (Supplemental Fig. S17A,B).

Further inspection identified Sec61b among the prominent disome transcripts. Sec61b encodes a tail-anchored protein and translational slowdown after its C-terminal transmembrane domain (TMD) is understood to provide time to recruit the machinery for membrane insertion before TMD release from the ribosomal exit tunnel (Mariappan et al. 2010). We observed several disome peaks on Sec61b (Fig. 5D), including on a Gly-Lys dipeptide (disome-prone codon usage GGCAAG) immediately adjacent to the TMD. Mariappan et al. had reported the same mechanism for Vamp2, which indeed showed a strong disome peak towards the 3′ end of the CDS that lay, however, within rather than after the TMD (Fig. 5E). Finally, we examined other documented translational pauses. Pausing on the Xbp1 transcript, which involves translation of a hydrophobic C-terminal region (CTR), facilitates cotranslational mRNA localization to the endoplasmic reticulum membrane (Yanagitani et al. 2011). Our data confirm multiple disome sites in this area (Fig. 5F). A strong site was specifically on Asn256, which is the last codon required for translational arrest (Yanagitani et al. 2011) and on which a pausing event was identified by Ingolia et al. (2011). Recently, two unusual cases of regulatory translational stalling were identified for Amd1 (Yordanova et al. 2018) and Azin1 (Ivanov et al. 2018), encoding components of the polyamine biosynthesis pathway. Low coverage precluded analysis of disomes on Amd1. For Azin1, a specialised uORF, termed the upstream conserved coding region (uCC), undergoes
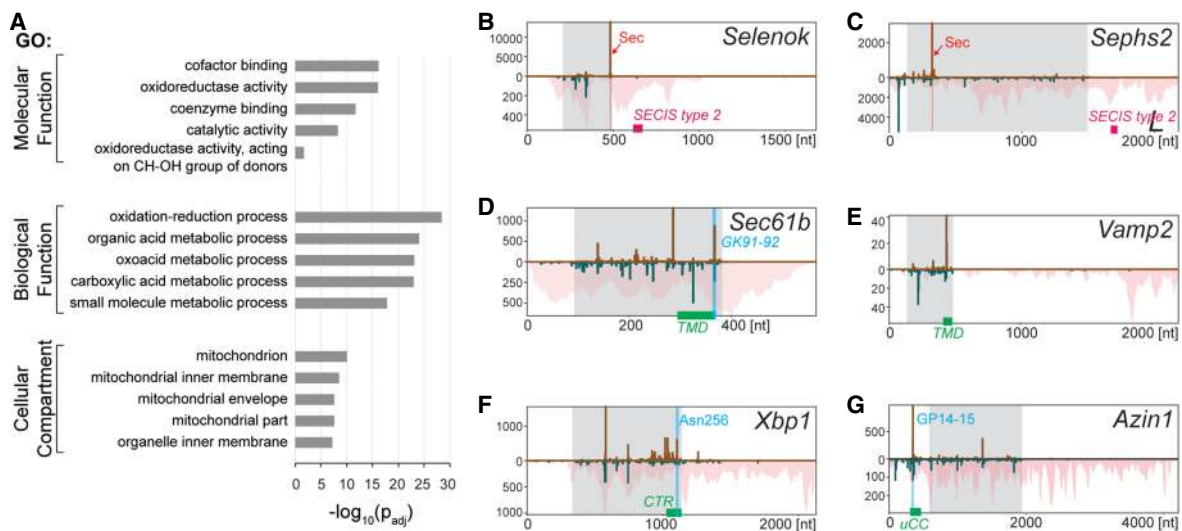

Figure 5. Disome sites are associated with specific pathways and with known pausing events. (A) Functional enrichment analysis of the top 200 genes from the prominent disome peak list. Five terms with the highest $-\log_{10}(p_{adj})$ values (horizontal bars) are shown from each Gene Ontology (GO) group: molecular function, cellular component, biological process. See Supplemental Table S4 for full analysis. (B–G) Distribution of normalized counts of monosome and disome footprints (per nt) and RNA (pileup) along selected transcripts, similar to Figure 2F–O. Selenok (B) and Sephs2 (C) show a strong disome peak on the selenocysteine codon (Sec, marked in pink). Position of the SECIS elements is indicated in pink. Sec61b (D) and Vamp2 (E) are tail-anchored proteins with a transmembrane domain (TMD, green). For Sec61b, a strong disome site is located on GK91-92 (marked in blue). Xbp1 (F) contains a C-terminal region (CTR, green) with several disome sites. The strong site on Asn256 is marked in blue; Azin1 (G) contains an upstream conserved coding region (uCC, green) that undergoes polyamine-dependent translational elongation. The main disome site is on a the uCC dipeptide GP14-15.

polyamine-dependent translational elongation, which leads to ribosome queuing and main CDS start site selection (Ivanov et al. 2018). We indeed observed strong disome signal precisely on the uCC (Fig. 5G). Ivanov et al. mapped a pausing event to the PPW tricodon (uCC amino acids 47–49), whereas our data revealed strongest disome accumulation on a site corresponding to ribosome pausing further upstream, on a Gly-Pro dipeptide (aa 14–15).

Evolutionary conservation at disome sites suggests an active, functional role for pausing

The above examples showed that disome sites were associated with several functionally characterized cases of ribosomal pausing. Globally, however, our analyses did not allow distinguishing whether the observed ribosomal 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, protein biosynthesis and folding could slow down translation and thus, as a downstream effect, lead to ribosome pausing and collisions, without being of functional relevance for the preceding event itself. We therefore sought a way to evaluate the two scenarios. We reasoned that, in the case of an active, functional role, the codons and dipeptides on which pausing occurred would show higher evolutionary conservation than expected.

RUST analysis using phyloP conservation scores revealed enrichment for highly conserved codons at the P- and A-sites of the stalled ribosome (Fig. 6A). Moreover, we observed that highly conserved transcripts generally showed high disome levels, while poorly conserved transcripts were rather disome-poor, especially for mRNAs with high TE (Fig. 6B). Although these analyses indicated connections between translational pausing and evolutionary conservation, they did not allow determining how direct this link was. Most of all, the further interpretation was rendered difficult due to the intrinsic selectivity of disome sites for specific amino acids, codons, and dipeptides (Fig. 3), which would represent a strong confounding factor for simple evolutionary analyses as the one shown in Figure 6A.

To uncouple the various effects on conservation, we used logistic regression analysis to determine what contribution specifically the disome sites made to overall conservation, against the background of general dipeptide and transcript conservation scores. This analysis showed that different dipeptides had very distinct conservation scores already on their own

(Fig. 6C). Therefore, the association of disome sites with high phyloP scores (Fig. 6A) was mostly attributable to the specific dipeptide bias at these positions. Furthermore, the regression analysis revealed that the presence of disomes increased the odds of having a highly conserved phyloP score (cutoff: phyloP > 5 for 60-way vertebrate and > 1.3 for euarchontoglires data sets, respectively) by approximately 10% (60-way vertebrate: $b = 0.1258$, P-value $< 1 \times 10^{-16}$, OR = 1.134 95% CI = 1.125, 1.143; euarchontoglires: $b = 0.1012$, P-value $< 1 \times 10^{-16}$, OR = 1.107, 95% CI = 1.098, 1.115). This outcome established that translational pauses identified by disome sites were significantly more conserved than expected by chance. While the thousands of pause sites that can be detected translatome-wide will span all categories (from deleterious to beneficial), the globally detectable signal of positive selection on stall site codons strongly argues for an active, functional role of ribosomal pausing.
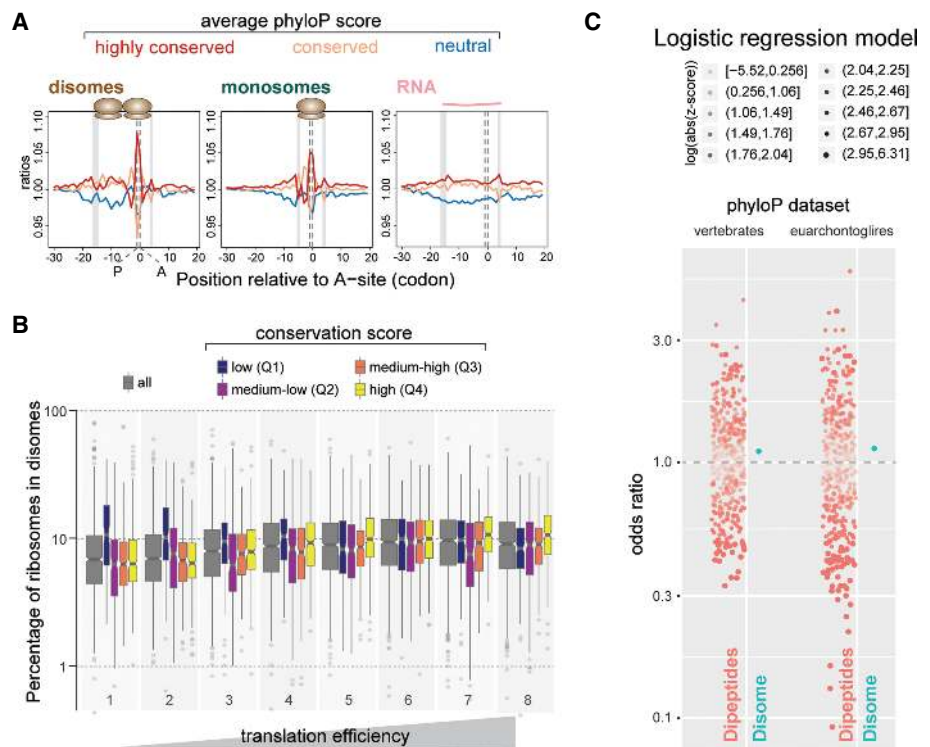

Figure 6. Evolutionary conservation at disome sites. (*A*) Association of highly conserved codons with the P- and A-sites of disome sites revealed by position-specific enrichment analysis. Along coding regions, phyloP conservation scores were grouped into categories: neutral - blue, [−3, 3), conserved - 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 (disomes, *left*), or of the individual ribosome (monosomes, *middle*). See Figure 3A for other elements. (*B*) Box-and-whiskers illustrating the estimated percentages of ribosomes that were in disomes for groups of transcripts with different overall evolutionary conservation. Groups included all detectable genes (all, gray, N = 7375), which were stratified into four groups (N = 2270 or 2271 for each; color code at the *top*) based on the quartiles of average phyloP scores with the following right-closed boundaries: −0.585, 2.327, 3.356, 4.239, 6.437. *x*-axis and other features are as in Figure 2P. (*C*) Odds ratio estimates of dipeptides and disome sites for increased phyloP scores. Odds ratios (OR) for having a high phyloP score at P-A dicodons were estimated for dipeptides encoded by the dicodon (orange dots, 399 levels relative to dipeptide VH, which had moderate phyloP scores in both models) and presence of a disome peak (green dots, A-position disome density > mean transcript density) using a logistic regression model. Confidence levels of estimates were represented by transparency levels that corresponded to deciles of the logarithm of absolute values of their *z*-scores (legend). Two separate regression models were fitted using phyloP scores from the 60-way vertebrate data set (*left*) and the Euarchontoglire subset (*right*). For disomes, OR is larger than 1 (dashed line) indicating that it is more likely to observe a high phyloP score when disome peaks are present than when they are absent.

Disome site codon usage affects protein output from a reporter gene

The globally detectable signature of evolutionary conservation is consistent with the idea that translational pausing events play active roles in polypeptide biosynthesis. Future experiments—ideally by creating pausing loss- and gain-of-function mutants through genetic knock-in of alternative codon usages at endogenous loci—will allow validating the biological functions of individual pausing events. In the framework of this study, we wished to gain, first, preliminary insights into how pause site modification affected protein output from a reporter gene. In order to select a suitable candidate, we speculated that, in cases where polypeptides assembled into large multiprotein complexes, poorly coordinated translation kinetics might affect protein abundance due to altered efficiency of incorporation. Excess unincorporated protein may be subject to degradation, mislocalization, or aggregation, potentially impacting steady-state protein levels. In the list of strong disome peaks (Supplemental Table S3), we noted several transcripts encoding ribosomal proteins. We selected Rps5 (Supplemental Fig. S18A,B) and cloned its cDNA in-frame with firefly luciferase in a lentiviral vector that allowed internal normalization to Renilla luciferase (Supplemental Fig. S18C). Change of codon usage at the Asp-Ile disome site from its natural GATATT (Rps5-wt; this codon usage is highly disome-prone: transcriptome-wide 46.6% of sites show disomes) to the disome-poorer GACATC (Rps5-mut2; 26.4% of sites have disomes) led to a significant change in steady-state protein output, although both constructs encoded for precisely the same protein at the amino acid level. Other variants of Rps5 did not show an effect, including such with disome-poorer codon usages. These observations indicated a complex relationship between pausing potential and protein abundance. The identification of functionally important disomes sites, and testing for such function, will be one of the future challenges.

Discussion

It has long been known that elongating ribosomes can slow down when they encounter obstacles. Presumably, most such pauses are only transitory and resolved in a productive manner (Schuller and Green 2018), and for certain cases, there is evidence that they are even an integral part of the mechanism of nascent polypeptide synthesis, as exemplified by the pausing seen on signal peptide-encoding transcripts that require targeting to the secretory pathway (Wolin and Walter 1988). Finally, pauses can be unresolvable, thus triggering a dedicated ribosome rescue program (Joazeiro 2019). While a number of previous studies have used monosome footprint intensities to infer pausing, there is clear benefit in tracking stalled ribosomes from 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 analyzed in this study indicate that they indeed represent a steady-state snapshot of the translational pausing and collision status in mouse liver in vivo. We view the sheer quantity of ribosomes that are trapped in the disome state (calculated from spike-based quantifications of disome vs. monosome signals) as quite remarkable. For a typical, highly translated mRNA, we estimate that ~10% of elongating ribosomes are affected by this phenomenon (Fig. 2P). Because certain stacking events will involve more than two collided ribosomes (Fig. 1A,B), we are likely even underestimating the overall ribosomal queuing phenomenon. Another potential source of underestimation could be

a loss of disomes through cleavage into two individual monosomes during the purification protocol. However, several observations argue against systematic loss by nuclease activity in the experiment. First, disomes show resilience to different conditions of nuclease treatment (Supplemental Fig. S1), and, second, there is no noticeable sequence bias within the footprint region that is located between the individual ribosomes (Supplemental Fig. S7A; Fig. 3E). Finally, when disomes are cleaved, we would expect the corresponding, 30-nt-spaced monosome footprints to appear. Previous studies have observed such 30-nt monosome footprint phasing, presumably reflecting queued ribosomes, upstream of stop codons (Andreev et al. 2015). In our data, monosome footprint phasing affects only a low number of very highly populated disome sites and is therefore overall rather limited in magnitude (Supplemental Fig. S19). Collectively, we view the 10% disome rate as a realistic estimate that is, moreover, of similar magnitude as the ~20% ribosome queuing rate recently calculated in budding yeast (Diament et al. 2018). While we consider many collisions to be a consequence of high ribosomal flux (i.e., "stochastic," yet more likely on certain codons/amino acids) rather than evidence of biological function, the loss of elongating ribosomes into queues poses challenges to the interpretation of conventional Ribo-seq data. Monosome footprint-based analyses 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 mRNA 3' ends engender small monosome footprints of ~21 nt and, when an incoming ribosome stacks onto the stalled one, of ~48 nt (Guydosh and Green 2014). Short footprints also occur in human cells (Wu et al. 2019), and data from HeLa cells suggests that both transient and hard stalls trigger an endonucleolytic cleavage that generates short footprints (Ibrahim et al. 2018). Conceivably, short footprints may reflect the more harmful pauses that provoke specific clearance pathways. The abundant ~60-nt footprints we describe here are distinct not only in size but likely also in the translational state that they represent. They match the length reported for SRP-related pausing in vitro (Wolin and Walter 1988). Moreover, footprints of this size were also noted in the above yeast study (Guydosh and Green 2014) and, although the authors did not follow up on them in greater detail, it is intriguing that a similar bimodal size distribution and depletion from the first codons post-initiation was reported, as in our liver data (Fig. 1E,F). We do not yet understand the significance of the two size populations: could they be associated with specific functional properties or collision states? The pattern of amino acid enrichment at the stalled ribosome is identical for the 59–60 nt and the 62–63 nt footprints (Supplemental Fig. S9A,B). Moreover, the translatome-wide most prominent disome sites always showed a mixture of both footprint sizes, yet the relative ratio of the two size classes differed widely across sites (Supplemental Table S3). Finally, we observed an unexpected distribution of the two size classes at the very 5' end of the CDS (Supplemental Fig. S20). Despite various open questions, we can conclude from the association with signal peptides, the high steady-state abundance, and the absence of signs of mRNA cleavage at the stall site that our disome profiling 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 largely governed by the P- and A-site amino acids of the downstream ribosome (Fig. 3E). There is little specificity at the E-site, which is notable because previous monosome

footprint-based pause site predictions identified a strong E-site bias for proline (e.g., Ingolia et al. 2011; Pop et al. 2014; Zhang et al. 2017). Due to their particular chemistry, prolines (especially in a poly-Pro context) are well-known for their difficult peptide bond formation and slow decoding, leading to stalls that can be resolved through the activity of HYP2 (also known as eIF5A) (Gutierrez et al. 2013). Yet, apart from a minor signal in the A-site codon (Fig. 3E; Supplemental Fig. S8), we do not see prolines associated with liver disome sites at all.

Recent disome data from mESCs (Tuck et al. 2020) do, however, show the expected poly-proline motif (Supplemental Fig. S9D). An explanation for these differences in disome patterns between cell types may be the high activity of EIF5A in liver. We have noted that, based on monosome footprint RPKMs, EIF5A is indeed synthesized at very high levels that even exceed, for example, those of the essential elongation factor EEF2. Moreover, it is curious that *Eif5a* is itself among the 200 genes with the strongest disome peaks, occurring on a conserved Gly-Ile position (Figs. 2P, 4I; Supplemental Table S3). It would be fascinating if translational pausing on *Eif5A* mRNA turned out to be part of a mechanism designed to autoregulate its own biosynthesis. Beyond the lack of proline signal in the liver data, we noted that termination codons were also absent from the disome data (see Fig. 1F). Of note, we deliberately did not pretreat our tissue samples with cycloheximide in order to avoid artifacts that this elongation inhibitor can cause. However, polysomal extract preparation and RNase I digestion occurred with cycloheximide to stabilize elongating ribosomes. We cannot exclude that terminating ribosomes may have been selectively lost at this stage, reducing disome signals at stop codons.

The specific amino acid and dipeptide motifs that we find enriched at the paused ribosomes show some resemblance as well as distinct differences to previous reports. For example, Asp and Glu have been associated with presumed pauses (Ingolia 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 association of pause sites with isoleucine is an unexpected outcome of our study, as, 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 some non-isoleucine dicodons as well) transcriptome-wide 35%–50% of such sites carry a strong disome footprint (Supplemental Table S2). These findings allow a bold speculation, which is that, on top of the simple three-letter codon table, a six-letter code punctuates translation to organize the biosynthesis of nascent polypeptides into segments separated by intermittent pause sites. Of note, the phenomenon that translation speed can be governed at the level of the dicodon is actually well known from work on bacterial translation (Irwin et al. 1995). Our analysis of mammalian translation identified thousands of such intermittent pauses (Supplemental Table S3), which, as an ensemble, likely reflect an array of different protein biosynthetic phenomena whose common denominator is local elongation slowdown. Even on the global set, an association of disome sites with structural features of the nascent polypeptide is evident (Fig. 4), as is the signature of evolutionary conservation (Fig. 6). Conceivably, this indicates that a major role for pausing could lie in the coordination of translation with the folding, assembly, or structural modification of nascent polypeptides. There is compelling evidence from yeast that many multiprotein complexes assemble cotranslationally (Shiber et al. 2018) and that the association of individual subunits involves translational pausing on nascent poly-

peptides (Panasenko et al. 2019). The showcase examples in the latter study are two proteins of the yeast proteasome regulatory particle, Rpt1 and Rpt2, whose elongation pausing leads to the association of the translating ribosomes into heavy particles ("assemblysomes") where the nascent peptides assemble into the multiprotein complex. In mouse liver, several proteasomal protein mRNAs carry high disome peaks as well (*Psmd5* in Fig. 2I; >10 other proteasome subunits in Supplemental Table S3), possibly indicating a conserved assembly pathway.

In conclusion, we view the disome profiling methodology as an important complementary technique to the already available ribosome profiling repertoire. Not unlike conventional Ribo-seq, it delivers a "snapshot" of the translation status, yet the cellular disome state provides specific, new information on translation kinetics. We deem it likely that the kinetics will show regulation in different organisms, tissues, cell types, and under different physiological conditions, which will manifest in distinct disome profiles. It will be exciting to collect and compare such data across experimental models, and to evaluate to what extent translational pausing represents an obligatory, potentially regulated event that contributes to physiological gene expression output. Through such new data sets, and already through the extensive data we have collected and analyzed in the framework of this study and in recent work in mESCs (Tuck et al. 2020), important new scientific questions are likely to become experimentally accessible.

Methods

Experimental models

Mouse liver extracts (from 12-wk C57BL/6 males) were the same as in Janich et al. (2015), and only those for spike-in experiments were prepared independently for this study (experiments approved by the Cantonal Veterinary Office, authorization VD2376). All details on extract preparation can be found in the Supplemental Material. Cell lines (NIH3T3, HEK293FT) were the same as described in Janich et al. (2015) and cultured under standard conditions (Supplemental Material).

Northern blots

The general protocol has been described in Gatfield et al. (2009). Briefly, RNAs purified from nuclease-treated tissue extracts were separated by polyacrylamide gel electrophoresis, electroblotted on membrane, immobilized, and hybridized with radioactively labeled oligonucleotides antisense to the *Alb* and *Mup7* transcripts. See Supplemental Material for details and probe sequences. Please note that the lower part of the northern blot panels shown in Supplemental Figure S1B was the same as in our previous publication (Janich et al. 2016).

Footprint and library generation

The original mouse liver data sets for monosome footprints and RNA-seq used in this current study were the same as reported in Janich et al. (2015), of which we used the three time points, ZT0, 2, 12 (two biological replicates per time point; each assembled from a pool of liver lysates from two mice). Disome footprints from the same samples had already been cut simultaneously, and from the same gels, as the monosome footprints in Janich et al. (2015), and the data sets were produced for the current study. Library generation occurred with a modified Ribo-seq protocol (in principle, according to Illumina's protocol for TruSeq Ribo Profile, using Ribo-Zero Gold rRNA Removal kit) with details (including for the spike-in experiment) in Supplemental Material. All libraries

were sequenced in-house on Illumina HiSeq 2500. mESC data were from Tuck et al. (2020).

Basic analysis of sequencing reads

Preprocessing of sequencing reads, mapping and quantification of mRNA, and footprint abundances largely followed protocols as described in Janich et al. (2015); a detailed description is given in Supplemental Material.

Spike-in normalization and global quantification of ribosomes retained in disomes

Design and sequences of the spike-in oligos (30-mers, and 60-mers that were concatamers of the 30-mer sequences), mapping procedure, and the counting algorithm to avoid counting degradation products of the 60-mers as 30-mers are all described in Supplemental Material. In the sequencing data, spike reads were mapped and processed similarly to all other reads. Spike counts were first normalized for library size with an upper-quantile method, and spike-in normalization factors were calculated as 60-mer/30-mer ratios per sample to correct the experimental biases between the disome and monosome counts. 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.

Observed-to-expected ratios for proximal sequence features

The calculation of observed-to-expected ratios for sequence features proximal to footprint sites was performed following the principles of the Ribo-seq Unit Step Transformation method (O'Connor et al. 2016). The described method was extended by including additional features such as 6-mer, dipeptide, charge, secondary structure, and phyloP conservation in addition to codon and amino acids. A margin of 30 nt was excluded from each end of the CDS. The analysis window (typically 50 codons wide) was moved along the CDS regions at single-nt or 3-nt steps. Enrichment was calculated as the observed-to-present ratio normalized to the expected ratio. All analyses were performed with in-house Python (creation of data matrices) and R software (R Core Team 2016) (visualization, statistical analysis). Unless stated otherwise, samples were treated as replicates and, in most analyses, were combined. More details are found in Supplemental Material.

Estimation of A-site positions

The A-site positions for monosome footprints were calculated as in Janich et al. (2015). For disome footprints, 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 footprint 5′ end. This approach was suitable for exploratory analyses (e.g., metatranscript analysis) and provided direct comparability to monosome results. In other analyses, we used an empirical method to estimate the A-site of the leading ribosome. In order to infer the optimal offsets for different footprint lengths, we first split the disome footprints by their size, from 55 to 64 nt. Within each size group, footprints were further split by frame (three bins, relative to the main CDS). For each group, position-specific (relative to their 5′ ends at nucleotide resolution) information content matrices were calculated using Kullback-Leibler divergence scores (O'Connor et al. 2016) of observed-to-expected ratios of codon analysis (Supplemental Material). For combinations of footprint size and reading frame, where the position of P-/A-sites could be identified as highest information positions (with approximately two peaks spaced by

3 nt) ~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. Offsets for 58, 59, 60, 62, and 63 nt disome footprints on different reading frames were, respectively: [45, 44, 43], [45, 44, 46], [45, 44, 46], [48, 47, 46], and [48, 47, 49]. Total RNA reads were offset with different methods to be consistent with the data set they were being compared to: by their center (default), +15 (when compared to monosomes, also selecting a similar size range of 26–35 nt), or disome offsetting (with size range 58–63 nt).

Other computational methods

The Supplemental Material contains detailed descriptions of other computational methods used in the study, including metatranscript analyses, analysis of footprint densities in relation to peptide secondary structures, analysis of evolutionary conservation at disome sites, mapping of disome amino acids onto protein structures, and functional enrichment analysis of genes with prominent disome peaks.

Data access

The raw sequencing data and processed quantification data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE134541. All scripts are available as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

References

Andreev DE, O'Connor PB, Zhdanov AV, Dmitriev RI, Shatsky IN, Papkovsky DB, Baranov PV. 2015. Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol* **16:** 90. doi:10.1186/s13059-015-0651-z

Charneski CA, Hurst LD. 2013. Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol* **11:** e1001508. doi:10.1371/journal.pbio.1001508

Dana A, Tuller T. 2012. Determinants of translation elongation speed and ribosomal profiling biases in mouse embryonic stem cells. *PLoS Comput Biol* **8:** e1002755. doi:10.1371/journal.pcbi.1002755

Dao Duc K, Song YS. 2018. The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. *PLoS Genet* **14:** e1007166. doi:10.1371/journal.pgen.1007166

Darnell AM, Subramaniam AR, O'Shea EK. 2018. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Mol Cell* **71:** 229–243.e11. doi:10.1016/j.molcel.2018.06.041

Diament A, Feldman A, Schochet E, Kupiec M, Arava Y, Tuller T. 2018. The extent of ribosome queuing in budding yeast. *PLoS Comput Biol* **14:** e1005951. doi:10.1371/journal.pcbi.1005951

Döring K, Ahmed N, Riemer T, Suresh HG, Vainshtein Y, Habich M, Riemer J, Mayer MP, O'Brien EP, Kramer G, et al. 2017. Profiling ssb-nascent

chain interactions reveals principles of Hsp70-assisted folding. *Cell* **170:** 298–311.e20. doi:10.1016/j.cell.2017.06.038

Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. 2016. Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell* **166:** 679–690. doi:10.1016/j.cell.2016.05.070

Gatfield D, Le Martelot G, Vejnar CE, Gerlach D, Schaad O, Fleury-Olela F, Ruskeepää AL, Oresic M, Esau CC, Zdobnov EM, et al. 2009. Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes Dev* **23:** 1313–1326. doi:10.1101/gad.1781009

Gobet C, Weger BD, Marquis J, Martin E, Neelagandan N, Gachon F, Naef F. 2020. Robust landscapes of ribosome dwell times and aminoacyl-tRNAs in response to nutrient stress in liver. *Proc. Natl Acad Sci* **117:** 9630–9641. doi:10.1073/pnas.1918145117

Gutierrez E, Shin BS, Woolstenhulme C, Kim JR, Saini P, Buskirk A, Dever T. 2013. eIF5A promotes translation of polyproline motifs. *Mol Cell* **51:** 35–45. doi:10.1016/j.molcel.2013.04.021

Guydosh NR, Green R. 2014. Dom34 rescues ribosomes in 3′ untranslated regions. *Cell* **156:** 950–962. doi:10.1016/j.cell.2014.02.006

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 Res* **24:** 842–851. doi:10.1038/cr.2014.74

Hinnebusch AG. 2014. The scanning mechanism of eukaryotic translation initiation. *Annu Rev Biochem* **83:** 779–812. doi:10.1146/annurev-biochem-060713-035802

Howard MT, Carlson BA, Anderson CB, Hatfield DL. 2013. Translational redefinition of UGA codons is regulated by selenium availability. *J. Biol. Chem.* **288:** 19401–19413. doi:10.1074/jbc.M113.481051

Ibrahim F, Maragkakis M, Alexiou P, Mourelatos Z. 2018. Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. *Nat Struct Mol Biol* **25:** 302–310. doi:10.1038/s41594-018-0042-8

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223. doi:10.1126/science.1168978

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. doi:10.1016/j.cell.2011.10.002

Ingolia NT, Hussmann JA, Weissman JS. 2019. Ribosome profiling: global views of translation. *Cold Spring Harb Perspect Biol* **11:** a032698. doi:10.1101/cshperspect.a032698

Irwin B, Heck JD, Hatfield GW. 1995. Codon pair utilization biases influence translational elongation step times. *J. Biol. Chem.* **270:** 22801–22806. doi:10.1074/jbc.270.39.22801

Ivanov IP, Shin BS, Loughran G, Tzani I, Young-Baird SK, Cao C, Atkins JF, Dever TE. 2018. Polyamine control of translation elongation regulates start site selection on antizyme inhibitor mRNA via ribosome queuing. *Mol Cell* **70:** 254–264.e6. doi:10.1016/j.molcel.2018.03.015

Janich P, Arpat A, Castelo-Szekely V, Lopes M, Gatfield D. 2015. Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. *Genome Res* **25:** 1848–1859. doi:10.1101/gr.195404.115

Janich P, Arpat AB, Castelo-Szekely V, Gatfield D. 2016. Analyzing the temporal regulation of translation efficiency in mouse liver. *Genom Data* **8:** 41–44. doi:10.1016/j.gdata.2016.03.004

Joazeiro CAP. 2019. Mechanisms and functions of ribosome-associated protein quality control. *Nat. Rev. Mol. Cell Biol.* **20:** 368–383. doi:10.1038/s41580-019-0118-2

Johnstone T, Bazzini A, Giraldez A. 2016. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J* **35:** 706–723. doi:10.15252/embj.201592759

Lesnik C, Golani-Armon A, Arava Y. 2015. Localized translation near the mitochondrial outer membrane: an update. *RNA Biol* **12:** 801–809. doi:10.1080/15476286.2015.1058686

Mariappan M, Li X, Stefanovic S, Sharma A, Mateja A, Keenan RJ, Hegde RS. 2010. A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466:** 1120–1124. doi:10.1038/nature09296

O'Connor PBF, Andreev DE, Baranov PV. 2016. Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. *Nat Commun* **7:** 12915. doi:10.1038/ncomms12915

Panasenko OO, Somasekharan SP, Villanyi Z, Zagatti M, Bezrukov F, Rashpa R, Cornut J, Iqbal J, Longis M, Carl SH, et al. 2019. Co-translational assembly of proteasome subunits in NOT1-containing assemblysomes. *Nat Struct Mol Biol* **26:** 110–120. doi:10.1038/s41594-018-0179-5

Pelechano V, Alepuz P. 2017. eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res* **45:** 7326–7338. doi:10.1093/nar/gkx479

Pop C, Rouskin S, Ingolia NT, Han L, Phizicky EM, Weissman JS, Koller D. 2014. Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol Syst Biol* **10:** 770. doi:10.15252/msb.20145524

R Core Team. 2016. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/.

Schuller AP, Green R. 2018. Roadblocks and resolutions in eukaryotic translation. *Nature* **19:** 526–541.

Schuller AP, Wu CCC, Dever TE, Buskirk AR, Green R. 2017. eIF5A functions globally in translation elongation and termination. *Mol Cell* **66:** 194–205.e5. doi:10.1016/j.molcel.2017.03.003

Shiber A, Döring 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. doi:10.1038/s41586-018-0462-y

Sinturel F, Gerber A, Mauvoisin D, Wang J, Gatfield D, Stubblefield JJ, Green CB, Gachon F, Schibler U. 2017. Diurnal oscillations in liver mass and cell size accompany ribosome assembly cycles. *Cell* **169:** 651–663.e14. doi:10.1016/j.cell.2017.04.015

Tuck AC, Rankova A, Arpat AB, Liechti LA, Hess D, Iesmantavicius V, Castelo-Szekely V, Gatfield D, Bühler M. 2020. Mammalian RNA decay pathways are highly specialized and widely linked to translation. *Mol Cell* **77:** 1222–1236.e13. doi:10.1016/j.molcel.2020.01.007

Vindry C, Ohlmann T, Chavatte L. 2018. Translation regulation of mammalian selenoproteins. *Biochim Biophys Acta* **1862:** 2480–2492. doi:10.1016/j.bbagen.2018.05.010

Wolin S, Walter P. 1988. Ribosome pausing and stacking during translation of a eukaryotic mrna. *EMBO J* **7:** 3559–3569. doi:10.1002/j.1460-2075.1988.tb03233.x

Wu CCC, Zinshteyn B, Wehner KA, Green R. 2019. High-resolution ribosome profiling defines discrete ribosome elongation states and translational regulation during cellular stress. *Mol Cell* **73:** 959–970.e5. doi:10.1016/j.molcel.2018.12.009

Yanagitani K, Kimata Y, Kadokura H, Kohno K. 2011. Translational pausing ensures membrane targeting and cytoplasmic splicing of XBP1u mRNA. *Science* **331:** 586–589. doi:10.1126/science.1197142

Yordanova, MM, Loughran, G, Zhdanov AV, Mariotti, M, Kiniry, SJ, O'Connor, PBF, Andreev, DE, Tzani, I, Saffert, P, Michel AM, et al. 2018. AMD1 mRNA employs ribosome stalling as a mechanism for molecular memory formation. *Nature* **553:** 356–360. doi:10.1038/nature25174

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. doi:10.1002/j.1460-2075.1986.tb04395.x

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 Syst* **5:** 212–220.e6. doi:10.1016/j.cels.2017.08.004